

Text as Data in Economic Analysis

Tarek A. Hassan, Stephan Hollander, Aakash Kalyani, Laurence van Lent, Markus Schwedeler, and Ahmed Tahoun

Advancements in information technology, coupled with the rapid growth of digitally accessible text, have greatly enhanced economists' ability to use text as data in their research (Baker, Bloom, and Davis 2016; Gentzkow, Kelly, and Taddy 2019). Today, the range of such text sources is virtually limitless, including books, contracts, transcripts, news articles, social media feeds, and many more. Of course, economists have long engaged in analyzing text, as evidenced, for example, by Romer and Romer (1989, 2023), who analyzed historical transcripts of Federal Reserve policymaking meetings. This article focuses on the technological advancements that have transformed this practice. We explore how these innovations enable faster, larger-scale, and more consistent processing of text, while also facilitating the measurement of multiple textual dimensions.

■ *Tarek A. Hassan is Professor of Economics, Boston University, and Co-founder and Chief Executive Officer, NL Analytics Inc., both in Boston, Massachusetts. Stephan Hollander is Professor of Financial Accounting, Tilburg University, Tilburg, the Netherlands. Aakash Kalyani is Research Economist, Federal Reserve Bank of St. Louis, St. Louis, Missouri. Laurence van Lent is Professor of Accounting and Economics, Frankfurt School of Finance and Management, Frankfurt, Germany. Markus Schwedeler is co-founder and Chief Executive Officer, NL Analytics Inc., Boston, Massachusetts. Ahmed Tahoun is Professor of Accounting at London Business School, London, United Kingdom. Hassan, Kalyani, Schwedeler, and Tahoun are co-founders of NL Analytics Inc., a startup that provides tools for earnings-call analysis to researchers. Their respective e-mail addresses are thassan@bu.edu, s.hollander@tilburguniversity.edu, aakash.kalyani@stls.frb.org, l.vanlent@fs.de, markus@nlanalytics.tech, and atahoun@london.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.20231365>.

Specifically, we focus on how analyzing largely unstructured corporate text—produced and utilized daily for decision-making by entrepreneurs and their firms’ shareholders, financial analysts, and other stakeholders—can enhance economic research for academic scholars, policymakers, and corporate management, as well as business and economics students. Traditional quantitative economic research has typically relied on survey data or accounting metrics, such as firms’ profit, investment, or employment figures. The use of text opens up a wealth of new data, offering insights into expectations, perceived risks, costs, and opportunities that are difficult or impossible to extract from conventional data sources.

To illustrate the power of text analysis, we will focus much of this paper on the example of analyzing text from corporate earnings calls, a particularly active area of current research. These calls provide a real-time window into how companies respond to unfolding events, offering insights into their strategies, concerns, and expectations. To illustrate, consider Russia’s large-scale invasion of Ukraine on February 24, 2022. Within hours of this escalation, Geox S.p.A., an Italian shoe manufacturer, held its quarterly earnings call. The call focused on the unfolding crisis, with analysts and investors inquiring about contingency plans, revenue dependencies, and strategic options in light of potential global embargoes, disrupted supply chains, and interrupted interbank communications. Geox was not unique in this regard—management of numerous other corporations engaged in similar discussions during earnings calls in the following days, anticipating many of the impacts economic policymakers had to confront over the coming years, ranging from high natural gas prices to increased defense budgets. Even beyond geopolitical crises, earnings calls routinely bring policy-relevant issues to the surface. Recent examples include ongoing concerns about the dangers of artificial intelligence (AI), the sources of persistent inflation, and trade policy uncertainty.

In this paper, we introduce a set of methods that allow researchers to systematically analyze the content of these earnings calls (and other texts) at scale. These methods enable researchers to measure the risks, costs, and opportunities that firms around the globe attribute to specific shocks, policies, and regulations. Building on these techniques, we show how to trace shock-induced risk transmission and potential contagion across firms, sectors, and regions; and how to measure present and past expectations regarding specific events, such as the 2003 SARS outbreak or the 2017 Tax Cuts and Jobs Act, in this way overcoming the limitations of retrospective surveys that attempt to reconstruct pre-event expectations (Gallemore et al. 2025).¹ Moreover, we demonstrate how to make cardinal comparisons of different risks, showing, for example, that in 2023, US-based firms allocated twice as much attention to inflation-related risks than to risks from geopolitics or trade wars—providing insights that are difficult or impossible to glean from other data.

¹ One possible reason why measuring past expectations related to historical events is important is provided by Hennessy and Strebulaev (2020). They demonstrate analytically that if firms’ behavior is influenced by expectations about a policy, neglecting to account for these expectations in empirical analyses can distort inferences about the causal effects of that policy.

Importantly, the transcripts of these earnings calls are widely available almost immediately, allowing quantitative analysis without the months-long lags typical of accounting or survey data.

Beyond the specific example of earnings calls, we argue that analyzing the full range of corporate text data—including job postings and patents—represents a sea change in our ability to track and understand economic shocks. At a narrow level, discerning why and how firms respond to proposed reforms or technological changes can help shape more effective policy design and facilitate smoother adaptation to economic shocks. On a broader scale, tracing the propagation of supply, demand, and other economic shocks across firms, sectors, and countries is key for testing and refining theories of how the economy functions.

In the subsequent sections, we begin by introducing earnings conference call transcripts as a data source, then outline the building blocks of our approach to creating measures of firms' exposure to specific shocks using these data. We discuss the options available to researchers in text classification tasks, including keyword-based approaches, the use of "training libraries," and large language models (LLMs). We then introduce and discuss straightforward techniques that enable the measurement of the risks, costs, and opportunities that decision-makers associate with specific shocks and policies, emphasizing the advantages they provide in terms of cardinal comparisons of risk, differentiating between the effects of bad news versus increased risks, and analyzing the transmission of shocks across countries and firms. Finally, we show how to expand these techniques to other corporate text sources, including patents and job postings. We demonstrate how to use these sources to trace the origin and spread of technological innovations across geography, industries, firms, and occupations.

In sum, this article offers a glimpse of how scholars, central banks, think tanks, and international organizations are already integrating text-analysis techniques for a range of purposes in their research and policymaking. Throughout, we contend that the main frontier of text as data in economic analysis is not in developing more advanced text-analysis techniques, but in making vast, untapped text corpora accessible to scholars, economists, and policymakers (Bae, Hung, and van Lent 2023).

Earnings Conference Calls: A Marketplace of Information

Corporate communication has changed substantially over the past decades. Before 2000, publicly listed firms primarily communicated through accounting numbers. However, these accounting numbers are inherently backward-looking and slow to reflect new information: they document past events, take months to prepare, and incorporate changes into earnings figures only gradually due to accounting principles that prioritize caution. For this reason, firms' management would typically provide more information and context for their accounting disclosures in private conversations with selected journalists, analysts, and investors.

In 2000, the Securities and Exchange Commission (SEC) made this practice illegal and instead required publicly traded companies to share any such context with the general public. While earnings conference calls—held each quarter after the release of a firm’s financial results—were practiced prior to this regulation, the new requirements for simultaneous public disclosure accelerated their widespread adoption for discussing a firm’s key financial information. This shift made earnings calls the primary channel through which market participants—institutional investors, financial analysts, and individual shareholders alike—could access management’s interpretation of their firms’ performance.

During a typical corporate earnings call, an executive of the firm (often the chief executive officer or the chief financial officer) reviews the firm’s recent financial performance, discusses operational challenges, and outlines any issues that impact the firm’s future prospects. Financial analysts then pose questions during a question-and-answer session, prompting further discussion of the company’s strategies. These conversations last up to an hour. They are livestreamed online, and then later transcribed and made available. Primarily due to this interactive nature, earnings calls offer substantial complementary textual narratives for nearly every set of accounting numbers available through databases like S&P’s Compustat, and provide valuable context and additional information about firms’ current and future situations.²

Economists have historically been skeptical of using information on what economic decision makers *say*, favoring instead evidence from what they *do* (choices and behaviors), because of concerns about “cheap talk”—the perception that decision makers can spin information or outright lie to gain an economic advantage. Notwithstanding this, we argue that earnings calls are a valuable venue where what participants say is highly informative about economic behavior. Indeed, unlike casual statements or survey responses, communications during earnings calls carry real consequences. For instance, securities law enforces significant penalties for making materially misleading statements, while the scrutiny of sophisticated analysts, who can challenge management’s claims, generates reputational costs for evasive or inconsistent responses (Hollander, Pronk, and Roelofsen 2010; Rogers, Van Buskirk, and Zechman 2011). Perhaps most importantly, market participants have the option (and are known) to follow up and verify statements made during earnings calls by comparing them to the firm’s subsequent behavior (for example, Suijs 2007).

These institutional features distinguish earnings calls from other contexts where talk may be cheap. Understanding this distinction helps in determining when text data will likely offer reliable economic insights. In general, corporate communication is most informative when statements involve explicit or implicit costs and can

² Even when accounting numbers are available, they provide limited insight into how or why shocks affect firms. For example, when profits drop due to lower sales, identifying the underlying causes—such as supply chain disruptions, demand contractions, or policy interventions—necessitates additional context. Earnings conference calls provide an opportunity to obtain this additional information by questioning management.

be cross-validated against observable outcomes, and when institutional structures create incentives for truthful communication.

Many international firms, listed on a stock exchange, either conduct their earnings calls in English or provide translations, enabling rapid assembly of directly comparable corpora across more than 12,500 firms from nearly 90 countries. Primary sources for earnings-call transcripts include corporate investor relations websites and public financial platforms like Yahoo! Finance and Seeking Alpha. Commercial data providers, including the London Stock Exchange Group (LSEG) and S&P, standardize and aggregate these transcripts, making them available through institutional subscriptions commonly held by universities. The earnings call data used in this paper come from LSEG through NL Analytics, a startup founded by some of the authors of this paper, which offers online tools for earnings call analysis to researchers, central banks, and international organizations.

Decoding Shocks: Exposure, Risk, and Sentiment

Exposure

The time people spend discussing a topic reveals its importance to them—a simple yet powerful insight that opens new possibilities for measuring the importance of different topics in corporate communication. This observation becomes particularly valuable in the context of earnings conference calls, where executives have less than an hour to address their most pressing concerns (Matsumoto, Pronk, and Roelofsen 2011). The time constraint creates a natural prioritization: topics that receive more attention in these discussions are, by necessity, those that executives and noncorporate participants consider most significant.

This relation between discussion time and importance allows us to calculate a precise measure of firms' exposure to specific events or issues, which we call "topic exposure." Specifically, by calculating the proportion of sentences devoted to a particular topic during an earnings call, we can quantify its relative significance to the firm. For example, counting sentences about Russia's invasion of Ukraine as a percentage of total sentences reveals which firms face meaningful exposure to the conflict and how this exposure compares to other concerns. Extensive audits with human coders strongly support the validity of this classification approach (Hassan et al. 2019, 2023, 2024a, 2024b).

Risk

One application of this idea is measuring the overall risk a given firm faces at a particular moment by calculating the proportion of the earnings-call discussion devoted to risk or uncertainty. While economists define "risk" as the variance of outcomes, participants in earnings calls may use the term more loosely, frequently focusing on adverse scenarios ("downside risk"). To bridge this gap, we construct a comprehensive dictionary of risk-related terms by identifying single-word synonyms of "risk," "risky," "uncertain," and "uncertainty" from the Oxford English

Dictionary, excluding only the terms “question” and “questions,” which of course carry a different meaning in earnings calls. This systematic approach yields terms across a spectrum, from those suggesting volatility (“variable,” “variability”), to those indicating upside and downside potential (“chance,” “prospect,” “pending”). This breadth of terms ensures we capture the full range of risks firms face, not just their concerns about adverse scenarios.³

Applying this dictionary to analyze earnings-call discussions yields a firm-level measure of “overall risk”: the percentage of sentences in a given firm’s earnings call that mention a synonym for risk or uncertainty. Figure 1, panel A, shows the average of overall risk across all firms with a US stock market listing. This measure of economy-wide risk reveals distinct trends. During the “Great Moderation,” from 2002 to 2007, risk mentions remained relatively subdued, with fewer than 1.6 percent of sentences containing risk-related terms (Benati and Surico 2009; Galí and Gambetti 2009). The global financial crisis, however, marked a sharp departure from this pattern, with risk discussions peaking at 2.2 percent of earnings call content in late 2008. Subsequent spikes emerged during the European sovereign debt crisis of 2011–2012, and risk-related discussions peaked at 2.8 percent during the Covid-19 pandemic.

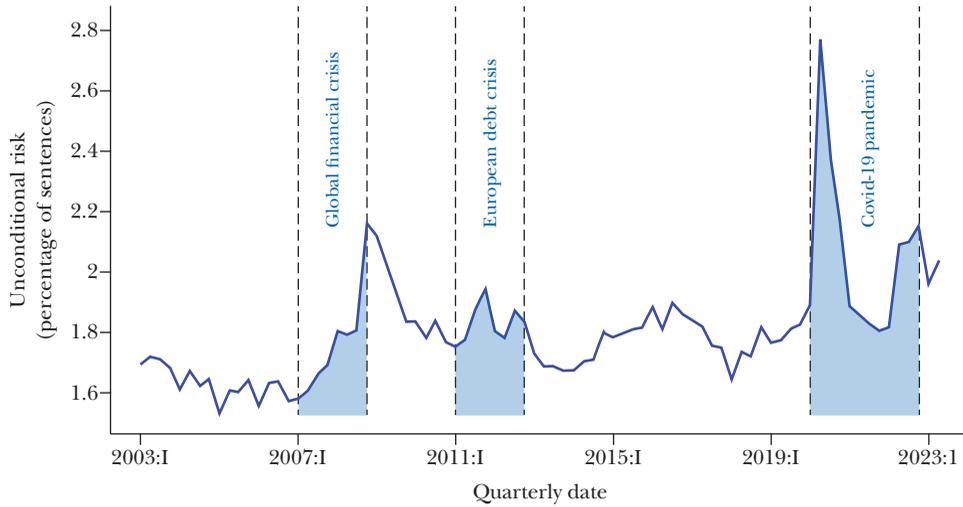
At this aggregate level, the pattern of risk-related discussions in earnings calls correlates strongly with market-based measures of risk such as the Chicago Board Options Exchange volatility index, known as the VIX. (The VIX reflects the likelihood of large price swings in the S&P 500 index over the next 30 days, based on prices of stock options.) However, compared to the VIX and other aggregate market-based indicators, a text-based measure offers a distinct advantage: straightforward disaggregation. Indeed, starting with the firm-level risk-related discussions, we can construct detailed, disaggregated metrics by sector and region (who is perceiving risk), and even analyze risk patterns for specific topics (what is the risk about).

To measure “topic risk,” we calculate the proportion of earnings-call sentences that mention both risk-related terms and a specific subject, relative to the total number of sentences in the call. Figure 1, panel B, illustrates this topic-specific approach by tracing Brexit-related risks discussed by firms around the world (Hassan et al. 2024a). The time-series pattern shows notable spikes in “Brexit risk” that correspond with key political developments during that period. To illustrate, discussions of Brexit-related risk surged immediately after the 2016 UK referendum vote to leave the European Union. While this initial uncertainty subsided somewhat in 2017, it intensified sharply in the second half of 2018 as the details of Prime Minister Theresa May’s agreement with the European Union became clear, along with the challenges of securing parliamentary approval. The prospect of a no-deal Brexit under the subsequent Prime Minister, Boris Johnson, kept uncertainty elevated through 2019, until Brexit finally took effect on January 31, 2020.

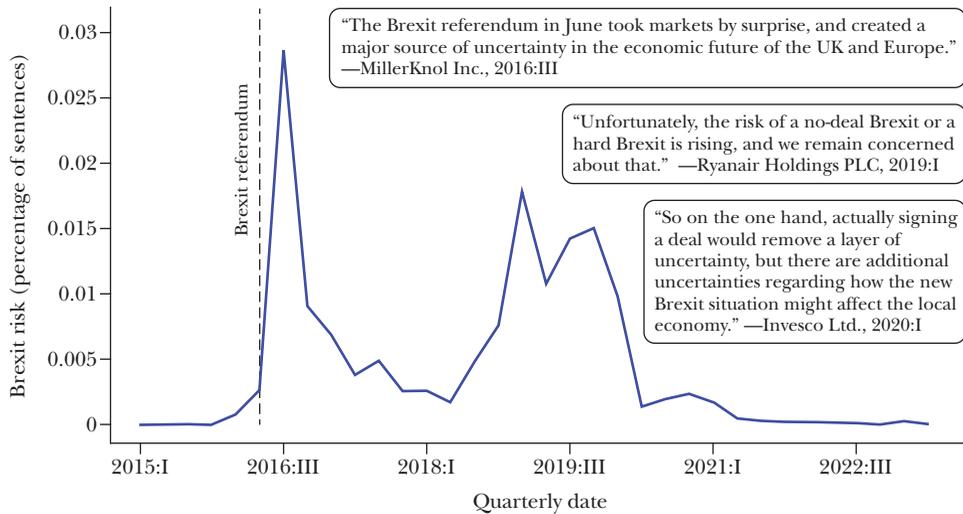
³ In comparison, Baker, Bloom, and Davis (2016) construct their measure of economic policy uncertainty by including only the terms “uncertain” and “uncertainty,” which focuses their analysis on uncertainty rather than broader notions of risk.

Figure 1
Text-Based Risk

Panel A. Overall risk



Panel B. Brexit risk



Source: Data sourced from www.nlanalytics.tech.

Note: This figure displays overall risk (panel A) and Brexit risk (panel B) as a percentage of sentences in earnings calls. The sample comprises 379,227 earnings conference calls, representing 12,805 firms headquartered in 89 countries.

By intersecting discussions of risk with discussions of specific topics, we can thus assess how much risk a given firm associates with a given specific topic (such as Brexit).

The figure highlights another key advantage of this text-based measurement approach: its capacity for self-labeling. In other words, the reasons behind observed fluctuations in text-based risk measures become apparent by examining the underlying discussions from the transcripts that contribute to each spike. As the figure illustrates, excerpts from high-scoring earnings-call transcripts enable us to directly link periods of elevated Brexit risk to specific political events and their anticipated economic impacts.

Sentiment

Measuring sentiment from text involves identifying terms that indicate how positively or negatively firms discuss specific topics. The most straightforward method to flag earnings-call sentences as positive is if they contain words like “good,” “strong,” or “great” without negative terms, and negative if they include words like “loss,” “decline,” or “difficult” without positive ones. Loughran and McDonald (2011) provide a comprehensive sentiment lexicon specifically calibrated for financial text.

Recent advances in natural language processing allow for understanding words in their full context without relying on specialized dictionaries. However, systematically identifying positive- and negative-toned discussions naturally complements the earlier discussion on measuring topics and risk through word frequency. Indeed, similar to how a simple count of topic-specific terms reveals firm exposure and risk-related terms capture uncertainty, tracking sentiment-related words indicates whether firms perceive economic shocks as having a positive or negative impact. Following this logic, “topic sentiment” counts sentences that mention a given topic with positive sentiment, subtracts the count of those with negative sentiment, and normalizes by the total number of sentences in the call.

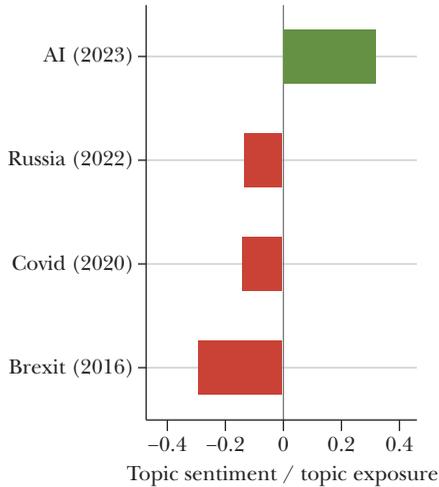
Figure 2 shows how topic-sentiment measures capture firms’ responses to major economic shocks. Specifically, panel A compares average sentiment across several significant events: Russia’s 2022 invasion of Ukraine, the Covid-19 outbreak in 2020, Brexit-related developments in 2016, and breakthroughs in artificial intelligence, particularly the release of large language models, in 2023. The measures indicate systematically negative sentiment in discussions related to Brexit, Covid-19, and Russia’s invasion, while discussions related to AI trend positive, with 42.1 percent of sentences related to AI containing positive terms, compared to 10.2 percent negative, yielding a net positive sentiment of +32 percent.

These aggregate patterns mask substantial heterogeneity at the firm, sector, and country levels. Panel B shows firm-level variation in discussions related to artificial intelligence. Technology firms heavily invested in development of AI, such as Alphabet, NVIDIA, and Microsoft, express strongly positive sentiment. In contrast, content producers like News Corp, Universal Music Group, and Warner Music Group discuss AI with notably negative sentiment, reflecting concerns about intellectual property protection.⁴

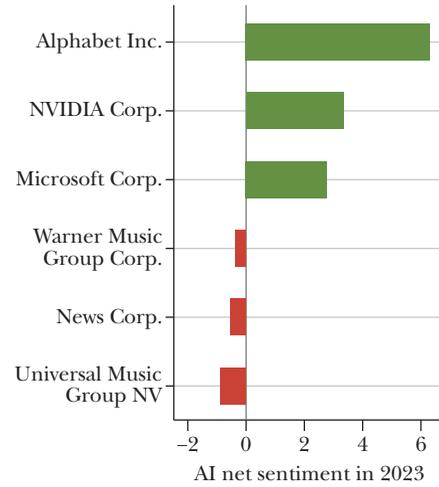
⁴ For additional examples of AI-related excerpts from earnings-call conversations, see Supplemental Appendix Table 1.

Figure 2
Text-Based Sentiment

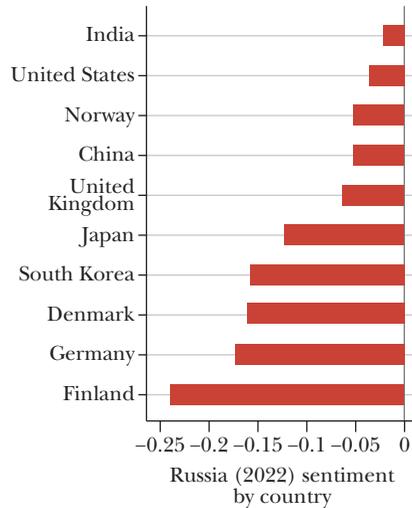
Panel A. Sentiment by event



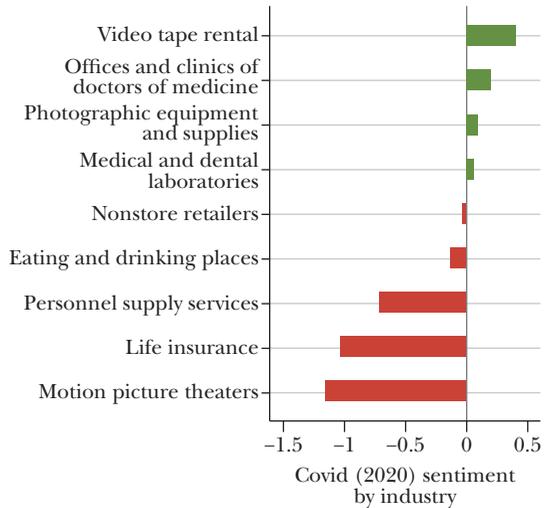
Panel B. AI (2023) sentiment by firm



Panel C. Russia (2022) sentiment by country



Panel D. Covid (2020) sentiment by industry



Source: Data sourced from www.nlanalytics.tech.

Note: This figure illustrates topic sentiment as follows: panel A shows topic-specific (namely, Russia (2022), Covid-19 (2020), Brexit (2016), and AI (2023)) sentiment divided by topic-specific exposure; panel B shows AI (2023) sentiment, as a percentage of sentences, by firm; panel C shows Russia (2022) sentiment, as a percentage of sentences, by country; and panel D shows Covid-19 (2020) sentiment, as a percentage of sentences, by three-digit SIC sector. To calculate topic-specific sentiment on the aggregate or by firm, country, or industry, we take an average across earnings calls of topic sentiment where comprises of topic-specific keywords. To illustrate, the topic Russia includes keywords that relate to the country, such as “Russia,” “Russian,” “Ruble,” and “Mr Putin.” Similarly, for the topic of AI, in addition to “AI” and “artificial intelligence” we included terms like “ChatGPT,” “LLM,” “deep learning,” and “machine learning.”

Panels C and D display variation in sentiment across geographic regions and sectors, respectively. Panel C shows that Finnish, German, and Danish firms expressed particularly negative sentiment about Russia-related developments in 2022, with discussions centered on regional security, energy markets, and economic relationships. The transcripts show Japanese and South Korean firms focusing their concerns specifically on trade access through Russian territory.

Even during the same economic shock, sentiment varies substantially across sectors. Panel D shows that early in the Covid-19 pandemic, the Video Tape Rental and Medical Services sectors anticipated benefits, reflecting expected increases in streaming and healthcare demand. In contrast, Life Insurance and Motion Picture Theater sectors projected strongly negative impacts, anticipating increased mortality-related payouts and restricted public gatherings. Collectively, these findings suggest that topic-sentiment analysis is a powerful tool, enabling us to characterize the heterogeneous impact of economic shocks at the firm, sector, and country levels.

One important caveat is that earnings-call language has become more positively toned over time, as executives learned to steer clear of words that might trigger algorithmic trades (Cao et al. 2023). This upward drift inflates raw sentiment, but it is straightforward to purge with time fixed effects.

Lexicon Construction

Each of our measures of topic-specific exposure, risk, and sentiment involves a systematic procedure for identifying sentences that mention a particular topic. For this task, researchers have three options: keyword-based searches, training libraries, and machine learning. These options differ in the degree to which they depend on human judgement, but they all share a common goal: systematically capturing all relevant content (that is, minimizing false negatives), while avoiding spurious mentions (that is, minimizing false positives).

Keyword-Based Approach

The simplest approach to identifying discussions on a given topic is by selecting relevant keywords (Baker, Bloom, and Davis 2016). In our experience, this approach is particularly effective when the topic of interest pertains to a specific, narrowly defined event or policy. For instance, in Hassan et al. (2024a), the single keyword “Brexit” sufficed to accurately capture conversations related to the United Kingdom’s exit from the European Union. Similarly effective examples include Covid-19, supply chain disruptions, and cyber risk, where a handful of keywords suffice.

The process of choosing keywords usually begins with a set of annotated sentences obtained from human readers, flagging those that discuss the topic of interest.⁵ Next, we select keywords or phrases that mimic this human classification

⁵ This systematic approach requires the topic to have existed long enough for researchers to annotate a sufficiently large number of sentences, which in some cases can present challenges. In the case of

within the sample of previously annotated text. These keywords or phrases are frequent in sentences discussing the topic of interest and infrequent in those that do not. To check their effectiveness out-of-sample, we sample new sentences from our corpus and assess the rate of false positives. This process may prompt changes to our set of keywords, followed by further sampling and iteration until we identify a set that performs well on our corpus. For example, to identify supply chain-related discussions in Hassan et al. (2023), we started with “supplier” and “supply” and, thus, iteratively built a comprehensive set of keywords.

Training Libraries

For a broader or less precisely defined event or policy, a keyword-based approach can become unwieldy. In such cases, relying on “training libraries” rather than human judgement can be a more effective solution. A straightforward approach in this context is to use the numerical statistic known as term frequency–inverse document frequency (Manning, Raghavan, and Schütze 2008; Song and Wu 2008). We apply this approach in Hassan et al. (2019, 2024b) to differentiate between political and nonpolitical risk discussions and to assess the risks firms associate with specific countries.

This approach requires a sufficiently large corpus of text representative of discussions on the given topic. For example, to differentiate between political and nonpolitical discussions, we use two textbooks: one focused on politics and the other on a nonpolitical subject (in our context, we opted for an accounting textbook). From these sources, we extract and tally the occurrences of bigrams (that is, two-word combinations), identifying those that are frequent in one training library but not in the other. For example, the bigrams “Joe Biden” and “Congress passed” might be frequent in discussions related to politics, but never or rarely occur in discussions on nonpolitical subjects. This typically results in a list of several thousands of topic-related bigrams, which are assigned weights proportional to their frequency in the training library.

Compared to a keyword-based approach, training libraries generally capture a much broader range of language used to discuss a topic, allowing for the detection of more subtle and indirect references to the topic in question.

Machine Learning

The third approach leverages machine learning to aid in keyword selection. In this method, the common practice is for a machine learning model to propose keywords to the researcher, who then decides whether to accept or reject them based on how well they align with their definition of the topic in question. For example, Sautner et al. (2023) utilize training libraries to train a machine learning model to distinguish between climate- and non-climate-related discussions in earnings calls.

Covid-19, as studied in Hassan et al. (2023), no such training data were available at the onset of the outbreak. As a result, we relied on a list of Covid-19 synonyms provided by the World Health Organization, supplemented by cross-referencing contemporaneous newspaper articles and online resources.

The model then proposes phrases that most effectively make this distinction. For each suggestion, they randomly sample from the text of earnings calls that mention the proposed phrase and accept it if the false positive rate falls below a certain threshold.

In contrast, Kalyani et. al. (2025a) uses a large language model to make suggestions. To identify job postings for research and development positions, the authors start with a set of seed phrases (for example, “research,” “and develop”) and let the LLM suggest additional phrases commonly used in similar contexts. They then randomly sample excerpts from job postings for each suggestion to check for false positives and include those suggested phrases that meet the threshold for true positives.

A more evolved version of this approach completely delegates the keyword selection-rejection decision to a large language model. In this approach, the researcher provides a verbal description of the topic and allows the LLM to determine whether a given sentence or paragraph from an earnings-call transcript aligns with this definition (Jha et al. 2024). The drawback of this fully automated approach, in our view, is that the researcher effectively relinquishes control over the definition of the topic and hands it over to the LLM. This can lead to issues if not handled carefully. One issue is replicability—different LLMs may make different selection-rejection decisions for reasons that are difficult, if not impossible, to interpret afterward. Moreover, particularly when dealing with historical data, LLMs are prone to hindsight bias. For example, an LLM might link “Fukushima” to a nuclear disaster, even when analyzing text from the time before the disaster occurred (Sarkar and Vafa 2024).⁶

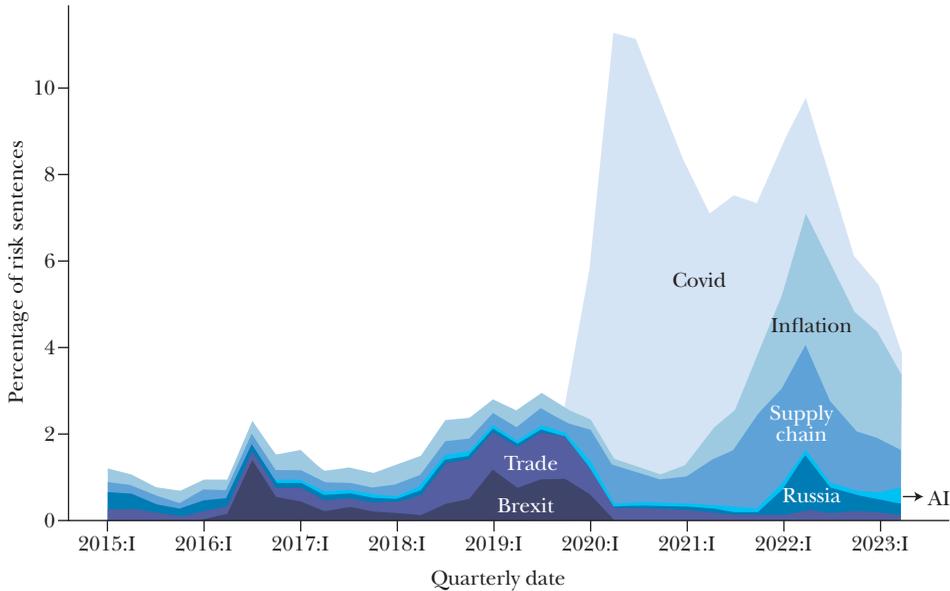
In conclusion, depending on the specific application, a researcher should thoughtfully choose one of the approaches outlined above to build a lexicon. Our experience suggests the virtues of erring on the side of simplicity and opting for a more straightforward approach whenever feasible. Sample code illustrating each approach is available in Hassan et al. (2025b), which can be found at the landing page for this article at the American Economic Association’s website.

Text Analysis Applications in Economic Research

We next present three applications that show how our text-based measures of exposure, risk, and sentiment allow us to track the impact of economic shocks in novel ways that would be hard or impossible to achieve with conventional data. These applications demonstrate how to decompose the relative importance of different risks, separate direct effects from uncertainty impacts, and track the transmission of shocks across global markets.

⁶ While current large language models offer promising capabilities for keyword selection, their relative advantage over lower-tech methods may evolve as both technologies advance. Future improvements in model efficiency, interpretability, and cost-effectiveness could shift the balance, making it important to regularly reassess their value in comparison to traditional approaches.

Figure 3
Risk Decomposition



Source: Data sourced from www.nlanalytics.tech.

Note: This figure shows, as a percentage of sentences, a breakdown of risk related to AI, Brexit, Covid-19, inflation, Russia, the supply chain, and trade for our sample of US-listed firms.

Cardinal Risk Decomposition

Our first application illustrates how text analysis facilitates a cardinal comparison of different sources of risk. This approach provides a more precise understanding of how each specific risk contributes to a firm's overall risk profile and can help policymakers and analysts identify which risks are most pressing.

To fix ideas, note from Figure 1 that in 2016:III, firms devoted 1.90 percent of all sentences to discussing risks and 0.029 percent specifically to discussing Brexit risk. We can thus conclude that Brexit accounted for 1.53 percent of overall risk-related discussions in that quarter. This decomposition of overall risk into specific topics is an important addition to the literature. The closest analog is the attribution of excess returns to risk factors in empirical asset pricing. However, these methods generally do not allow an assessment of which specific shock contributes to the overall uncertainty faced by a firm at a given time, even when paired with an event study.

Figure 3 applies this decomposition more broadly to track the evolution of different sources of risk over time. The data reveal that trade policy uncertainty reached its peak during the trade war in the first term of the Trump administration, representing 0.98 percent of risk mentions in 2018:III—comparable to Brexit-related concerns. The Covid-19 crisis subsequently emerged as the dominant source of risk, accounting for 9.88 percent of the risk-related discussions in 2020:II.

As the pandemic's direct effects subsided, firms' focus shifted toward inflation and supply chain disruptions. Although Russia's invasion of Ukraine garnered substantial media coverage, its effect on firm-level risk discussions was generally limited, with notable exceptions in certain regions. For example, German firms devoted 4.02 percent of their risk discussions to Russia in 2022:I.

This cardinal decomposition provides two key advantages. First, it enables a systematic comparison of the relative importance of different (sources of) risks at any given time, helping policymakers understand which forces are the primary drivers of risk for specific (groups of) firms. Second, the ability to track the evolution of these proportions over time allows for the analysis of how the importance of different risk sources rises and falls. The consistent measurement framework allows us to make meaningful statements about how firms' concerns with one source of risk compare to another, both at a given point in time (that is, cross-sectionally) and over time, even as the overall volume of risk discussion fluctuates.

Distinguish Bad News from Uncertainty Effects

Our second application demonstrates how to separate the effects of "bad news" from those of uncertainty in firm-level data. To illustrate the point, consider the management of a music company learning about advancements in artificial intelligence capable of imitating artists' voices. These advancements represent both bad news (that is, a potential decline in the company's expected future cash flows due to competition from content generated by AI) and increased uncertainty (for example, from the possibility of legal action or changes in legislation to protect copyright). If we observe a subsequent decline in this music company's hiring or planned investments, it raises a question: Is this decrease driven by the anticipated negative impact of the AI advancements on its business, or is it the result of increased uncertainty regarding future legislation or court rulings? Separating both effects in the data is typically challenging, as negative news about a firm's prospects often coincides with increased uncertainty about its future (Berger, Dew-Becker, and Giglio 2020).

With our text-based measures of AI sentiment and AI risk, we can readily separate between the two. For example, Supplemental Appendix Table 1 shows an excerpt from one of Chegg's earnings calls, in which the education technology company stresses the negative impact of ChatGPT's launch: students were abandoning the firm's subscription products for it. For Chegg, the boom in AI is unambiguously bad news, yielding a strongly negative measure of AI sentiment. In contrast, Spotify and News Corp emphasized the uncertainty effects in their earnings calls, particularly focusing on the future intellectual property landscape and potential regulatory changes—yielding higher levels of AI risk.

Suppose a researcher wants to examine how firms adjust their investment decisions in response to the artificial intelligence boom. The dependent variable is the investment rate for a given firm in a given quarter. Textual classification of firms' earnings-call transcripts provides three key independent variables: a measure of overall risk (the percentage of sentences discussing risk, irrespective of AI), topic-specific risk measures for AI, and AI sentiment expressed in discussions of AI.

Table 1
Risk, Sentiment, and Investment Rate

	<i>Investment Rate</i> _{<i>i,t</i>} × 100		
	(1)	(2)	(3)
<i>Unconditional Risk</i> _{<i>i,t</i>}	−0.039 (0.006)	−0.039 (0.006)	−0.039 (0.006)
<i>AI Risk</i> _{<i>i,t</i>}		−0.333 (0.125)	
<i>AI Positive Sentiment</i> _{<i>i,t</i>}		0.073 (0.041)	
<i>AI Negative Sentiment</i> _{<i>i,t</i>}		−0.002 (0.077)	
<i>Brexit Risk</i> _{<i>i,t</i>} pre-2019			−0.298 (0.136)
<i>Brexit Risk</i> _{<i>i,t</i>} post-2019			0.119 (0.326)
<i>Brexit Negative Sentiment</i> _{<i>i,t</i>} pre-2019			0.095 (0.118)
<i>Brexit Negative Sentiment</i> _{<i>i,t</i>} post-2019			−0.256 (0.154)
Sector fixed effects	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
Observations	84,500	84,500	84,500

Note: This table reports estimation results from a regression of firm *i*'s investment rate at time (quarterly) *t* on text-derived risk and sentiment variables. Investment rate is calculated as the ratio of capital expenditures reported by firm *i* at time *t* to the stock of property, plant, and equipment reported at time *t*−1. All risk and sentiment variables are in units of percentage of the total number of sentences in earnings calls. *Unconditional Risk*_{*i,t*} is the percentage of sentences with risk synonyms in an earnings call hosted by firm *i* at time *t*. *AI (Brexit) Risk* and *Sentiment* is the percentage of sentences with AI (Brexit) keywords mentioned in conjunction with a risk synonym as defined on page 5. *Brexit Risk* pre-2019 (similarly for post-2019) is the interaction of *Brexit Risk* with a dummy equal to one for the period before (post) 2019; zero otherwise. Standard errors are clustered by firm. All specifications include controls for sector (2 digit SIC) fixed effects and time (quarter) fixed effects. Only firms with at least ten earnings calls during 2015–2023 are included.

Regression results for such an analysis can be found in Table 1. As a general pattern, firms tend to reduce investment significantly when faced with higher overall risk, supporting the notion that they prefer to wait for more clarity before committing resources. The response to AI, however, reveals more nuanced patterns. While firms that discuss AI-related risks typically invest less, those expressing positive sentiments about AI's impact on their business significantly increase their investment rates.

The following column of Table 1 repeats the same analysis for Brexit risk and sentiment. Up until 2019, higher Brexit-related risk was associated with significantly lower investment across firms. However, after Brexit's implementation on January 31, 2020, it no longer exhibits a statistically significant relation with investment patterns. In contrast, negative sentiment about Brexit's impact on business operations now emerges as the key explanatory variable: firms that view Brexit as detrimental to their

business reduce investment, while other firms do not. This pattern suggests that, once Brexit's form became clearer, firms' assessment of its future cash flow effects, rather than uncertainty surrounding the policy, guided their investment decisions.

Transmission of Shock-Induced Risks across Countries

Textual analysis can also be used to assess the transmission of economic shocks across the globe. By analyzing how firms from different countries discuss risks related to specific foreign countries in their earnings calls, we can trace the propagation of shocks from one country to another. Figure 4 depicts this by presenting two contrasting examples: discussions of risks relating to the United Kingdom ("UK Risk") during the Brexit crisis in 2016 and discussions of risks relating to the US economy ("US Risk") during the global financial crisis of 2008.

Panel A of Figure 4 shows how discussions on UK-related risk evolved across countries after the Brexit referendum.⁷ The x -axis represents the average level of UK-related risk discussions among firms in each country in noncrisis times, while the y -axis indicates the level of UK-related risk in the two quarters after the 2016 Brexit referendum. To analyze these data systematically, we focus on three key metrics. First, the "global impact" measure of 1.31 quantifies how much the median country's UK-related risk discussion surpassed expected precrisis average, measured by the median vertical distance from the 45-degree line (that is, the median country increased its UK-related risk discussions by 31 percent.) Second, the steep slope coefficient of 1.51 signifies the strength of "bilateral transmission," indicating that countries where firms previously engaged more in UK-related risk discussions saw significantly greater increases in such discussions. Finally, the high R^2 of 0.86 reflects strong "regularity" in transmission, demonstrating how predictably the increase in risk discussions followed patterns based on preexisting economic relationships.

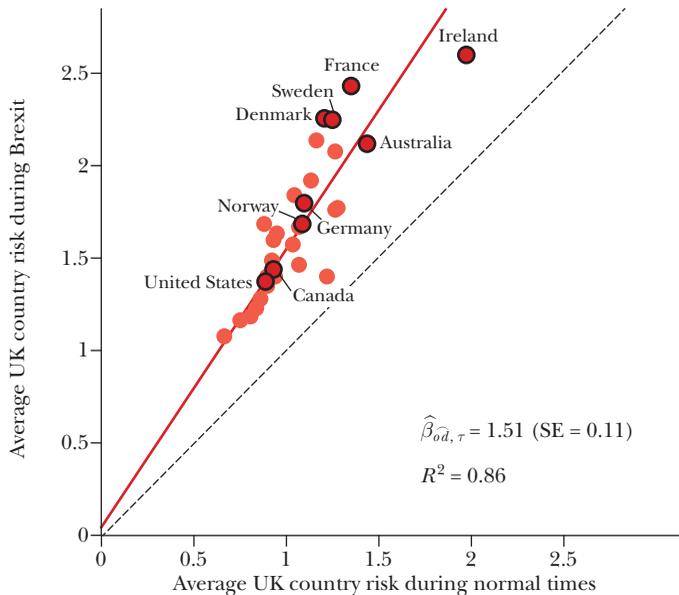
Together, these three metrics provide insights into the nature of Brexit as an economic shock. The relatively moderate global impact combined with strong bilateral transmission suggests that the shock propagated primarily through existing economic relationships rather than leading to widespread global disruption. The high regularity indicates that firms' precrisis discussions of UK-related risks largely predicted their post-referendum UK-related concerns. This pattern is evident in the data: Irish firms, with significant precrisis exposure, show the highest risk concerns due to their deep integration with the United Kingdom, while firms in France, Germany, and Sweden, with more limited precrisis exposure, exhibit increases in risk in sync with their pre-referendum exposure. That is, the Brexit shock largely impacted firms through established economic channels rather than creating systemic global risk.

Panel B of Figure 4 shows how US-related risk discussions spread during the global financial crisis of 2008. The x -axis represents precrisis levels of US-related

⁷ To identify discussions of specific foreign countries, we use training libraries for each of 45 countries sourced from the Economist Intelligence Unit Country Commerce Reports, in combination with term frequency-inverse document frequency as described above.

Figure 4
Country Risk Transmission

Panel A. Transmission of UK country risk after the Brexit referendum



(continued)

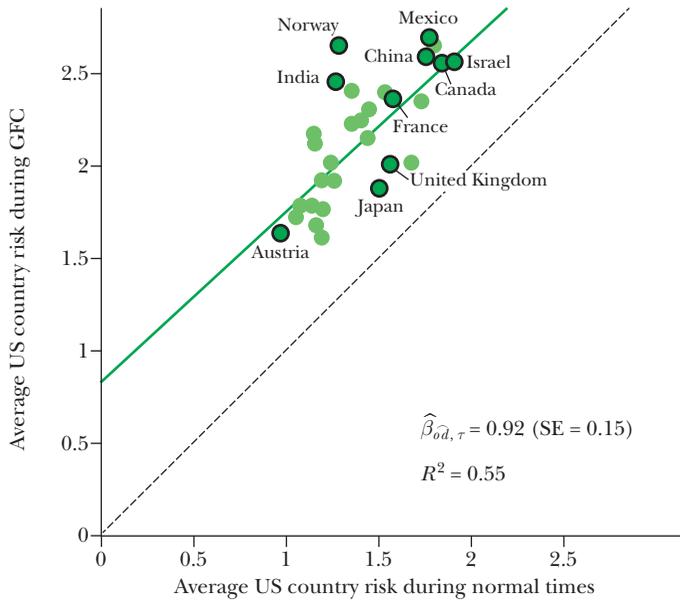
risk discussion among firms in each country, while the y-axis indicates the levels during Q1–Q3 of 2008. (We focus on US-related risk because the 2008 financial crisis notably accelerated with the US housing bubble collapse and the subsequent subprime mortgage crisis, eventually leading to a recession in the United States beginning in December 2007.) The pattern differs markedly from the Brexit case shown in panel A. The reduced slope coefficient (0.92) indicates weaker bilateral transmission, meaning that a country's precrisis exposure to US-related risks has little predictive power for its increase in risk discussion during the crisis. The lower R^2 of 0.55 suggests much less regularity in transmission, implying that the crisis spread through channels that precrisis economic relationships were unable to predict.

Despite weaker bilateral transmission, the 2008 financial crisis had a larger global effect than Brexit (2.27), with substantial increases in US-related risk discussion across *all* countries, regardless of their precrisis exposure. This is evident in the data: countries like China, Mexico, and Norway are grouped at higher levels of risk discussion during the crisis, despite differing widely in their precrisis exposure. The more scattered pattern of observations and their overall upward shift, compared to the concentrated pattern seen with Brexit (panel A), suggests that the financial crisis generated systemic global risk concerns that transcended traditional economic relationships.

Figure 4 (continued)

Country Risk Transmission

Panel B. Transmission of US country risk during the global financial crisis



Source: Hassan et al. (2024b).

Note: Panel A plots each country's average UK risk from earnings calls in the two quarters after the Brexit referendum (June 23) against its average UK risk in all other periods. In panel B, we repeat the same plot for US risk with and without the first three quarters of the global financial crisis: Q1–Q3 of 2008. To identify US-related risk (and similarly for UK risk) we use terms from Hassan et al. (2024b) that identify US-related conversations—such as “United States,” “US patent,” “USA,” and “the DoJ”—using training libraries of archetypical text for multiple different countries. For each panel, we report the coefficient along with standard errors and the R^2 .

This contrast between the two crises, shown in panels A and B of Figure 4, illustrates how these simple text-based metrics can capture fundamental differences in shock transmission. While the Brexit shock event primarily affected firms through established dependencies, the financial crisis spread through intricate financial linkages, creating widespread uncertainty that existing economic relationships could not predict.

Text as Data: Lessons and Pitfalls

In this section, we identify three common challenges researchers face when working with text as data and discuss effective strategies for overcoming them.

Preprocessing

Textual analysis starts with preprocessing—cleaning and transforming raw text into a structured format suitable for analysis. The first challenge revolves around preprocessing choices that can affect the quality of measurements. Unlike many other text-analysis applications in computational linguistics, our experience with earnings-call transcripts suggests that invasive preprocessing techniques such as stemming (treating words with the same stem as identical), lemmatizing (standardizing variant forms of a word), and removing punctuation are generally not beneficial. Furthermore, although we usually remove capitalization, we find that preserving it can sometimes be useful—for example, when identifying discussions of different countries, we might want to distinguish the animal “turkey” from the country “Turkey.”

Because of these considerations, we prefer to preserve the original text as much as possible. In fact, we structure our code to allow us to access and read the text underlying each data point at any time. This practice has been particularly useful when interpreting over-time variation in our constructed measures, as it enables us to review any fluctuations in text-based measures through a targeted reading of the underlying text.

Cheap Talk

The second challenge relates to the potential for strategic behavior in management communications. Although the structure of earnings calls imposes institutional constraints on cheap talk, as previously discussed, researchers must still account for any potential residual strategic behavior in their empirical estimations. For example, when firms face declining operational performance, executives may expand their discussion to include external factors like climate change or political uncertainty in an effort to shift the focus of financial analysts away from company-specific issues. Researchers can address this through various methods. One approach involves examining whether estimation effects of text-based measures remain robust when controlling for earnings surprises and pre-call stock returns (for applications, see Hassan et al. 2019; Sautner et al. 2023). Other approaches aim to detect potential strategic communication through linguistic patterns (Dikolli et al. 2020) or vocal traits during the call (Hobson, Mayew, and Venkatachalam 2012), each offering distinct signals of possible misrepresentation.

Measurement Error

The third challenge concerns how accurately text-derived measures reflect the economic reality of a given firm. Like any measurement method (including those using conventional data sources), our text-derived measures of exposure, risk, and sentiment are subject to measurement error—random deviations from the true values. Generally, we can expect measurement error to be most pronounced at the granular (firm-quarter) level and to reduce considerably when aggregating across multiple observations, such as by averaging over multiple time periods for a given firm, all firms in a sector, or all firms in a country.

Several methods have been developed to quantify the remaining measurement error directly. For instance, Hassan et al. (2019) construct two distinct measures of political risk: one based on earnings-call transcripts and another derived from text in regulatory filings—specifically, the Form 10-K that publicly traded firms must file annually with the Securities and Exchange Commission.⁸ One can then serve as an instrument for the other in an instrumental variables regression to gauge the degree of random noise in each. Under the assumption that measurement errors are uncorrelated between the earnings-call-based and the Form 10K-based measures, we find that approximately half of the firm-quarter variation in political risk measures results from measurement error, indicating a signal-to-noise ratio of about 1:1. This level of precision is comparable to the one reported by Bloom et al. (2018) when using accounting data to measure total factor productivity at the firm-quarter level.⁹ While the extent of measurement error will always depend on the specific application, we can conclude that text-based measurements are not inherently noisier than those based on conventional data sources.

Other Sources of Text: Patents and Job Postings

In addition to corporate earnings call transcripts and 10-K filings, economists increasingly draw on other sources of text, including the use of patent documents and online job postings to study innovation and technological change.¹⁰ Patents have long been used to measure innovation—for example, by counting filings or citations (Lerner and Seru 2022; Hall, Jaffe, and Trajtenberg 2005). Recently, researchers have shifted toward analyzing the content of the patent text itself, thereby measuring how patents differ or how they incorporate new technical vocabulary. Job postings, by contrast, can help to capture the labor market dimension of innovation and technology adoption. As job advertisements have steadily moved online, data vendors have accumulated large databases of postings containing rich information on skills demanded and job locations (Hershbein and Kahn 2018). By analyzing the wording of these postings, economists can study

⁸ Earnings calls differ from regulatory filings, like the Form 10-K filing (Handley and Li 2020), in a number of ways. While the latter provide standardized quantitative information under strict accounting rules, earnings calls create an interactive forum where market participants can probe management's understanding and interpretation of firm performance. The question-and-answer format enables analysts to press management for clarification on vague statements, challenge overly optimistic projections, and request specific details about business conditions—imposing reputational costs for evasive or inconsistent responses. Under the protection of “safe harbor” provisions, management can discuss forward-looking information more extensively than in mandated filings, addressing emerging market conditions, anticipated economic shocks, and strategic responses not yet reflected in accounting metrics.

⁹ Sautner et al. (2023) find a lower level of measurement error in their text-based measures of climate change risk, possibly due to the more standardized vocabulary used in climate-related discussions.

¹⁰ We source full text of online job postings from Burning Glass, which aggregates them from online job boards, employer websites, and other sources. Patent text is obtained from the United States Patent and Trademark Office (USPTO).

which jobs are involved with using and developing which technologies, and where the jobs are located.

Broadly speaking, there are two complementary ways to use text analysis in these domains. One approach classifies the content of patents themselves, grouping them by linguistic similarity or by alignment with specific occupations or tasks. Another approach identifies specific technological concepts from within patent text, and then tracks their subsequent diffusion across other textual sources.

Classifying Patent Text

In the first category, for instance, Kelly et al. (2021) use a term frequency–inverse document frequency based measure of textual similarity between patents to identify “breakthrough” patents by comparing each patent’s language to both prior and subsequent patents. Patents featuring low backward similarity (previously uncommon language) but high forward similarity (strong influence on subsequent patents) are labeled breakthroughs. Kalyani (2023) instead focuses on the number of new technical phrases within the patent’s text to distinguish more creative patents (substantial new language) from derivative ones. He finds that creative patents—those introducing many new technical concepts—are strongly linked to firm- and sector-level growth in total factor productivity, as well as to abnormally positive stock returns. In contrast, derivative patents, which lack new technical language, show no such associations.¹¹

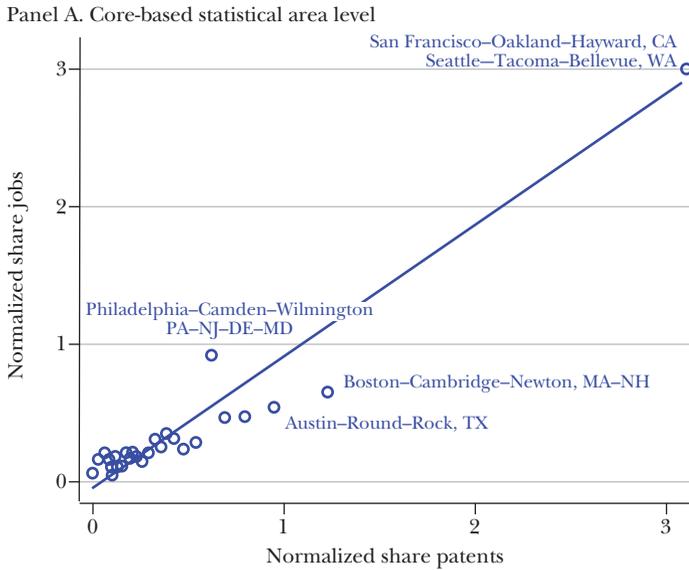
Autor et al. (2024) and Kogan et al. (2021) use embeddings-based textual similarity between patents and descriptions of tasks from the Dictionary of Occupational Titles to classify inventions as “labor-saving” or “labor-augmenting.” Inventions with higher overlap with routine tasks are classified as labor-saving, and those with higher overlap with nonroutine tasks as labor-augmenting. They then construct an occupation-by-year “exposure” measure to examine the wage and employment effects of these inventions.

Identifying Novel Technologies and Tracking Them across Multiple Corpora

Rather than classifying patents, the second approach uses patent text to identify a list of specific new technologies. In Kalyani et al. (2025a), we first analyze the full text of US patents from 1976 to 2014 to identify phrases that appear in multiple patents but did not exist before 1970, thus pinpointing new terminology specific to influential innovations over the past decades. Next, to map these phrases to specific new technologies, we search for these terms on Wikipedia, leveraging the standardized structure of Wikipedia technology pages to identify how a new phrase maps to a specific technology. This procedure leads us to identify a set of phrases associated with each of 1,286 unique new technologies developed since 1976. Examples include Computer Program, Monoclonal Antibody, Molecular Cloning, and Mobile Device, each linked to a set of new phrases associated with these new technologies.

¹¹ Mann and Püttmann (2023) classify patents as either automation-related or nonautomation, helping measure the extent to which industries face labor-replacing technologies.

Figure 5

Machine Learning/AI Jobs and Patents*(continued)*

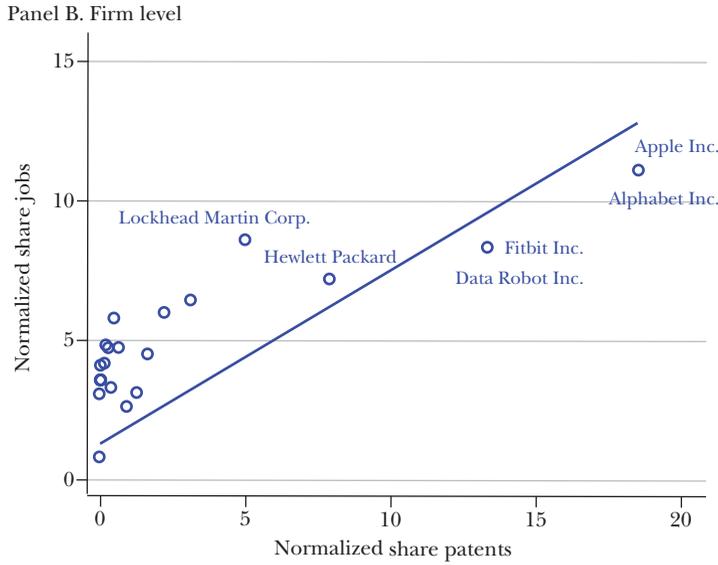
With this list of technologies in hand, we then search for them in patents and job postings. By examining inventor addresses of patents with early mentions of the technology, we determine the locations where each technology was developed and, by tracking patent application years, identify when these technologies saw significant increases in patent citations (their “emergence year”). Finally, we cross-reference the technology phrases with the full text of online job postings, identifying millions of job listings from 2010 to 2019 that mention the new technologies.

These detailed data enable us to track the number, location, and skill requirements of job postings linked to each new technology. From these data, we observe that “pioneer locations”—areas where a technology was invented—enjoy large and persistent job growth in roles using or producing that technology. This advantage for pioneering locations is especially strong in high-skill jobs and persists for decades.

Figure 5 illustrates these relations for one of the 1,286 new technologies: artificial intelligence. Panel A plots a given region’s share of patents mentioning AI prior to 2014 (AI emerged in 2014 as a new technology in patents) against its share of jobs mentioning AI post-2014. In each case, we normalize by the size of the local labor market so that the “normalized share” of jobs and patents indicates the local over- or underrepresentation of AI patents and jobs in a given area. We observe a strong association between AI’s early development and subsequent job creation related to the technology.

Figure 5 (continued)

Machine Learning/AI Jobs and Patents



Source: Hassan et. al. (2024b).

Note: Panel A shows a binned scatter plot depicting the relation between the normalized share of machine learning/AI-related job postings and patents across CBSAs (core-based statistical areas, defined by the Office of Management and Budget). The x-axis represents the normalized share of patents filed in a CBSA between 1976 and 2014 that mention machine learning/AI keywords, while the y-axis represents the normalized share of job postings mentioning these keywords in the same CBSA from 2015 to 2019. Specifically, the normalized share of jobs for each CBSA is calculated as:

$$Normalized\ share\ jobs_{c,t} = \frac{Share\ of\ machine\ learning\ or\ AI\ related\ job\ postings_{c,t}}{Share\ of\ all\ job\ postings_{c,t}}$$

This represents the proportion of machine learning/AI-related job postings within a CBSA relative to the national total, adjusted for the CBSA’s overall share of job postings in the United States during 2015–2019. Similarly, the normalized share of patents for each CBSA is calculated as the proportion of patents filed between 1976 and 2014 that mention machine learning/AI keywords, relative to the total number of patents filed in that CBSA during the same period. Panel B conducts the same analysis, but at the firm level instead of the level of the core-based statistical area.

Places like San Francisco, Boston, Philadelphia, and Austin pioneered artificial intelligence patents, and later hosted the lion’s share of AI-related jobs. A similar pattern holds true for the remaining 1,285 other technologies, where the vast majority of cases show a strong link between early patenting in the technology and subsequent job growth in using and producing that technology.

Panel B illustrates the same relationship at the firm level. Companies like Apple and Alphabet, which pioneered early development of the technology, show high levels of artificial intelligence patenting before 2014 and subsequent job creation. However, the slope of the regression line is now significantly flatter, indicating

positive spillovers from these companies to other local companies using and producing the new technology. This finding suggests that the relationship is steeper at the region level rather than at that of the firm.

This combined textual source approach thus offers valuable insight into how technology is adopted and spreads: Hassan, Kalyani, and Restrepo (2025) use these data to show a distinct boom-bust pattern in the number of technologies emerging over time and large regional discrepancies in the age of the median technology used. For example, the most frequently mentioned technology in job postings for positions in New York is about 34 years old, whereas the most frequently mentioned technologies in rural areas tend to be around 52 years old. These patterns have important implications for job opportunities and wage inequality across regions and over time.

Concluding Remarks

A text-based approach to economic analysis offers flexibility in tackling a variety of new and previously unexplored research questions. Our discussion has focused on corporate texts—such as earnings-call transcripts, but also patents and job postings—that are increasingly used in economic research. These sources offer valuable insights to researchers and policymakers on both historical and current events. Perhaps the most appealing feature of earnings calls is that they systematically capture the opinions, preferences, and expectations of key corporate decision-makers. They are available not only for a large sample of US publicly traded companies but also for numerous international firms, which makes earnings-call analysis a useful resource for gaining a global perspective on corporate strategies, risks, and expectations. Likewise, patents and job postings provide detailed and systematic insights into emerging technologies and their spread.

A wide array of other corporate text sources are available, which have so far received less attention, including but not limited to regulatory filings (such as annual or quarterly financial reports), investor presentations, press releases, analyst reports, corporate blogs, and contracts. Systematic research efforts will be needed to bring these additional corpora online for economic research and to determine how best to parse, structure, and interpret the information they contain.

The text-as-data techniques we discussed in this paper are increasingly informing policy work at central banks and international organizations. Researchers now use earnings calls to quantify trade-policy uncertainty (Correa et al. 2023), firms' cross-country inflation expectations (Albrizio, Dizioli, and Simon 2023), and supply chain "friendshoring" or "reshoring" (Alfaro and Chor 2023). Central bank economists build realtime dashboards of supply/demand imbalances, corporate sentiment, and risk from the same data (Gosselin and Taskin 2023; Andersson, Neves, and Nunes 2023), and exploit patent text to gauge how monetary shocks shape innovation (Ma and Zimmermann 2023). Several recent issues of the IMF's *World Economic Outlook* and similar publications include chapters that incorporate earnings-call

data to analyze supply chains, inflation expectations, and other macroeconomic risks (Faella et al. 2022; International Monetary Fund 2022; Ahn et al. 2023; Albrizio et al. 2023).

These applications are generally straightforward and easily interpretable, demonstrating that meaningful results in this field do not always require intricate linguistic and econometric models. Given the wealth of information available, through various widely accessible text sources, we anticipate that text analysis will become a standard part of the future tool kit for economic research.

While our analysis highlights the power of textual analysis in extracting meaningful signals from unstructured data, a crucial question is how to integrate these signals into economic theories, whether through structural modeling, causal inference, or predictive frameworks. Key open questions remain, particularly in refining textual measures to better align with economic concepts (such as the measurement of “risk”), and ensuring robustness in cross-context applications. Addressing these challenges will be essential for advancing the field.

■ *This article reflects the views of the authors and not necessarily those of the Federal Reserve Bank of St. Louis or the Federal Reserve System. We thank Jiarui Wang for providing excellent research assistance. We appreciate helpful suggestions from Erik Hurst, Jonathan Parker, Nina Pavcnik, Timothy Taylor, and Heidi Williams. Tahoun’s research is supported by grants from the Institute for New Economic Thinking, the Wheeler Institute for Business and Development, and the London Business School. Van Lent gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft Project ID 403041268-TRR 266. Hassan, Kalyani, Schwedeler, and Tahoun are cofounders of NL Analytics Inc., a startup that provides tools for earnings-call analysis to researchers, central banks, and international organizations.*

References

- Ahn, JaeBin, Benjamin Carton, Ashique Habib, Davide Malacrino, Dirk Muir, and Andrea Presbitero. 2023. “Goeconomic Fragmentation and Foreign Direct Investment.” In *World Economic Outlook*, 91–114. International Monetary Fund.
- Albrizio, Silvia, John Bluedorn, Allan Dizioli, Christopher Koch, and Philippe Wingender. 2023. “Managing Expectations: Inflation and Monetary Policy.” In *World Economic Outlook*, 49–69. International Monetary Fund.
- Albrizio, Silvia, Allan Gloe Dizioli, and Pedro Vitale Simon. 2023. “Mining the Gap: Extracting Firms’ Inflation Expectations from Earnings Calls.” International Monetary Fund Working Paper 202.
- Alfaro, Laura, and Davin Chor. 2023. “Global Supply Chains: The Looming ‘Great Reallocation.’” NBER Working Paper 31661.
- Andersson, Malin, Pedro Neves, and Carolina Nunes. 2023. “Earnings Calls: New Evidence on Corporate Profits, Investment, and Financing Conditions.” *European Central Bank Economic Bulletin* 4: 56–62.

- Autor, David, Caroline Chin, Anna Salomons, and Brayn Seegmiller.** 2024. “New Frontiers: The Origins and Content of New Work, 1940–2018.” *Quarterly Journal of Economics* 139 (3): 1399–1465.
- Bae, Jihun, Chung Yu Hung, and Laurence van Lent.** 2023. “Mobilizing Text as Data.” *European Accounting Review* 32 (5): 1085–1106.
- Baker, Scott R., Nicholas Bloom, and Steven J. Davis.** 2016. “Measuring Economic Policy Uncertainty.” *Quarterly Journal of Economics* 131 (4): 1593–1636.
- Benati, Luca, and Paolo Surico.** 2009. “VAR Analysis and the Great Moderation.” *American Economic Review* 99 (4): 1636–52.
- Berger, David, Ian Dew-Becker, and Stefano Giglio.** 2020. “Uncertainty Shocks as Second-Moment News Shocks.” *Review of Economic Studies* 87 (1): 40–76.
- Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J. Terry.** 2018. “Really Uncertain Business Cycles.” *Econometrica* 86 (3): 1031–65.
- Bloom, Nicholas, Tarek Alexander Hassan, Aakash Kalyani, Josh Lerner, and Ahmed Tahoun.** 2021a. “The Diffusion of Disruptive Technologies.” NBER Working Paper 28999.
- Bloom, Nicholas, Tarek Alexander Hassan, Aakash Kalyani, Josh Lerner, and Ahmed Tahoun.** 2021b. *Data and Code for: “The Diffusion of Disruptive Technologies.”* <https://www.techdiffusion.net>.
- Bochkay, Khrystyna, Jeffrey Hales, and Sudheer Chava.** 2020. “Hyperbole or Reality? Investor Response to Extreme Language in Earnings Conference Calls.” *Accounting Review* 95 (2): 31–60.
- Cao, Sean, Wei Jiang, Baozhong Yang, and Alan L. Zhang.** 2023. “How to Talk When Machines Are Listening: Corporate Disclosure in the Age of AI.” *Review of Financial Studies* 36 (9): 3603–42.
- Correa, Ricardo, Julian di Giovanni, Linda S. Goldberg, and Camelia Minoiu.** 2023. “Trade Uncertainty and U.S. Bank Lending.” Federal Reserve Bank of New York Staff Reports 1076.
- Crawford, Vincent P., and Joel Sobel.** 1982. “Strategic Information Transmission.” *Econometrica* 50 (6): 1431–51.
- Dikolli, Shane S., Thomas Keusch, William J. Mayew, and Thomas D. Steffen.** 2020. “CEO Behavioral Integrity, Auditor Responses, and Firm Outcomes.” *Accounting Review* 95 (2): 61–88.
- Dixit, Avinash K., and Robert S. Pindyck.** 1994. *Investment under Uncertainty*. Princeton University Press.
- Faella, F., B. Javorcik, L. Kitzmüller, V. Langella, and H. Schweiger.** 2022. “Global Supply Chains in Turbulence.” In *European Bank for Reconstruction and Development Transition Report*, 59–61. European Bank for Reconstruction and Development.
- Flynn, Joel P., and Karthik Sastry.** 2024. “Attention Cycles.” NBER Working Paper 32553.
- Frankel, Richard, Marilyn Johnson, and Douglas J. Skinner.** 1999. “An Empirical Examination of Conference Calls as a Voluntary Disclosure Medium.” *Journal of Accounting Research* 37 (1): 133–50.
- Gali, Jordi, and Luca Gambetti.** 2009. “On the Sources of the Great Moderation.” *American Economic Journal: Macroeconomics* 1 (1): 26–57.
- Gallemore, John, Stephan Hollander, Martin Jacob, and Xiang Zheng.** 2025. “Tax Policy Expectations and Investment.” *Journal of Accounting Research* 63 (1): 363–412.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy.** 2019. “Text as Data.” *Journal of Economic Literature* 57 (3): 535–74.
- Gosselin, Marc-André, and Temel Taskin.** 2023. “What Can Earnings Calls Tell Us about the Output Gap and Inflation in Canada? Bank of Canada Staff Discussion Paper 2023-13.
- Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg.** 2005. “Market Value and Patent Citations.” *RAND Journal of Economics* 36 (1): 16–38.
- Handley, Kyle, and J. Frank Li.** 2020. “Measuring the Effects of Firm Uncertainty on Economic Activity: New Evidence from One Million Documents.” NBER Working Paper 27896.
- Hassan, Tarek A., Stephan Hollander, Aakash Kalyani, Laurence van Lent, Markus Schwedeler, and Ahmed Tahoun.** 2025a. *Data and Code for: “Text as Data in Economic Analysis.”* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. <https://doi.org/10.3886/E228102V1>.
- Hassan, Tarek A., Stephan Hollander, Aakash Kalyani, Laurence van Lent, Markus Schwedeler, and Ahmed Tahoun.** 2025b. *Tutorial Package for: “Text as Data in Economic Analysis.”* <https://doi.org/10.34894/KNDZ9T>, Dataverse NL, VI.
- Hassan, Tarek A., Stephan Hollander, Laurence van Lent, Markus Schwedeler, and Ahmed Tahoun.** 2023. “Firm-Level Exposure to Epidemic Diseases: Covid-19, SARS, and H1N1.” *Review of Financial Studies* 36 (12): 4919–64.
- Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun.** 2019. “Firm-Level Political Risk: Measurement and Effects.” *Quarterly Journal of Economics* 134 (4): 2135–202.

- Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. 2024a. "The Global Impact of Brexit Uncertainty." *Journal of Finance* 79 (1): 413–58.
- Hassan, Tarek A., Jesse Schreger, Markus Schwedeler, and Ahmed Tahoun. 2024b. "Sources and Transmission of Country Risk." *Review of Economic Studies* 91 (4): 2307–46.
- Hassan, Tarek A., Jesse Schreger, Markus Schwedeler, and Ahmed Tahoun. 2024c. *Replication Package for: "Sources and Transmission of Country Risk."* <https://doi.org/10.5281/zenodo.7783429>.
- Hassan, Tarek A., Aakash Kalyani, and Pascual Restrepo. 2025. "New Technologies and the College Premium." Working Paper.
- Hennessy, Christopher A., and Ilya A. Strebulaev. 2020. "Beyond Random Assignment: Credible Inference and Extrapolation in Dynamic Economies." *Journal of Finance* 75 (2): 825–66.
- Hershbein, Brad, and Lisa B. Kahn. 2018. "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings." *American Economic Review* 108 (7): 1737–72.
- Hobson, Jessen L., William J. Mayew, and Mohan Venkatachalam. 2012. "Analyzing Speech to Detect Financial Misreporting." *Journal of Accounting Research* 50 (2): 349–92.
- Hollander, Stephan, Maarten Pronk, and Erik Roelofsen. 2010. "Does Silence Speak? An Empirical Analysis of Disclosure Choices during Conference Calls." *Journal of Accounting Research* 48 (3): 531–63.
- Huang, Allen H., Reuven Lehavy, Amy Y. Zang, and Rong Zheng. 2018. "Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach." *Management Science* 64 (6): 2833–55.
- International Monetary Fund. 2022. "Global Prospects and Policies." In *World Economic Outlook: Countering the Cost-of-Living Crisis*, 1–49. International Monetary Fund.
- Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang. 2024. "ChatGPT and Corporate Policies." NBER Working Paper 32161.
- Kalyani, Aakash. 2023. "The Creativity Decline: Evidence from US Patents." Preprint, SSRN. <http://dx.doi.org/10.2139/ssrn.4318158>.
- Kalyani, Aakash, Nicholas Bloom, Marcela Carvalho, Tarek Alexander Hassan, Josh Lerner, and Ahmed Tahoun. 2025a. "The Diffusion of New Technologies." *Quarterly Journal of Economics* 140 (2): 1299–365.
- Kalyani, Aakash, Nicholas Bloom, Marcela Carvalho, Tarek Alexander Hassan, Josh Lerner, and Ahmed Tahoun. 2025b. *Replication Package for: "The Diffusion of New Technologies."* <https://doi.org/10.7910/DVN/AYRHYN>.
- Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2021. "Measuring Technological Innovation over the Long Run." *American Economic Review: Insights* 3 (3): 303–20.
- Kim, Alex G., Maximilian Muhn, and Valeri V. Nikolaev. 2023. "Bloated Disclosures: Can ChatGPT Help Investors Process Information?" Chicago Booth Research Paper 23-07.
- Kogan, Leonid, Dimitris Papanikolaou, Lawrence D. W. Schmidt, and Bryan Seegmiller. 2021. "Technology-Skill Complementarity and Labor Displacement: Evidence from Linking Two Centuries of Patents with Occupations." NBER Working Paper 29552.
- Lee, Joshua. 2016. "Can Investors Detect Managers' Lack of Spontaneity? Adherence to Predetermined Scripts during Earnings Conference Calls." *Accounting Review* 91 (1): 229–50.
- Lerner, Josh, and Amit Seru. 2022. "The Use and Misuse of Patent Data: Issues for Finance and Beyond." *Review of Financial Studies* 35 (6): 2667–704.
- Loughran, Tim, and Bill McDonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance* 66 (1): 35–65.
- Ma, Yueran, and Kaspar Zimmermann. 2023. "Monetary Policy and Innovation." NBER Working Paper 31698.
- Mann, Katja, and Lukas Püttmann. 2023. "Benign Effects of Automation: New Evidence from Patent Texts." *Review of Economics and Statistics* 105 (3): 562–79.
- Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press.
- Matsumoto, Dawn, Maarten Pronk, and Erik Roelofsen. 2011. "What Makes Conference Calls Useful? The Information Content of Managers' Presentations and Analysts' Discussion Sessions." *Accounting Review* 86 (4): 1383–414.
- Mayew, William J., Mani Sethuraman, and Mohan Venkatachalam. 2020. "Individual Analysts' Stock Recommendations, Earnings Forecasts, and the Informativeness of Conference Call Question and Answer Sessions." *Accounting Review* 95 (6): 311–37.
- Min, Sarah. 2023. "Chegg Shares Drop More than 40 Percent after Company Says ChatGPT Is Killing Its

- Business." *CNBC*, May 2. <https://www.cnn.com/2023/05/02/chegg-drops-more-than-40percent-after-saying-chatgpt-is-killing-its-business.html>.
- Pindyck, Robert S.** 1991. "Irreversibility, Uncertainty, and Investment." *Journal of Economic Literature* 29 (3): 1110–48.
- Rogers, Jonathan L., Andrew Van Buskirk, and Sarah L. C. Zechman.** 2011. "Disclosure Tone and Shareholder Litigation." *Accounting Review* 86 (6): 2155–83.
- Romer, Christian D., and David H. Romer.** 1989. "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." In *NBER Macroeconomics Annual*, Vol. 4, 121–70. MIT Press.
- Romer, Christina D., and David H. Romer.** 2023. "Presidential Address: Does Monetary Policy Matter? The Narrative Approach after 35 Years." *American Economic Review* 113 (6): 1395–423.
- Sautner, Zacharias, Laurence Van Lent, Grigory Vilkov, and Ruishen Zhang.** 2023. "Firm-Level Climate Change Exposure." *Journal of Finance* 78 (3): 1449–98.
- Sarkar, Suproteem K., and Keyon Vafa.** 2024. "Lookahead Bias in Pretrained Language Models." <http://dx.doi.org/10.2139/ssrn.4754678>.
- Song, Min, and Yi-fing Brook Wu.** 2008. *Handbook of Research on Text and Web Mining Technologies*. Information Science Reference, IGI Global.
- Suijs, Jeroen.** 2007. "Voluntary Disclosure of Information When Firms Are Uncertain of Investor Response." *Journal of Accounting and Economics* 43 (2–3): 391–410.
- Taskin, Temel, and Franz Ulrich Ruch.** 2023. "Global Demand and Supply Sentiment: Evidence from Earnings Calls." Bank of Canada Staff Working Paper 2023-37.
- Vafa, Keyon, Justin Y. Chen, Ashesh Rambachan, Jon Kleinberg, and Sendhil Mullainathan.** 2024. "Evaluating the World Model Implicit in a Generative Model." Preprint, arXiv. <https://doi.org/10.48550/arXiv.2406.03689>.