Noisy News: Theory and Evidence from AI*

Christoph E. Boehm
University of Texas at Austin
and NBER

Zhen Huo
Yale University

Andrei A. Levchenko
University of Michigan
NBER and CEPR

Nitya Pandalai-Nayar
University of Texas at Austin
and NBER

May 2026

Abstract

Using a new dataset of firm mentions of Artificial Intelligence (AI) in earnings calls, we show substantial heterogeneity in firm beliefs about AI. We find that a large share of firms do not mention AI as late as 2025, and the intensity with which firms mention AI varies substantially across firms. Among firms that mention AI, there is wide variation in beliefs about the impact of AI on their own efficiency and performance. Following the release of ChatGPT, there is an increase in mentions of AI across all sectors, and an increase in the similarity in how firms talk about AI. We develop a dynamic model of noisy news about future technology improvements. Initially, firms have heterogeneous private signals about the productivity of a future technology. At some point in time they receive a common public signal about the technology, that in the data we associate with the ChatGPT release. We calibrate the model to (i) the increased similarity in AI mentions after the ChatGPT announcement, and (ii) the correlation of firm AI mentions with concrete AI-related actions, measured by AI job posting intensities. The model implies that with only noisy private information, investment in the new technology is initially muted, but sharply increases with arrival of the public signal. The increase in investment triggered by the public signal is larger when the new technology is believed to be arriving further in the future, as the public signal plays a bigger role in synchronizing higher-order beliefs with a later arrival date. These patterns are consistent with trends in AI job posting shares and narratives about AI investment.

Keywords: Information Frictions, News shocks, Noise shocks, AI, Investment
JEL Codes: C72, D82, E22, E32

*We thank Wouter den Haan and participants at the 57th Konstanz Seminar for helpful discussions, and Joost Haddinga for outstanding research assistance. Email: chris.e.boehm@utexas.edu, zhen.huo@yale.edu, alev@umich.edu and npnayar@utexas.edu.

1. Introduction

The world is currently experiencing the arrival of a new general purpose technology, AI. There is increasing investment by firms in AI-related infrastructure, and widespread belief that the technology will be transformative for society. However, transformative technologies also arrive slowly. There is often a transition period during which there is awareness of a new technology, but it is not yet sufficiently mature to fully realize its productivity gains.[1] Relatedly, a common thread in the public discourse is that firms and workers are not certain of how transformative the technology will be, and when it will impact them. Indeed, today exhibits some parallels with the preceding episode of a new transformative technology – the IT revolution and the economic and asset price boom of the late 1990s. Those developments spurred an active research agenda that attempted to understand how the economy responds to an anticipated arrival of a future productivity shock (e.g. Beaudry and Portier, 2006), and the inherent uncertainty about the size of that shock (e.g. Lorenzoni, 2009).

This paper thinks of AI as a news shock. We use both measurement and theory to understand how the arrival of news about this transformative technology impacts the economy. As a laboratory to understand the dynamic impact of future technologies, the arrival of AI has at least two advantages relative to the 1990s. The first is the much greater availability of text-as-data and methods of processing it to elicit agents' information sets and beliefs about this technology. The second is the presence of a clearly defined watershed event: the release of ChatGPT in November 2022.

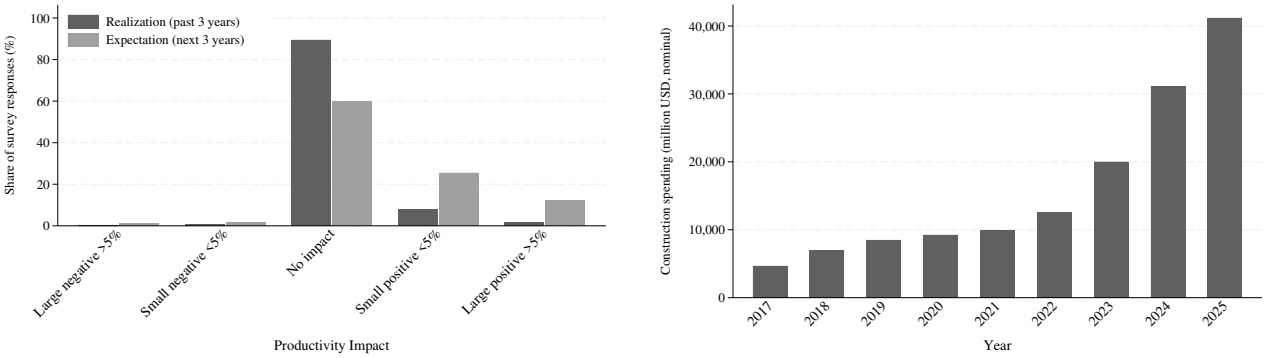
There is, however, a key difference of the arrival of AI from the standard news shock model. While agents in the standard news shock model all receive the same (potentially noisy) signal, beliefs about the productivity impact of AI vary widely. This disagreement exists among economists as well as firms. Panel (a) of Figure 1 illustrates firms' views of the productivity impact of AI over the past three years and the expected productivity impact in the next three years. The interviews took place in late 2025 or early 2026. Almost 90 percent of respondents report no productivity impact of AI over the preceding three years, consistent with our interpretation of AI as a news shock. For the next three years about 60 percent of firms expect no effect on productivity, a little over 20 percent expect small positive gains of less than 5 percent, and about 10 percent of firms expect gains greater than 5 percent. Clearly, there is substantial disagreement among firms about the likely impact of AI on productivity – at least over the coming three years.

At the same time, however, construction spending on data centers has shot up (see Panel (b) of Figure 1). These expenditures approximately doubled from 2021 to 2023, and then again from 2023 to 2025. The combination of these two facts is surprising. How is it possible that the large majority of firms expect no meaningful effect on productivity in the near future, and yet we see a quadrupling of data center construction expenditures, among other investments into AI?

We propose an explanation for this type of behavior based on a business cycle model with belief

[1] Yotzov et al. (2026) use firm surveys from five countries to document that current productivity impacts are modest on average, and expected future productivity impacts are larger, with substantial heterogeneity across firms.

Figure 1: Firm Responses About Productivity Impact of AI and Aggregate AI Investment



(a) Productivity Impact

(b) Data Center Construction Expenditure (US)

Notes: Panel (a): Data from Yotzov et al. (2026). Survey responses among 5,878 firms across the UK, US, Germany, and Australia in late 2025 to early 2026 to the questions: "How has the adoption of artificial intelligence technologies affected volume of sales per employee in your business over he past three years? And how do you expect this to affect your volume of sales per employee over the next 3 years?" Panel (b): Data from U.S. Census Bureau, Construction Spending. Annual private construction spending (value put in place) on the item *Data Centers* between 2017-2025.

dispersion.

Our first contribution is to assemble a new panel dataset of firms' AI-related actions and beliefs. Starting with transcripts of global earnings calls available in the English language, we identify AI-related sentences using Natural Language Processing (NLP) and machine learning methods. We subsequently use Large Language Models (LLMs) and other techniques to extract features from these sentences that are informative about firms' actual or intended AI-related actions, their beliefs about how AI will impact their operational efficiency and financial performance, as well as their general sentiment towards AI. We then aggregate this sentence-level information to a new dataset at the firm-quarter level from 2017q1 to 2025q4.

The data reveal that AI mentions in earnings calls were infrequent up to the release of ChatGPT in 2022q4, but surged afterwards and now stand at over one percent of all sentences. Talk about AI reflects generally positive sentiment and positive or neutral views on how AI will affect operational efficiency and financial performance. The majority of AI mentions in earnings calls suggest that firms engage in AI-related actions, such as investments or AI-related hiring.

Heterogeneity across firms is large. Most firms in our sample do not mention AI in a given quarter, although the share firms mentioning AI has increased from about 10 percent before ChatGPT to about 30 percent at the end of 2025. Conditional on mentioning AI, most firms dedicate only small shares of their discussion to AI, while others – post ChatGPT – dedicate more than 10 percent of sentences to AI. Firms' views of AI regarding performance and efficiency range from mildly negative to very positive. While the cross-sectional dispersion of all measures of beliefs, sentiments, and actions is high, similarity measures of AI-related talk rise post ChatGPT, suggesting that firms' views of AI converge somewhat.

Actions related to AI and beliefs about AI are strongly positively correlated. We measure AI-

related actions by actual or intended actions from earnings call transcripts, as well as by the share of AI-related vacancies, which we construct from the Lightcast job vacancy database. The correlation is also robust to using different measures of firms' beliefs, including beliefs about how AI impacts operational efficiency, how it impacts financial performance, and firms' general sentiment towards AI. In addition, the correlation between AI-related actions and beliefs increases substantially after ChatGPT.

Our second contribution is to develop a dynamic model of a future arrival of a productivity shock featuring both uncertainty about the size of this shock and dispersed information. Thus, our framework unifies the explicitly dynamic and forward-looking models of technology news shocks (Beaudry and Portier, 2006) with the predominantly static models of noise/sentiment shocks (Angeletos and La'O, 2013). Firms know that in some future period T, productivity of their capital will rise. However, they are not sure about the size of the productivity change. This setup captures the combination of awareness of AI with uncertainty about the magnitude of AI's ultimate impact. Because capital is accumulated slowly, firms make forward-looking investment decisions under incomplete information about the size of the future productivity shock. The economic environment features strategic complementarity: a firm wants to have a high level of capital in a future period if other firms also have a lot of capital, and vice versa. The combination of dispersed information and strategic complementarity introduces a role for higher-order beliefs: to make decisions, each firm needs to forecast the beliefs of others, the beliefs of others about the beliefs of others, and so on. Unlike earlier contributions such as Angeletos and La'O (2013), (part of) the complementarity is forward-looking: firms make investment decisions based on their forecast of the future path of other firms' capital.

The information structure is as follows: each firm receives a private signal of the productivity improvement at some future time T. It continues making investment decisions based on this private signal until, in a particular period, all firms also receive a noisy public signal regarding the value of that future productivity. The arrival of this noisy public signal captures the first release of ChatGPT in November 2022. The noisy public signal improves the precision of each firms' belief about the magnitude of the future productivity shock. At the same time, by its public nature, it also changes higher-order beliefs.

In spite of rich dynamics, the model admits an analytical solution to first order. The key endogenous variable is the path of aggregate capital from the present to T, when the productivity shock is revealed. This path can be expressed as a function of matrices of structural parameters and an infinite sum of higher-order expectations of the future productivity shock. The model implies that under dispersed information, reliance on higher-order expectations is greater the further in the future is the productivity shock. This is because when the shock is relatively far away, it is relatively more important to predict others' beliefs – higher order expectations – than to predict actual productivity.

When the public signal arrives, it changes higher-order expectations by more than the first-order expectations. Because higher-order expectations matter more in earlier periods, the arrival of the

public signal changes behavior closer to the period of the public signal by more than the later behavior. Thus, the public signal serves to coordinate actions, and by relatively more in the short run. Its arrival can precipitate a "frenzy" or a "bubble"-like event, when many firms simultaneously embark on investment based on the promise of a future productivity boom, even though there is considerable uncertainty about the size of the shock itself.

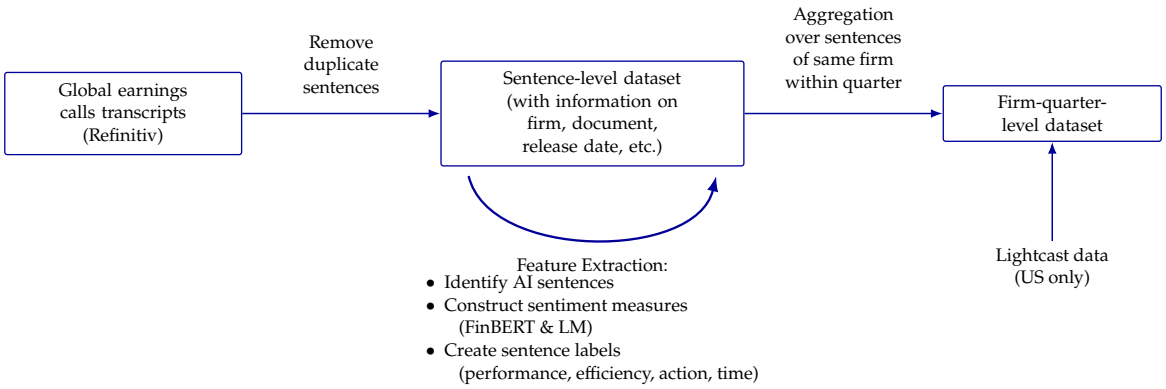
Related literature. Our project builds first and foremost on literature on the role of imperfect information and noise shocks in the business cycle (a very partial list includes Beaudry and Portier, 2006; Lorenzoni, 2009; Barsky and Sims, 2011; Blanchard, L'Huillier, and Lorenzoni, 2013; Angeletos and La'O, 2013; Nimark, 2014; Benhabib, Wang, and Wen, 2015; Huo and Takayama, 2015; Acharya, Benhabib, and Huo, 2021; Bybee et al., 2023; Hébert and La'O, 2023; Angeletos, Lorenzoni, and Pavan, 2022); and on network games with incomplete information (Bergemann, Heumann, and Morris, 2017; Atolia and Chahrour, 2020; Chahrour, Nimark, and Pitschner, 2021; Lian, 2021). Our theoretical contribution is to generalize the framework with strategic complementarities and dispersed information to a dynamic setting, in which forward-looking decisions depend on forecasts of the path of future beliefs of others. Our empirical contribution is to measure firms' AI-related beliefs over time with text data.

Previous work has documented that while the adoption of AI often increases individual worker productivity for specific tasks (e.g., Noy and Zhang, 2023; Brynjolfsson, Li, and Raymond, 2025), the effects on firm productivity are often absent or small (e.g., Babina et al., 2024; Yotzov et al., 2026; Baslandze et al., 2026). At the same time, firms often expect future productivity gains (Yotzov et al., 2026; Baslandze et al., 2026). These patterns in the data are consistent with our interpretation of the release of ChatGPT and the subsequent wave of AI-optimism as a news shock. Indeed, other research has likened this episode to news shocks, including Gazzani and Natoli (2024), Goertz, Gunn, and Lubik (2025), Faria-e Castro and Ozkan (2026).

Our paper also builds on the empirical research program that uses text as data (see, e.g. Gentzkow, Kelly, and Taddy, 2019; Hassan et al., 2025, for overviews), especially the branch that works with firms' earnings calls (e.g. Hassan et al., 2019, 2023, 2024; Sautner et al., 2023), and linking earnings calls to vacancy data to study the spread of new technologies (Kalyani et al., 2025). We apply these tools to measure beliefs about AI in particular.

The rest of the paper is organized as follows. Section 2 describes our data collection effort. Section 3 presents several motivating facts suggestive of dispersed information about AI among firms, and documents some reduced-form patterns in AI related sentiment and hiring before and after the ChatGPT announcement. firms. Section 4 sets up and solves a business cycle model with dispersed information about future productivity shocks. Section 5 calibrates and quantifies the model using the patterns in Section 3. Section 6 concludes. The appendices collect proofs of propositions and additional details on theory, data, and robustness.

Figure 2: Overview: From Earnings Calls to Datasets



Notes: The figure depicts schematically how we create the sentence-level and firm-quarter-level datasets from the earnings call transcripts. FinBERT (Araci, 2019; Huang, Wang, and Yang, 2023) is a large language model for sentiment classification trained on financial text, and LM stands for the Loughran-McDonald (LM) measure (Loughran and Mcdonald, 2011; Loughran and McDonald, 2020) to capture sentiment.

2. Data and Measurement

Our primary data source are transcripts of corporate earnings calls from Refinitiv between January 2001 and December 31, 2025. We work with all earnings calls that are available in English. The dataset includes 465,531 transcripts containing 124,031,096 unique sentences from 17,034 publicly listed firms around the world (of which 7,720 are US firms, see Appendix Table A1). All earnings calls are time-stamped.

We use these earnings call transcripts to create two datasets. In a first step, we create a dataset at the sentence level, with information on the firm, the document, the release date, the position of the sentence within the earnings call, the total number of sentences in the earnings call, the adjacent sentences, and so forth. Since firms occasionally have two earnings calls on the same day and repeat part of their scripted message, we delete duplicate sentences at this stage. We supplement the earnings call data with information on the firms' industries, which also comes from Refinitiv.

We then construct – at the sentence level – several measures related to firms' beliefs about AI and statements about activities related to AI. These measures aim to gauge, among other things, firms' sentiment towards AI, how firms believe AI will affect their operational efficiency and financial performance, and whether the firm engages or intends to engage in AI-related action such as AI-specific investments. We subsequently aggregate this dataset to the firm-quarter level. This dataset has 449,889 observations and is the second dataset we use in our analysis. Figure 2 provides an overview of this process. We next describe it in greater detail.

2.1 Sentence-level Data

AI-related sentences. After reading in the earnings call transcripts, we begin with classifying each sentence as AI-related or not AI-related. We use Natural Language Processing (NLP) and machine

learning for this classification. Specifically, we classify each sentence with supervised machine learning by fine-tuning the FinBERT large language model, an LLM trained on financial text (Araci, 2019; Huang, Wang, and Yang, 2023). In the first step, we manually label a training dataset of AI sentences and fine-tune FinBERT to develop a decision boundary to differentiate between obvious non-AI and potential AI sentences. As this leads to overfitting, we repeat this stage using a training sample of false positive and ambiguous sentences to achieve higher precision in the classification. Lastly, we feed the FinBERT probabilities and sentence properties into a logistic regression to discipline the decision boundary for all sentences where FinBERT was not highly confident (sentences that received a probability of 0.3-0.9 of being an AI sentence in the first stage). We provide further details in Appendix A. With this approach, we classify 207,241 sentences from 7,046 different firms in 37,114 distinct earnings calls as AI-related. This corresponds to a share of 0.162 percent of all sentences.

Sentiment. Next, we construct two sentiment measures for each AI-related sentence. Both are widely used metrics in the finance literature: the Loughran-McDonald (LM) measure (Loughran and Mcdonald, 2011; Loughran and McDonald, 2020) and a measure based on FinBERT. We map these measures into three numerical values: 1 if sentiment is positive, 0 if neutral, and -1 if negative. Additionally, the LM dictionary allows for classifying sentences as uncertain or not based on the presence of uncertainty-signaling words in a sentence.

Action, performance, and efficiency labels. To measure firms' beliefs about AI and actions related to AI, we use an LLM-based approach. We design detailed prompts to categorize each sentence based on its content related to action, performance, and efficiency. Action refers to sentences in which the company talks about undertaking activities concerned with AI, performance refers to immediate business and financial effects (e.g., profits, costs), whereas efficiency refers to operational effects (e.g., productivity). For each of these categories, we classify AI-related sentences as positive mentions (1), negative mentions (-1), and neutral mentions (0). Sentences that are not related to AI all receive a 0 label. More formally, for a generic variable x_{fit} of firm f, sentence i, at time t:

$$x_{fit} = \begin{cases} 1 & \text{if AI-related and positive} \\ 0 & \text{if AI-related and neutral or not AI related} \\ -1 & \text{if AI-related and negative} \end{cases}. \tag{2.1}$$

We write separate prompts for each category (action, performance, efficiency) in which we describe the different available labels, a definition of each label, and specify examples for each label. Further, we include specific rules for how the model should handle conflicting information (e.g., take the dominant signal if different directions of performance or efficiency are described) or confounding information (e.g., distinguish between performance and efficiency). More information on the prompts is available in Appendix A.

We then feed each sentence into the ChatGPT-4o-API twice to classify it according to these cate-

gories. First, we supply only the identified sentence, and second, we add the preceding and succeeding sentence to add context in order to capture implicit or hidden details in the AI sentence. We instruct ChatGPT to only consider context for its classification if it detects explicit or implicit references to the central sentence, a logical continuation or repetition of named entities or nouns, or pronoun references. Otherwise, the model is instructed to base the classification solely on the information in the central sentence. We find that for about 15% of sentences signals from context are used in the final label. For each category, we apply prompt engineering to select the most accurate version and specify instructions for few-shot classification. We also test our baseline prompts against alternative prompts and cross-check our classification against alternative open-source models. Further details, together with LLM-robustness checks are available in Appendix A and the accompanying technical note.[2]

Time labels. Sentences can also receive time labels. These labels refer to any explicit or implicit timing associated with AI in the sentence. Here, we group firms' AI-related sentences into past (more than one quarter ago), present (less than a quarter ago to less than a quarter ahead), short-term future (more than a quarter ahead but less than three years ahead), and long-term future (more than three years ahead). Formally,

$$\ell_{fit} = \begin{cases} \text{past} & \text{if AI-related and references past} \\ \text{present} & \text{if AI-related and references present} \\ \text{near future} & \text{if AI-related and references near-term future} \\ \text{distant future} & \text{if AI-related and references more distant future} \\ & \text{if not AI-related or does not reference time} \end{cases}. \tag{2.2}$$

For example, a company describing current actions or ongoing operations related to AI, these are classified as present. Only sentences that explicitly express future expectations and actions or discuss the future impact of AI are classified as short-term or long-term depending on the cited horizon. The label will be applied whenever the firm makes any mention of timing in an AI-related sentence.

Examples. We discuss several example sentences and their labels in Appendix A.1.6.

2.2 Aggregation to the firm-quarter level

In a next step we aggregate the sentence-level dataset to the firm-quarter level. Specifically, let Ω_{ft} be the set of sentences in the earnings call(s) of firm f at time t, and N_{ft} the number of such sentences. We aggregate the measure x_{fit} (sentiment, performance, efficiency, or action) as follows:

$$x_{ft} = \frac{1}{N_{ft}} \sum_{i \in \Omega_{ft}} x_{fit}. \tag{2.3}$$

[2] Available here.

The interpretation of this measure is that it is the net share of positive AI sentences in the number of total sentences. Accordingly, we will refer to this measure as firms' net action, net sentiment, net performance or net efficiency going forward. It is easy to see that this measure is bounded from above and below by (minus) the share of AI sentences in total sentences in a firm's earnings call. We will report these measures in percent. Note that both neutral AI sentences and non-AI-related sentences enter with a zero in the sum (see equation 2.1). Note also that this measure captures a "sentence-level extensive margin" in the sense that this measure will move if a firm uses more AI-sentences in the earnings call unless the average of x_{fit} is zero. Below we show that the average is positive for all variables we use (both sentiment measures, performance, efficiency, or action), and hence a greater number of AI-sentences will typically raise the firm-level measure x_{ft}. The firm-quarter level is the aggregation level at which we perform most of our analysis.

To elicit firms' expectations of future AI-related productivity, we also construct a measure that conditions on specific time labels. In particular, we construct a measure of expected future AI-related efficiency as

$$x_{ft}^{\text{future}} = \frac{1}{N_{ft}} \sum_{i \in \Omega_{ft}} x_{fit} \cdot 1\left\{\ell_{fit} = \text{near or distant future}\right\}, \tag{2.4}$$

where x_{fit} is the efficiency label of firm f and sentence i.

2.3 Lightcast Data

Finally, we construct a measure of firm actions related to AI that is not obtained from earnings calls, but from job postings from Lightcast (formerly Burning Glass Technologies). Each posting contains the raw text of the job ad, the raw and cleaned firm name, the skills listed in the posting, the associated occupation classification, the associated Lightcast sector, and the NAICS industry classification of the firm. We use the Lightcast sector "Artificial Intelligence" as our classification of AI-related job postings, using the skills-based classification approach of online vacancies by Acemoglu et al. (2022). If a job posting mentions at least one skill that is relevant to AI, the job is classified as AI-related.[3] For each firm, we calculate the total and AI-related job postings per quarter, and define the ratio as the AI vacancy share. We then link these data to the earnings call data. Details are available in Appendix A.

To obtain an intuition of what constitutes an AI-related vacancy, consider the following two job postings from January 2019. First, Apple posted a job vacancy with the title "Siri - Senior Software Engineer -Voice Experiences, Domains" for its voice assistant technology section in its headquarters in Cupertino, California. The job is classified as a software developer following the O*NET classification and requires skills such as object-oriented programming, algorithms, user assistance, Java, software development and workflow management. As none of the listed skills is present in the list of tentative AI skills, the vacancy does not fall into the Lightcast sector Artificial Intelligence. Second, Amazon

[3]The list of skills identified by Lightcast is updated annually for the preparation of the Stanford AI Index and applied retroactively to all job postings in the data, and the list of skills up to 2025 is reported in the 2025 edition of the AI Index (Maslej et al., 2025, pp. 432-434)

Web Services (part of Amazon Inc) posted a vacancy for a "Software Development Engineer" in its King, WA office to explicitly work on machine learning algorithms in connection to voice assistance technologies. The vacancy is classified as software developer by O*NET and lists skills such as Python, C, C++, algorithm design, and machine learning, among others. Since machine learning is on the list of skills related to artificial intelligence, the job is classified into the respective Lightcast sector.

3. Motivating patterns

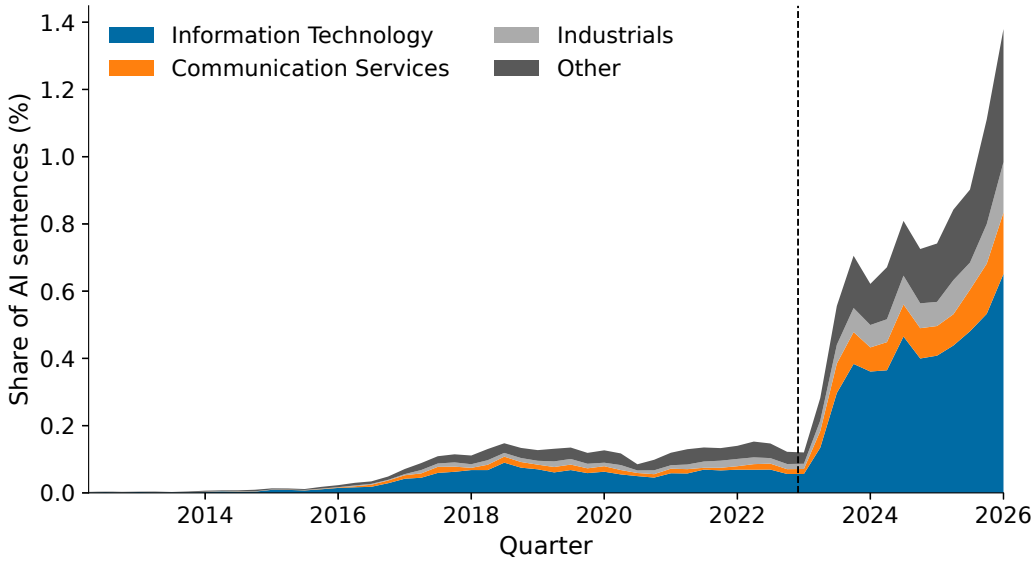
We next document salient patterns in the sentence- and firm-level data. We show that firm mentions of AI and AI-related actions surge following the announcement of the release of ChatGPT. The majority of AI mentions reflect a net positive perception of AI and suggests that firms engage in AI related actions. We also show that there is substantial heterogeneity in firm mentions of AI, as well as in their sentiment about it and in their associations of AI with efficiency and performance. We further document that the similarity in how firms talk about AI increases following the release of ChatGPT, suggesting that firms' views of AI converge somewhat. Lastly, we find evidence that firms' perceptions of AI are associated with stated and measured actions such as increasing the share of AI vacancies in job postings.

Surging AI mentions after ChatGPT. Figure 3 shows the share of AI sentences in all sentences over time. References of artificial intelligence do not appear at meaningful frequencies on earnings calls before 2017. Between 2017 and 2022 mentions of AI are positive but infrequent, accounting for between 0.1 and 0.15 percent of sentences. A notable increase occurs after the fourth quarter of 2022, when ChatGPT was released. Most recently, AI-related sentences account for more than one percent of all sentences in earnings calls. Firms classified as operating within the information technology sector account for the majority of these mentions.

Quantitatively, the increase in AI mentions after the release of ChatGPT is striking. Within a single year AI mentions approximately quintuple, with further increases thereafter. The increase is also broad based; it occurs within the information technology sector and outside of it. In the context of our model we will interpret the arrival of ChatGPT as a public signal about the future productivity of AI.

For the remainder of this paper our main analysis focuses on the period starting January 2017. We choose this starting point for two reasons. First, AI mentions are rare and unsystematic up to 2016. Although one can track the early onset of AI from roughly 2012, it is still part of very few earnings calls during that time. Second, the measurement error component is higher during the earlier period, since the acronym *AI* carries multiple interpretations (e.g., Autoplan International, Avian Influenza) – although we correct against this in our use of NLP by adding such examples to the training data. The data from 2017 onward include 200,749 AI-sentences from 34,455 documents across 6,604 firms.

Figure 3: Total AI Mentions by Industry



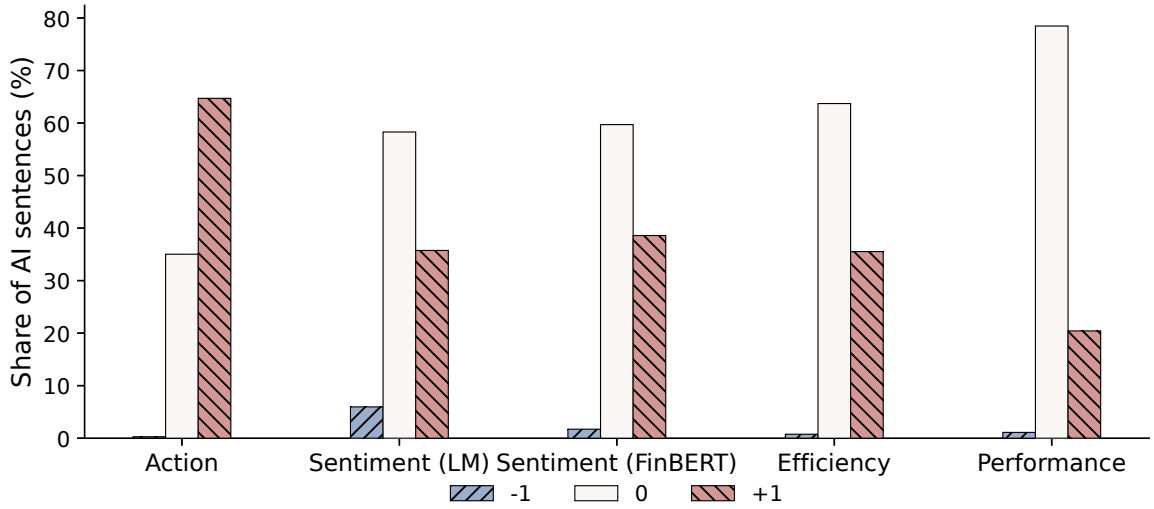
Notes: The figure shows total quarterly AI mentions divided by total sentences per quarter with a breakdown by industry. Vertical dashed line indicates ChatGPT release (Q4 2022). See Appendix A for details on the classification. Industry follows 2-digit GICS business classification

Mentions reflect positive perceptions and suggest action. Figure 4 illustrates the distribution of the labels of AI sentences at the sentence level. Several features are worth noting. First, there are few negative labels. The vast majority of AI related sentences do not express negative sentiment or associate AI negatively with efficiency, performance, or action. Around 35 percent of sentences receive positive sentiment and efficiency labels while only around 20 percent of sentences receive a positive performance label. This relatively small share might suggest that AI does not yet affect profitability or costs much. As our interpretation of AI as a news shock would suggest, this fact is consistent with the notion that the benefits of AI will only meaningfully affect such outcomes in the future. The large outlier is the action variable. More than 60 percent of the sentences suggest some form of AI-related action. Taken at face value, this suggests that firms are engaging in substantial AI-related activities.[4]

Substantial heterogeneity across firms. Figure 5 shows the distributions of x_{ft}, as defined in equation (2.3), for the firms' net action, net performance, net sentiment, and net efficiency measures before and after the release of ChatGPT. For all four labels more than 90 percent of firm-quarter observations are zero before ChatGPT. This reflects the fact that about 90 percent of firms do not mention AI before the release of ChatGPT as we document in Appendix Figure A6. Although the share of firms not mentioning AI in their earnings calls falls in the last quarter of 2025, it remains high at around 70

[4]Note that this does not imply firms are taking actions without positive AI sentiment, for instance. Within an earnings-call document, firms might discuss AI-action and AI-sentiment in different sentences, and so the sentence-level joint distribution of sentiment and action can differ from the document-level one. Appendix table A6 contains the pairwise correlations, by firm-quarter, across labels. The correlations are positive and high overall between all pairs of labels. In particular, firms' mentioning AI-related action is positively correlated with firms' sentiment towards AI with a correlation of nearly 0.9.

Figure 4: Label Distribution Across AI Sentences



Notes: The figure shows the distributions of various variables of AI sentences for the time period Q1 2017–Q4 2025. For each variable labels are mutually exclusive and therefore sum to one.

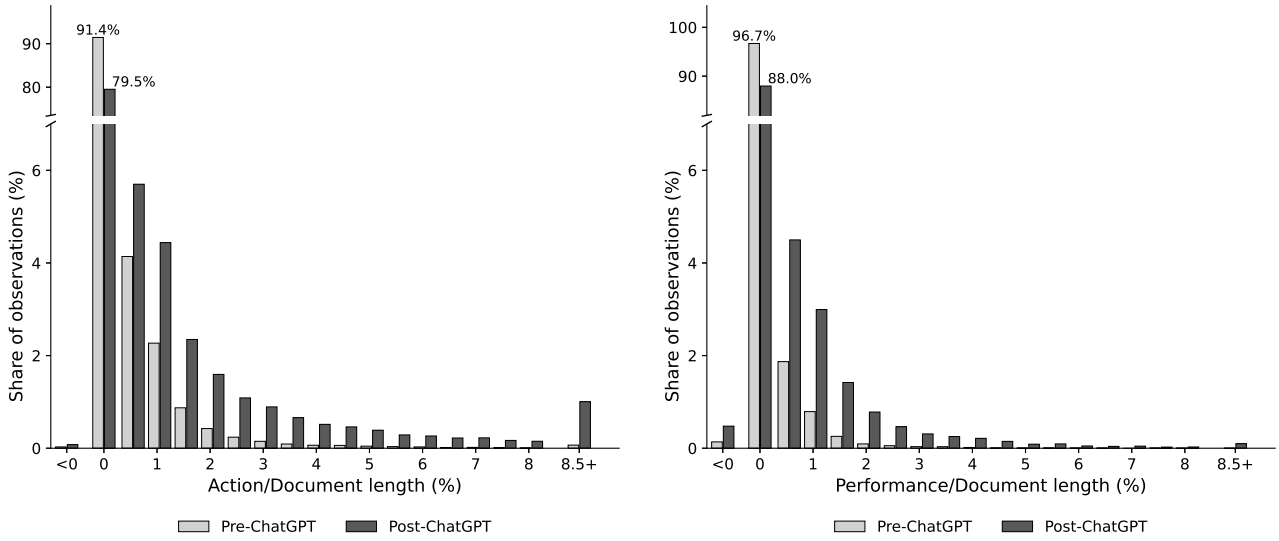
percent. Figure 5 also shows that, conditional on mentioning AI, x_{ft} varies substantially because some firms dedicate a larger share of the discussion in their earnings calls to AI (see Appendix Figure A7 for details). Although not easily visible in Figure 5, firms also vary substantially in the extent to which their AI-related sentences suggest action, effects on performance or efficiency, and in their sentiment towards AI. We document this in detail in Appendix Figure A8. Taken together, heterogeneity in these margins results in substantial heterogeneity in firms' measures of net action, net performance, net sentiment, and net efficiency. For instance, about one percent of firm-quarter observations have a net action measure equal to or above 8 percent after the release of ChatGPT, suggesting elaborate AI-related action. This contrasts with the large majority of other firms whose net action measure is below one percent.

AI-talk similarity. Despite the evident cross-firm heterogeneity that persists after the release of ChatGPT, there is a sense in which firms' views of AI are partially converging. To measure how firms' views of AI differ, we construct measures of AI-talk similarity at the firm level. Each AI-related sentence in period t is first mapped into a numerical vector representation $\mathbf{v}_{fit} \in \mathbb{R}^K$ using a sentence-embedding model, where f is the firm and i is the AI-sentence. A sentence embedding is a transformation of a sentence into a numeric vector containing meaningful information, using the approach in Mikolov et al. (2013). All sentence embeddings are normalized to unit length ($|\mathbf{v}_{fit}| = 1$).

For a given firm f containing n_f AI-related sentences with embeddings $\mathbf{v}_{f1t}, \ldots, \mathbf{v}_{fn_ft}$, we first construct a document-level embedding by averaging the sentence embeddings and then renormalizing to unit length.[5] The vector represents the overall direction of AI-related discourse in the transcript.

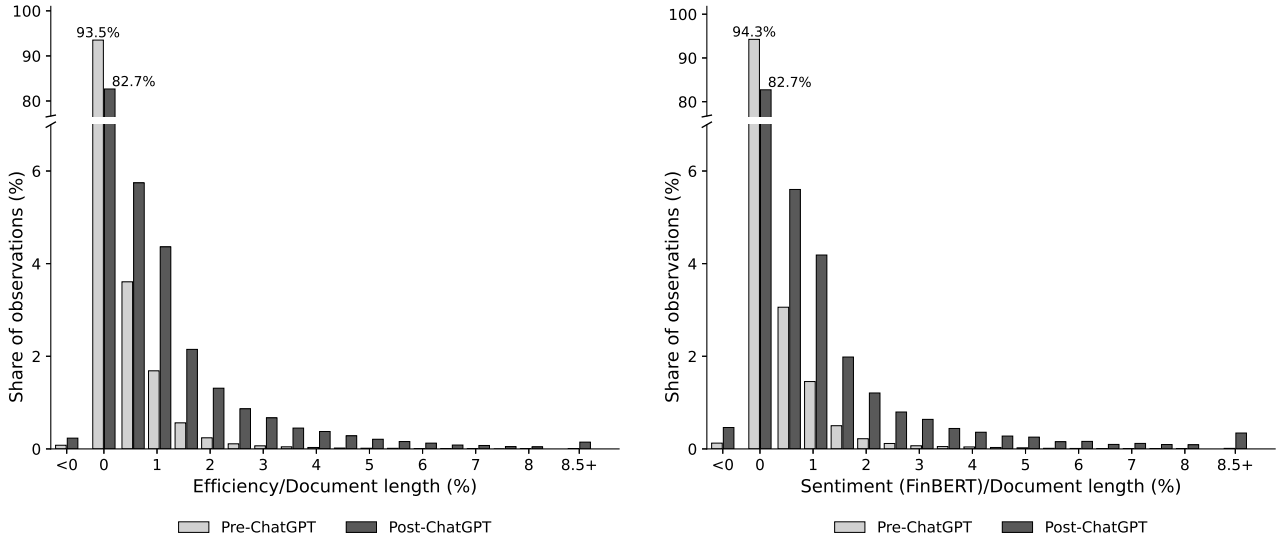
[5]Documents are defined for this instance only as the union of all AI-sentences of firm f in period t as other unrelated sentences do not contain relevant information about the AI content and only add noise when computing similarity scores (given companies operate in different industries and carry other firm-specific information that is not informative about

Figure 5: Heterogeneity in AI-related action and beliefs



(a) Net Action

(b) Net Performance



(c) Net Efficiency

(d) Net Sentiment (FinBERT)

Notes: The figure shows the distributions of the net action, net performance, net efficiency, and net FinBERT sentiment pre and post ChatGPT. For details on these measures, see Section 2. The pre-ChatGPT period runs from 2017q1 to 2022q3 and the post-ChatGPT period from 2022q4 to 2025q4. The binwidth is set to 0.5 percent and bars for the same bin pre and post ChatGPT are shown next to one other. Observation of less than 0 and more than 8 percent are grouped together. Observations that are exactly 0 receive own bar. The vertical axis is cut at 7 percent to improve legibility.

We then construct document AI cosine similarity. In some cases, we weight each firm by the number of AI sentences n_{ft}. Further details are in Appendix B. This measure helps assess whether firms in a cross-section are thinking similarly about AI, and its evolution over time helps assess whether beliefs about AI impacts are converging in firm discourse or not.

Figure 6 shows 2 versions of the cosine similarity measure, that differ with respect to the sample over which they are constructed. A key issue in measuring pairwise cosine similarity is that a large fraction of firms do not mention AI, as discussed above. We must therefore make assumptions on the unobserved AI-narratives of these firms, or exclude them from the sample. In the figure, we consider two cases. First, we focus on a sample of firms that have a minimum of one mention of AI in a quarter, and construct the (unweighted) pairwise AI-narrative similarity between this subset of documents. Second, we assume that if a firm does not discuss AI, its AI-discourse similarity with other firms' documents is 0. This sample therefore includes all firms with earnings calls in a quarter.[6] Alternatively, in Appendix B, we assume the firms that are not mentioning AI are similar in distribution to the firms that have the fewest AI mentions. We consider the distribution of similarities of firms in the bottom 10 percentile of AI mentions, with other firms mentioning AI and between themselves. We sample from the bottom 25 percentile of this distribution, and randomly assign these similarities to firms that are not observed mentioning AI. This approach assumes that AI mentions and AI beliefs are continuous and monotonically increasing, and the firms with no AI mentions are likely close to the bottom threshold of firms in terms of AI mentions and their similarities.

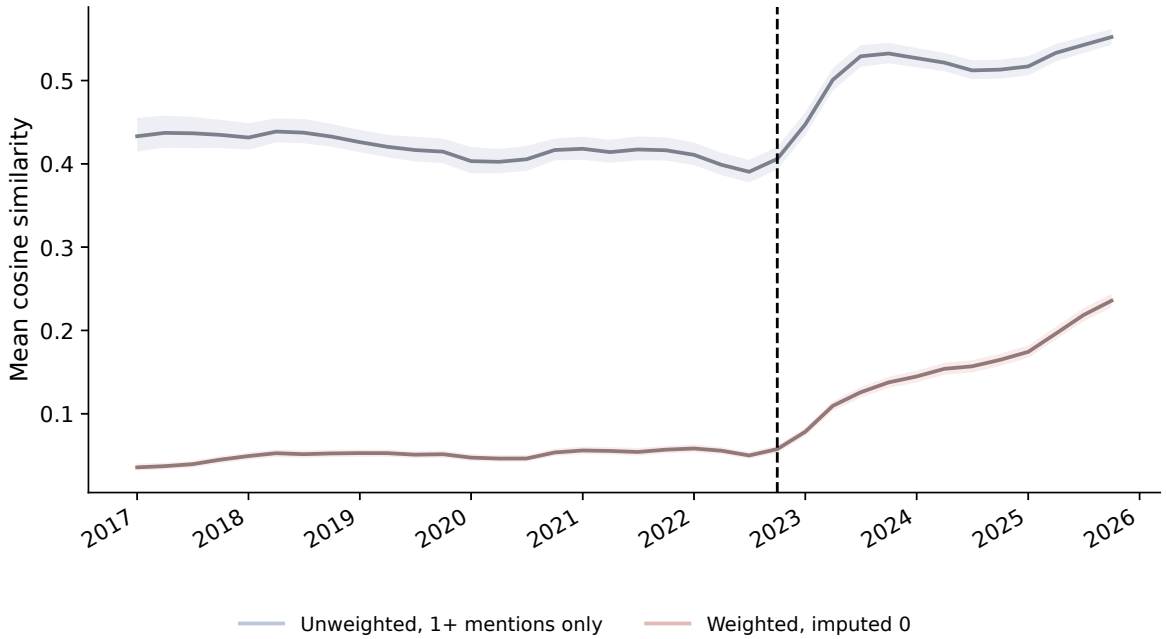
Notably, both versions of the measure are relatively flat initially, and then increase following the release of ChatGPT. For the intensive margin unweighted sample with ≥ 1 AI mentions, the average similarity in narratives increases by 21.4% (from 0.42 to 0.51) after ChatGPT. Assigning zeros to firms not mentioning AI results in an increase in similarity from 0.05 to 0.15 following ChatGPT. Appendix Figure A9 shows additional similarity measures using different quantile cutoffs for imputation as discussed above, restricting to only the sample of firms with a minimum of 5 AI mentions, and weighted and unweighted measures. In all cases, the similarity in AI-related narratives in firm earnings calls increases noticeably following the release of ChatGPT. For instance, with the 25 percentile imputation for firms without AI mentions, AI narrative similarity increases from 0.27 to 0.34. As we will see below, this is a feature consistent with interpreting the release of ChatGPT as the arrival of a public signal about AI's effect on future productivity, as this results in a synchronization of firm beliefs about AI.

AI-related actions are correlated with beliefs about AI. We next show that AI-related actions are positively correlated with various measures of beliefs about AI. Specifically, we estimate regressions

AI-related talk.)

[6]We only consider the weighted version of this sample. When firm-documents are weighted by the number of AI sentences in the document, we do not have to take a stance on the pairwise similarity between firms that both do not mention AI.

Figure 6: AI talk similarity



Notes: The figure displays semantic similarity of AI-related discussion at the document (firm-quarter) level, constructed as discussed in the text. Document-level represents the average embedding of all AI-sentences inside an earnings call. All embeddings are L2-normalized and the sample contains AI-flagged sentences only. Weighting follows the approach described in B.2. Under unweighted similarity all documents receive the same weight. *1+ mentions* only considers firm-quarter observations with at least 1 AI-mention. Under the *Imputed 0* measure, all pairwise combination that include any documents without AI sentences always receive a similarity of 0. Vertical marker indicates the ChatGPT release (Q4 2022). 95% confidence intervals are obtained via a transcript-cluster bootstrap per quarter. Mean and CI are smoothed with a centered three-quarter MA.

of the form

$$\text{Action}_{ft} = \beta \cdot \text{Belief}_{ft} + \delta_t + \delta_f + \varepsilon_{ft}, \tag{3.1}$$

where δ_t and δ_f are time and firm fixed effects, Action_{ft} is a particular measure of AI-related action and Belief_{ft} is a measure of beliefs. We estimate this specification on the quarterly firm panel dataset discussed above, considering the following measures. For Action_{ft} we consider the net action measure introduced above, but also the share of AI-related vacancies. Both of these measures are expressed as percent deviations from the average which average? the cross-sectional average within a quarter? one big pooled average?. For Belief_{ft} we consider the net efficiency, net performance, and net sentiment measures constructed above, as well as versions of these that explicitly condition on future time labels as in equation (2.4). These measures are also expressed as percent deviations from the average.

As shown in Panel A of Table 1, firms' actions and beliefs are robustly positively related. This is the case for all measures of beliefs – net efficiency, net performance, and the two net sentiment measures – as well as for both action measures – the net action measure constructed above and the AI-related vacancy share. We emphasize that the AI vacancy share is constructed from an entirely different source (Lightcast), and there is no mechanical relationship between this action measure and the beliefs measures. The elasticities we estimate with the AI-vacancy share as the dependent variable

Table 1: Actions and Beliefs

	Net Action				AI Vacancy Share			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Full sample								
Net Efficiency	0.790				0.071			
	(0.012)				(0.007)			
Net Performance		0.488				0.044		
		(0.016)				(0.005)		
Net Sentiment (LM)			0.670				0.058	
			(0.009)				(0.005)	
Net Sentiment (FinBERT)				0.745				0.062
				(0.013)				(0.006)
R^2	0.847	0.760	0.823	0.876	0.521	0.519	0.520	0.520
Observations	210,251	210,251	210,251	210,251	79,795	79,795	79,795	79,795
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: At least one AI-related sentence								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Net Efficiency	0.743				0.054			
	(0.014)				(0.007)			
Net Performance		0.431				0.030		
		(0.015)				(0.006)		
Net Sentiment (LM)			0.605				0.042	
			(0.012)				(0.005)	
Net Sentiment (FinBERT)				0.696				0.044
				(0.013)				(0.006)
R^2	0.835	0.764	0.810	0.868	0.573	0.571	0.572	0.572
Observations	32,518	32,518	32,518	32,518	15,202	15,202	15,202	15,202
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows estimates of specification (3.1). The AI Vacancy share is constructed from Lightcast data and only available for US firms. For the definitions of the remaining measures, see Section 2. Panel A shows estimates for the full sample while Panel B conditions on observations for which the underlying firms' earnings calls had at least one AI-related sentence. Standard errors clustered at the firm level and reported in parentheses. The sample period runs from 2017q1 to 2025q4.

are smaller than those for the generic action variable. This is consistent with the interpretation that the posting of AI-related vacancies is just one of many ways to take AI-related actions.

Panel B of Table 1 reports the same set of estimates after restricting the sample to firm-quarter observations for which the earnings calls have at least one AI-related sentence. This check is motivated by the possibility that the results in Panel A are potentially driven by the extensive margin of whether firms mention AI or not. As Panel B of Table 1 shows, this is not the case. The estimates are very similar to those on the full sample reported in Panel A. In Appendix Table A8 we report estimates that condition explicitly on future time labels, that is, if the time labels take the value near future or distant future. These specifications aim to approximate relationships between current actions

and expectations about future AI-related productivity. All coefficients are positive and significant, suggesting that higher expectations of future productivity and action are positively related.

Stronger correlation of actions and beliefs after ChatGPT. Next, we document that the association between AI-related actions and beliefs strengthens after the release of ChatGPT. To do so, we estimate the specifications of the form

$$\text{Action}_{ft} = \alpha_t + \beta \cdot \text{Belief}_{ft} + \delta \cdot \text{Belief}_{ft} \cdot \text{Post}_t + \gamma_f + \varepsilon_{ft}, \tag{3.2}$$

where post_t is an indicator that switches on after the release of ChatGPT in the last quarter of 2022, i.e. $\text{post}_t = 1\left\{t \geq 2022q4\right\}$.

Table 2 shows the estimates. The estimates in Panel A are on the full sample while the estimates in Panel B condition on at least one AI related sentence during the observation quarter. In all specifications, the coefficients on the interaction term are positive and significant. The economic magnitudes of these estimates are also sizable. On the full sample and with the net action measure as the dependent variable, the increase ranges from 11.8 percent of the main effect (= 0.079/0.668) for FinBert net sentiment to 52.2 percent of the main effect in the case of LM net sentiment (= 0.235/0.450). With the AI vacancy share as the dependent variable, it turns out that the coefficient on the main effect becomes insignificant. This suggests that firms began hiring AI workers systematically only after the release of ChatGPT.

Table 2: Actions and Beliefs pre and post ChatGPT

	Net Action				AI Vacancy Share			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Full sample								
Net Efficiency	0.597				-0.016			
	(0.019)				(0.021)			
Post × Net Efficiency	0.208				0.095			
	(0.025)				(0.023)			
Net Performance		0.337				-0.013		
		(0.014)				(0.024)		
Post × Net Performance		0.159				0.061		
		(0.021)				(0.024)		
Net Sentiment (LM)			0.450				-0.021	
			(0.019)				(0.018)	
Post × Net Sentiment (LM)			0.235				0.085	
			(0.021)				(0.020)	
Net Sentiment (FinBERT)				0.668				-0.019
				(0.017)				(0.026)
Post × Net Sentiment (FinBERT)				0.079				0.085
				(0.019)				(0.026)
R^2	0.850	0.763	0.828	0.876	0.523	0.520	0.522	0.521
Observations	210,251	210,251	210,251	210,251	79,795	79,795	79,795	79,795
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: At least one AI-related sentence								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Net Efficiency	0.572				-0.014			
	(0.023)				(0.024)			
Post × Net Efficiency	0.185				0.077			
	(0.031)				(0.026)			
Net Performance		0.302				-0.009		
		(0.016)				(0.025)		
Post × Net Performance		0.138				0.043		
		(0.023)				(0.025)		
Net Sentiment (LM)			0.382				-0.019	
			(0.023)				(0.020)	
Post × Net Sentiment (LM)			0.242				0.067	
			(0.026)				(0.021)	
Net Sentiment (FinBERT)				0.621				-0.019
				(0.020)				(0.027)
Post × Net Sentiment (FinBERT)				0.079				0.069
				(0.023)				(0.029)
R^2	0.837	0.767	0.816	0.869	0.575	0.572	0.574	0.574
Observations	32,518	32,518	32,518	32,518	15,202	15,202	15,202	15,202
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows estimates of specification (3.2), where the indicator variable post_t is defined as $\text{post}_t = 1\{t \geq 2022q4\}$. The AI Vacancy share is constructed from Lightcast data and only available for US firms. For the definitions of the remaining measures, see Section 2. Panel A shows estimates for the full sample while Panel B conditions on observations for which the underlying firms' earnings calls had at least one AI-related sentence. Standard errors clustered at the firm level and reported in parentheses. The sample period runs from 2017q1 to 2025q4.

4. Theoretical Framework

This section lays out a dynamic model of AI investment under dispersed information. Consider an economy comprised of a continuum of information islands indexed by i. Each island is populated by firms and workers that share an information set, but available information differs across islands. Firms use production workers and AI capital to produce output. AI capital is accumulated through investment undertaken by AI workers. Time is discrete and indexed by t. Each period contains two sub-periods. In the first, firms make production decisions and workers make labor supply decisions under incomplete information. In the second sub-period production takes place and goods markets clear.

4.1 Environment

Representative household. The economy contains a representative "big family" household that contains a single consumer and a continuum of workers, and owns the firms in the economy. The household solves

$$\max \mathbb{E}_t \left[\sum_{k=0}^{\infty} \beta^k \left(U(C_{t+k}) - \int_i \Gamma(L_{i,t+k}) di - \int_i V(N_{i,t+k}) di \right) \right],$$

s.t.

$$P_t C_t + B_{t+1} = R_t B_t + \int W_{it}^A L_{it} di + \int W_{it}^P N_{it} di + \int \pi_{it} di.$$

where C_t is the aggregate consumption, N_{it} production labor supplied to island i earning wage W_{it}^P, L_{it} is AI investment labor supplied to island i earning wage W_{it}^A, π_{it} is the profit of the firm on island i, and B_t are bond holdings earning return R_t. The instantaneous (dis)utilities have the following functional forms:

$$U(C) = \frac{C^{1-\gamma}}{1-\gamma}, \quad \Gamma(L) = \frac{L^{1+\eta}}{1+\eta}, \quad V(N) = \frac{N^{1+\varkappa}}{1+\varkappa},$$

where γ is the coefficient of relative risk aversion, $\varkappa$ is the inverse Frisch elasticity of labor supply, and η is the inverse Frisch elasticity of supply of AI investment labor.

The consumption bundle is a CES aggregate of the products of all islands. The bundle and its price index are:

$$C_t = \left[\int C_{it}^{\frac{\rho-1}{\rho}} di \right]^{\frac{\rho}{\rho-1}}, \qquad P_t = \left[\int P_{it}^{1-\rho} di \right]^{\frac{1}{1-\rho}},$$

where ρ is the elasticity of substitution. Household optimization leads to the following labor supply

curves:

$$V'(N_{it}) = \mathbb{E}_{it}\left[\frac{U'(C_t)}{P_t}\right] W_{it}^P, \qquad \Gamma'(L_{it}) = \mathbb{E}_{it}\left[\frac{U'(C_t)}{P_t}\right] W_{it}^A. \tag{4.1}$$

Firms and local labor market. Each firm is on an information island. The production function of firm on island i is given by

$$Y_{it} = F(A_t, M_{it}, N_{it}) = \left[\varpi(A_t M_{it})^{\frac{\xi-1}{\xi}} + (1-\varpi) N_{it}^{\frac{\xi-1}{\xi}}\right]^{\frac{\xi}{\xi-1}}, \tag{4.2}$$

where A_t is the AI-related exogenous technology level and M_{it} is the firm-level endogenous stock of AI capital, that evolves according to:

$$M_{it} = (1-\delta) M_{it-1} + L_{it}. \tag{4.3}$$

The firm is monopolistically competitive and faces downward-sloping demand for its product given by

$$P_{it}(Y_{it}, Y_t) = \left(\frac{Y_{it}}{Y_t}\right)^{-\frac{1}{\rho}} P_t. \tag{4.4}$$

We assume that on each information island, both firms and workers take wages as given. As the firm is owned by the representative household, it applies the household's stochastic discount factor when discounting future profits. The firm's problem can be written as

$$\max \mathbb{E}_{it}\left[\sum_{k=0}^{\infty} \beta^k \frac{U'(C_{t+k})}{P_{t+k}} \left(P_{it}(Y_{it+k}, Y_{t+K}) Y_{it+k} - W_{it+k}^P N_{it+k} - W_{it+k}^A L_{it+k}\right)\right]$$

subject to (4.2), (4.3), and (4.4) $\forall t$.

The first-order condition with respect to production labor is:

$$\mathbb{E}_{it}\left[\frac{U'(C_t)}{P_t} F_N(A_t, M_{it}, N_{it}) \left(1 - \frac{1}{\rho}\right) P_{it}\right] = \mathbb{E}_{it}\left[\frac{U'(C_t)}{P_t} W_{it}^P\right].$$

The first-order condition with respect to AI capital is:

$$\begin{aligned}
&\beta \mathbb{E}_{it}\left[\frac{U'(C_{t+1})}{P_{t+1}} F_M(A_{it+1}, M_{it+1}, N_{it+1}) \left(1 - \frac{1}{\rho}\right) P_{it+1} + (1-\delta)\frac{U'(C_{t+1})}{P_{t+1}} W_{it+1}^A\right] \\
=&\mathbb{E}_{it}\left[\frac{U'(C_t)}{P_t} W_{it}^A\right].
\end{aligned}$$

Combining these with workers' optimal labor supplies (4.1) yields

$$\mathbb{E}_{it}\left[\frac{U'(C_t)}{P_t} F_N(A_t, M_{it}, N_{it}) \left(1 - \frac{1}{\rho}\right) P_{it}\right] = V'(N_{it})$$

$$\beta \mathbb{E}_{it}\left[\frac{U'(C_{t+1})}{P_{t+1}} F_M(A_{t+1}, M_{it+1}, N_{it+1})\left(1-\frac{1}{\rho}\right) P_{it+1} + (1-\delta)\Gamma'(L_{it+1})\right] = \Gamma'(L_{it}).$$

4.2 Best Responses and Equilibrium Characterization

We focus on the linearized system and use small letters to denote log-deviations from steady state.

We start with firms' output choices. Given the amount of capital, this decision is essentially static and resembles Angeletos and La'O (2010). The following lemma provides the characterization of the static output response.

Lemma 1. *The static best response is*

$$y_{it} = \omega^p (a_t + m_{it}) + \omega^g \mathbb{E}_{it}[y_t], \tag{4.5}$$

where

$$\omega^p = \frac{\theta\left(\varkappa + \frac{1}{\xi}\right)}{\left(\varkappa + \frac{1}{\xi}\right) - (1-\theta)\left(\frac{1}{\xi} - \frac{1}{\rho}\right)} \qquad \text{and} \qquad \omega^g = \frac{(1-\theta)\left(\frac{1}{\rho} - \gamma\right)}{\left(\varkappa + \frac{1}{\xi}\right) - (1-\theta)\left(\frac{1}{\xi} - \frac{1}{\rho}\right)}.$$

and $\theta = \frac{\alpha(\bar{A}\bar{M})^{\xi}}{\alpha(\bar{A}\bar{M})^{\xi} + (1-\alpha)\bar{N}^{\xi}}$ *depends on the steady state levels of capital and labor.*

Proof. See Appendix C. □

Condition (4.5) illustrates that firms' output depends on the productivity a_t, its own stock of accumulated capital m_{it}, as well as the expected output choices of other firms. The degree of strategic complementarity is captured by the composite parameter ω^g, and the dependence on the fundamentals is captured by ω^p. Here and throughout, the superscript g denotes the general equilibrium force, while p denotes the partial equilibrium force, in which others' actions are held fixed.

Different from Angeletos and La'O (2010), firms must also choose AI investment labor, a decision that incorporates dynamic considerations.

Lemma 2. *The best response of* m_{it+1} *satisfies*

$$f_1 m_{it} + f_2 m_{it+1} + f_3 \mathbb{E}_{it}[m_{it+2}] + f_4 \mathbb{E}_{it}[y_{t+1}] + f_5 \mathbb{E}_{it}[a_{t+1}] = 0 \tag{4.6}$$

where

$$f_1 = \frac{\eta(1-\delta)}{\delta} \qquad f_4 = \eta\left(\frac{1}{\xi} - \frac{1}{\rho}\right)\omega^g + (1-\beta(1-\delta))\left(\frac{1}{\rho} - \gamma\right)$$

$$f_2 = (1-\beta(1-\delta))\left[\left(\frac{1}{\xi} - \frac{1}{\rho}\right)\omega^p - \frac{1}{\xi}\right] - \frac{\eta\left(1+\beta(1-\delta)^2\right)}{\delta} \qquad f_5 = (1-\beta(1-\delta))\left(\left(\frac{1}{\xi} - \frac{1}{\rho}\right)\omega^p + \frac{\xi-1}{\xi}\right).$$

$$f_3 = \frac{\eta\beta(1-\delta)}{\delta}$$

Proof. See Appendix C. □

Condition (4.6) characterizes the firm's dynamic investment responses m_{it+1}. This decision is both backward-looking as it depends on predetermined m_{it}, as well as forward-looking as it depends on expectations of future aggregate output and productivity shocks. Particularly, as future aggregate output is a function of aggregate capital, the investment decision features dynamic strategic complementarity. The following lemma expresses the aggregate outcomes in a semi-recursive form and highlights the nature of this dynamic game.

Lemma 3. *The aggregate outcomes satisfy the following dynamic system*

$$y_t = \omega^p \overline{\mathbb{E}}_t[a_t] + \omega^g \overline{\mathbb{E}}_t[y_t] + \omega^p m_t \tag{4.7}$$

$$m_{t+1} = \vartheta m_t + \sum_{k=0}^{\infty} \mu^k \left(k^g \overline{\mathbb{E}}_t[y_{t+k+1}] + k^p \overline{\mathbb{E}}_t[a_{t+k+1}] \right) \tag{4.8}$$

where $\overline{\mathbb{E}}_t[\cdot]$ *denotes the average expectation across information islands,* ϑ *is the inside root that satisfies*

$$f_1 + f_2 x + f_3 x^2 = 0,$$

and

$$\mu = -\frac{f_3}{f_2 + f_3 \vartheta}, \quad k^g = -\frac{f_4}{f_2 + f_3 \vartheta}, \quad k^p = -\frac{f_5}{f_2 + f_3 \vartheta}.$$

Proof. See Appendix C. □

Condition (4.7) is simply the aggregate version of condition (4.5). Without the dependence on m_t, it characterizes the static type of fixed point problem as in Morris and Shin (2002), where the aggregate outcome depends on not only the fundamentals, but also the beliefs about the fundamentals, the beliefs about other firms' beliefs about the fundamental and so on. With the presence of m_t, the aggregate output further hinges on the dynamics of the AI capital m_t, as well as the associated beliefs about these dynamics.

Condition (4.8), on the other hand, underscores the dynamic aspects of the coordination motive. First, the composite parameter ϑ summarizes the degree of dependence on past AI capital accumulation. This backward-looking component depends solely on the firm's own capital accumulation dynamics and is independent of both expectations about other firms' decisions and the process governing technology shocks. On the other hand, the forward-looking component depends on the entire expected future paths of $\{y_{t+k}\}$ and $\{a_{t+k}\}$. Importantly, the dependence of future aggregate output implies that firms need to form beliefs about other firms' future AI capital stock, which themselves depend on other firms' past capital stock and their expectations about future aggregate output and productivity.

The effects of intertemporal dependence on other firms' investment decisions depend on both the structure of information and the horizon over which firms are forward-looking. When information about future productivity shocks is common knowledge, firms understand that all other firms are aware of the shocks. The response to the shocks will internalize the effects of strategic complementarity fully and the implicit intertemporal coordination will be perfect. In contrast, when each firm receives some private information about future productivity shocks, their response will be arrested as they are not quite sure whether other firms are aware of the shocks. In addition, they understand that other firms also share the same concern and the effects of uncertainty will be amplified.

The aforementioned channel also depends on how far in the future the productivity shock is realized. If the shock occurs in the next period, firms only need to consider how other firms make decisions in the current period. By contrast, if the shock lies further in the future, firms must account for how other firms will decide in the next period, how those firms anticipate others' decisions in subsequent periods, and so on. The required coordination becomes substantially more complex.

In what follows, we will specify a particular information structure and provide a more transparent analysis of how firms react to noisy news.

4.3 News, Information, and Productivity Shock

We focus on a scenario in which agents know that at some future period T, there will be an AI productivity shock a_T. Once the shock realizes, future technology remains at that level, i.e.,

$$a_t = \begin{cases} 0, & \text{for } t < T, \\ a_T, & \text{for } t \geq T. \end{cases} \tag{4.9}$$

While T is common knowledge, the value of a_T is uncertain for agents. At $t = 0$, firm i receives a private noisy news about a_T

$$x_i = a_T + u_i, \qquad u_i \sim \mathbb{N}(0, \tau_u^{-1}),$$

where $\tau_u = \frac{\mathbb{V}(a_T)}{\mathbb{V}(u_i)}$ represents the precision of the private news. Importantly, the news signal x_i is privately observed by firm i. This captures the notion that, at the early stage of awareness of the future AI productivity shock, its prospects are highly uncertain, and firms form heterogeneous beliefs about its potential magnitude.

Additionally, we also consider the possibility that at some intermediate period $0 < t < T$, all firms receive a public signal about a_T

$$z = a_T + \varepsilon, \qquad \varepsilon \sim \mathbb{N}(0, \tau_\varepsilon^{-1}),$$

where $\tau_\varepsilon = \frac{\mathbb{V}(a_T)}{\mathbb{V}(\varepsilon)}$ represents the precision of the public news. This captures the idea that the release

of ChatGPT in late 2022 constituted a widely observed public signal that AI would become a major driver of productivity growth.

In periods $t < T$, we use $\mathbb{E}_i[\cdot]$ to denote firm i's expectation conditional on x_i and z, and $\overline{\mathbb{E}}[\cdot]$ to denote the average expectation. We assume that no additional learning occurs between periods 0 and T. This is a simplifying assumption adopted for tractability. Allowing firms to learn from noisy observations of past aggregate outcomes would introduce the additional complication that signal informativeness becomes endogenous. For the purpose of isolating the key effects of incomplete information, it is sufficient to work with an information structure where the information flow is exogenous.

Following the realization of the shock a_T, for $t \geq T$, information is assumed to be fully revealed, and the economy evolves under perfect information without uncertainty.

4.4 Dynamic Network Interpretation

To proceed, we will make one additional simplifying assumption on the parameters: $\omega^g = 0$. One parameter configuration under which this will occur is if the Frisch elasticity on the production workers is 0, i.e., $\varkappa \to \infty$. In this case labor supply of production workers to each island is fixed, and thus the firm cannot adjust output within a period, as the other factor m_{it} is predetermined within a period. This sidelines the static within-period interdependence of firm output on aggregate output. This interdependence – which usually takes the form of strategic complementarity – is well-studied in the literature and its implications are well-known.[7]

Lemma 4. *Assume $\omega^g = 0$ and the shock process satisfies (4.9). The system (4.7)-(4.8) simplifies to*

$$m_{t+1} = \vartheta m_t + \sum_{k=0}^{\infty} \mu^k \left(\kappa^g m_{t+k+1} + \kappa^p a_T\right), \qquad \text{for} \quad t \geq T \tag{4.10}$$

and

$$m_{t+1} = \vartheta m_t + \sum_{k=0}^{\infty} \mu^k \kappa^g \overline{\mathbb{E}}[m_{t+k+1}] + \sum_{k=T}^{\infty} \mu^{k-(t+1)} \kappa^p \overline{\mathbb{E}}[a_T], \qquad \text{for} \quad t < T, \tag{4.11}$$

where

$$\kappa^g = \omega^p k^g, \quad \text{and} \quad \kappa^p = k^p + \omega^p k^g.$$

By restricting $\omega^g = 0$, one can focus on the univariate system that involves only the capital stock dynamics. When $t \geq T$, the current and future technology shocks have all been realized, and this is common knowledge in the economy. As a result, the inference about future capital stock is perfect

[7] Note we make this assumption for clear analytical results regarding the dynamics of the model, but we relax it in the quantification in Section 5.

in equation (4.10). In contrast, when $t < T$, the technology shock has not realized yet, reflected in the additional discounting in equation (4.11), and firms have to form beliefs about the paths of the aggregate capital stock.

Regime after $t = T$. In period $t \geq T$, information is perfect and the technology remains constant at $a_t = a_T$. As a result, condition (4.10) implies that m_t admits a simple recursive formulation, stated in the following lemma.

Lemma 5. *When $t \geq T$, m_{t+1} obeys*

$$m_{t+1} = \alpha m_t + \nu a_T, \tag{4.12}$$

where $\alpha \in (0,1)$ and ν satisfy

$$\alpha = \vartheta + \frac{\alpha\kappa^g}{1-\alpha\mu}, \qquad \nu = \frac{\kappa^p(1-\alpha\mu)}{(1-\mu)(1-\alpha\mu)-\kappa^g}.$$

This recursive formulation is standard in macroeconomics. Firms understand that once period T is reached, there is no role for beliefs about the technology shock and the capital paths. The interesting model dynamics occur before the shock realization.

Regime before $t = T$. When $t < T$, firms in general have incomplete information about a_T and the path of $\{m_t\}$. Firms' investment decisions depend on their expectations about the investment choices of other firms. On the one hand, this requires forecasting other firms' future actions. On the other hand, because current investment depends on the existing capital stock, firms must also infer other firms' past capital accumulation. These intertemporal considerations admit the following network representation.

Lemma 6. *The sequence of capital stock $\boldsymbol{m}^T = \begin{bmatrix} m_1 & \dots & m_T \end{bmatrix}'$ satisfies*

$$\begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_T \end{bmatrix} = \mathcal{L}\boldsymbol{\varphi}\overline{\mathbb{E}}[a_T] + \mathcal{L}\boldsymbol{\gamma} \begin{bmatrix} \overline{\mathbb{E}}[m_1] \\ \overline{\mathbb{E}}[m_2] \\ \vdots \\ \overline{\mathbb{E}}[m_T] \end{bmatrix}$$

where

$$\mathcal{L} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \vartheta & 1 & 0 & \cdots & 0 \\ \vartheta^2 & \vartheta & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vartheta^{T-1} & \dots & \vartheta^2 & \vartheta & 1 \end{bmatrix}, \quad \boldsymbol{\gamma} = \kappa^g \begin{bmatrix} 1 & \mu & \cdots & \mu^{T-2} & \mu^{T-1}(1-\alpha\mu)^{-1} \\ 0 & 1 & \cdots & \mu^{T-3} & \mu^{T-2}(1-\alpha\mu)^{-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \mu(1-\alpha\mu)^{-1} \\ 0 & 0 & \cdots & 0 & (1-\alpha\mu)^{-1} \end{bmatrix}, \quad \boldsymbol{\varphi} = \left(\kappa^p + \frac{\nu\kappa^g\mu}{1-\alpha\mu}\right) \begin{bmatrix} \mu^{T-1} \\ \mu^{T-2} \\ \vdots \\ \mu \\ 1 \end{bmatrix}$$

This representation recasts the dynamic equation (4.10) into a network-like static problem. One can view the sequence $\{m_1, m_2, \ldots, m_T\}$ as a system of simultaneous decisions. The vector φ summarizes the forward-looking dependence on future realized technology shock a_T, taking into account the partial equilibrium impact before T (via κ^p), the general equilibrium impact after T (via $\kappa^g \frac{\nu\mu}{1-\alpha\mu}$), and the proper horizon-dependent discounting (via μ^t). The upper triangular matrix γ summarizes the forward-looking dependence on future aggregate m_t, and the lower triangular matrix $\mathcal{L}$ summarizes the backward-looking dependence on past aggregate m_t.

Under perfect information, all firms share the same perfect belief about a_T. In this case, all firms are identical and can perfectly infer the aggregate path $\overline{\mathbb{E}}[m_t] = m_t$. Therefore, the solution to the transition path $\{m_1, m_2, \ldots, m_T\}$ recovers the familiar Leontief inverse matrix.

Lemma 7. *Under perfect information, the solution is* $m_t^* = \mathcal{M}_t^* a_T$

$$\begin{bmatrix} \mathcal{M}_1^* & \mathcal{M}_2^* & \ldots & \mathcal{M}_T^* \end{bmatrix}' = (\mathbf{I} - \mathcal{L}\gamma)^{-1}\mathcal{L}\varphi.$$

In principle, the aggregate outcomes hinge on the dynamic coordination among firms. The expansion of the solution in Lemma 7 can be written as

$$\begin{bmatrix} m_1^* & \ldots & m_T^* \end{bmatrix}' = \mathcal{L}\varphi\, a_T + (\mathcal{L}\gamma)\mathcal{L}\varphi\, a_T + (\mathcal{L}\gamma)^2\mathcal{L}\varphi\, a_T + \ldots \tag{4.13}$$

where the $i-$th term represents the $i-$th order network effects. The assumption that information is perfect has the implication that the dynamic coordination among firms is also perfect, in the sense that the inference about other firms' decision is perfect, and the inference about other firms' inference about other firms' decision is perfect, and so on. Clearly, this result relies on both a stringent notion of rationality and strong requirements on the information structure. In what follows, we will relax the assumption on the information structure, which will also make the dynamic coordination imperfect as a byproduct.

Now we allow firms to receive noisy information about a_T as specified in Section 4.3. In this case, the beliefs about a_T is imperfect and asymmetric. As a result, firms need to form beliefs about other firms' beliefs, and the coordination is no longer frictionless. The following result characterizes how the network effects are modified under incomplete information.

Proposition 4.1. *With incomplete information, the sequence* $\{m_1, \ldots, m_T\}$ *satisfies*

$$m_t = \mathcal{M}_t^* \sum_{k=0}^{\infty} r_t^k \overline{\mathbb{E}}^{k+1}[a_T], \tag{4.14}$$

where $\overline{\mathbb{E}}^k[\cdot] = \overline{\mathbb{E}}\left[\overline{\mathbb{E}}^{k-1}[\cdot]\right]$ *are higher-order expectations recursively defined, and* r_t^k *for* $t \in \{1, \ldots, T\}$ *is the*

$t-$th element of the matrix $(\mathcal{L}\gamma)^k\mathcal{L}\varphi$. Jointly the r_t^k's satisfy

$$\sum_{k=0}^{\infty} r_t^k = 1.$$

Proposition 4.1 establishes that the counterpart of the network effects expansion (4.13) under incomplete information becomes

$$\begin{bmatrix} m_1 & \dots & m_T \end{bmatrix}' = \mathcal{L}\varphi\overline{\mathbb{E}}[a_T] + (\mathcal{L}\gamma)\mathcal{L}\varphi\overline{\mathbb{E}}^2[a_T] + (\mathcal{L}\gamma)^2\mathcal{L}\varphi\overline{\mathbb{E}}^3[a_T] + \dots \tag{4.15}$$

where the $k-$th order network effects are tied with $k-$th order higher-order expectations as in Bui et al. (2022). This is intuitive: higher-order network effects require firms to engage in deeper general equilibrium reasoning, effectively forming expectations about others' expectations of the fundamental, rather than simply expectations about the fundamental itself.

The solution (4.14) deviates from the perfect information benchmark $\mathcal{M}_t^*$ in two ways. First, with common but imperfect information, first-order and higher-order expectations are identical but differ from the technology shock, i.e., $\overline{\mathbb{E}}^k[a_T] = \overline{\mathbb{E}}^j[a_T] \neq a_T$. This corresponds, for example, to the case in which all firms observe only the public signal. Second, with incomplete information, first-order and higher-order expectations are different, and the response of m_t relies on a weighted sum of these expectations, where the weights given by $\{r_t^k\}$ are horizon $t-$dependent.

Therefore, Proposition 4.1 makes clear that understanding the investment decision requires two key elements: (1) the behavior of the weighting scheme $\{r_t^k\}$; and (2) the behavior of higher-order expectations. We now proceed to characterize both.

Weighting scheme. How does the weighting scheme $\{r_t^k\}$ evolve over time t, or equivalently, with the distance to the realization date T? To answer this question, it is useful to momentarily assume that the dependence on past capital stock is muted, i.e., $\vartheta = 0$ and $\mathcal{L}$ becomes the identity matrix. In this case, once the shock is realized, all first-order and higher-order uncertainty will be resolved. This implies that in period $t = T - 1$, firms only need to form beliefs about other firms' decisions in the current period, while future paths are recursively determined as functions of these decisions, as characterized by condition (4.12). In period $t = T - 2$, however, firms must form beliefs not only about other firms' decisions in period $t = T - 2$, but also about other firms' beliefs regarding others' decisions in period $t = T - 1$. As a result, the equilibrium places greater weight on higher-order expectations. More generally, the further the economy is from period T, the more important the general equilibrium feedback effects become, and the weighting scheme $\{r_t^k\}$ shifts toward higher-order terms. This argument remains valid when the backward-looking motive is present, and the following proposition formalizes it.

Proposition 4.2. *Fix the realization period T. The weights have the following properties*

- *The loading on first-order expectations, r_t^0, is increasing in t.*

- *The loading on higher-order expectations, r_t^k, is decreasing in t for all $k > 1$.*

As calendar time approaches the realization date T, the weight placed on first-order expectations increases, whereas the weights on higher-order expectations decline. In a similar vein, as T increases, the investment at a fixed time t will involve more higher-order beliefs, as the distance between t and T increases.

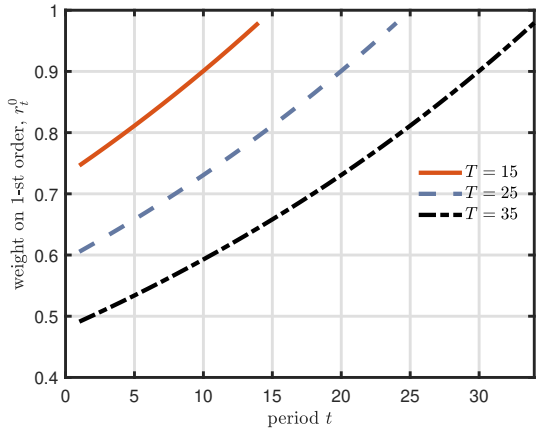
Proposition 4.3. *Fix a time t. The weights have the following properties*

- *The loading on first-order expectations, r_t^0, is decreasing in T.*
- *The loading on higher-order expectations, r_t^k, is increasing in T for all $k > 1$.*

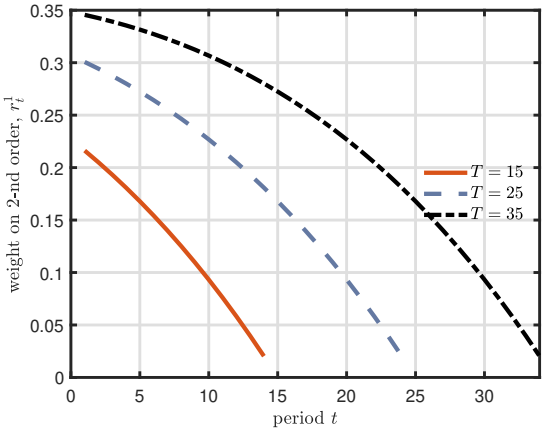
Figure 7 helps visualize the patterns in Proposition 4.2 and Proposition 4.3. The left panel shows that for a fixed realization date T, the weight on first-order expectation r_t^0 is increasing in t (moving along a line). But for a fixed calendar time t, r_t^0 is decreasing in T. The right panel shows that the exact opposite pattern shows up for the weight on the second-order expectation.

Figure 7: Patterns of Weighting Scheme

(a) patterns of r_t^0



(b) patterns of r_t^1



Expectations and macro implications. Now we move to the properties of the higher-order expectations. We start with the case where only noisy private information is available, that is, firm i only observes x_i when forming beliefs about a_T. In this case, the average first-order and higher-order expectations are

$$\bar{\mathbb{E}}^k[a_T] = \lambda^k a_T, \qquad \text{where} \quad \lambda = \frac{1}{1+\tau_u}. \tag{4.16}$$

Clearly, higher-order expectations are less responsive to the technology shock, a property that is common in models with incomplete information. In our environment, this property has direct

implications on the aggregate outcome. Applying the expectations according to (4.16), Proposition 4.1 becomes

$$m_t = \mathcal{M}_t^* \sum_{k=0}^{\infty} r_t^k \lambda^{k+1} a_T.$$

As $\lambda \in (0,1)$, it is clear that the responses relatively to $\mathcal{M}_t^*$ is more dampened as the weights $\{r_t^k\}$ shifts more towards terms with higher k. Propositions 4.2 and 4.3 then imply the following results.

Proposition 4.4. *With only private information, the response of m_t to a_T is attenuated relative to the perfect-information benchmark m_t^*, with the attenuation being stronger in earlier periods t and as T becomes larger.*

Next, we turn to understanding the effects of public news on aggregate investment. We consider the case in which the additional public signal z_t is also available in addition to private signals. How does this public signal alter the responses of $\{m_t\}$ relative to the previous case, in which firms observe only private information about a_T? This question boils down to the change of the first-order and higher-order expectations.

Lemma 8. *Denote $\Delta\overline{\mathbb{E}}^k[a_T]$ as the difference between the $k-$th order expectation conditional on both private and public signals and the one conditional on only private signals*

$$\Delta\overline{\mathbb{E}}^k[a_T] = \mathcal{A}_k a_T + \mathcal{B}_k \varepsilon.$$

Both $\mathcal{A}_k$ and $\mathcal{B}_k$ are positive and increasing with k.

The introduction of the public signal has three distinctive effects. First, it provides additional information about the technology shock a_T, which is reflected by a positive $\mathcal{A}_1$ or a responsive first-order expectation. Secondly, it helps to reach common knowledge by reducing higher-order uncertainty. This is captured by the fact that $\mathcal{A}_k$ is increasing in the order k. Thirdly, it introduces the dependence on the sentiment shock, $\mathcal{B}_k$, and this dependence becomes stronger with a higher k, as higher-order expectations rely more on public signals.

Taking these insights together, we obtain the following implications for aggregate outcomes.

Proposition 4.5. *Consider the impact when an additional public signal is introduced.*

- *Its relative effect on the response of m_t to the technology shock a_T, measured by $\Delta m_t / m_t^*$, is smaller in earlier periods t and for larger values of T.*
- *Its relative effect on the response of m_t to the sentiment shock ε, measured by $\Delta m_t / m_t^*$, is larger in earlier periods t and for larger values of T.*

In contrast, suppose the additional signal is not a public signal but another private signal with the same precision τ_ε. Though the change in first-order expectation is identical, the changes in

higher-order expectations are smaller, as there is less reduction in higher-order uncertainty. Similarly, suppose initially the information is complete, then the impact of a public signal will be smaller as well, as the additional role of eliminating lack of common knowledge will be absent.

5. Quantification

5.1 Calibration strategy

Param.	Value	Data Source	Description
Exogenously determined parameters			
$\varkappa$	0.5	—	production labor Frisch elasticity
ρ	3.5	—	elasticity of substitution
γ	0.2	—	income elasticity
β	0.98	—	discount factor
δ	0.02	—	depreciation rate
θ	0.40	—	capital share in production
T	50	—	distance to shock realization
Endogenously determined parameters			
τ_u	0.11	text similarity	private signal noise
τ_ε	0.08	text similarity	public signal noise
η	1.25	AI sentiment and action regression	R&D Frisch elasticity

We need to use two facts to calibrate the endogenous parameters.

1. We use the similarity of AI sentences before the ChatGPT release to pin down the precision of the private signal about a_T, and the similarity of AI sentences after the release to further pin down that of the public signal. The similarity is 0.10 and 0.35, respectively.

   With only private signal, the cosine similarity is given by

   $$\cos\big(\mathbb{E}_i[a_T], \mathbb{E}_j[a_T]\big) = \frac{\tau_u}{1+\tau_u}$$

   With the additional public signal, the cosine similarity becomes

   $$\cos\big(\mathbb{E}_i[a_T], \mathbb{E}_j[a_T]\big) = \frac{(\tau_u+\tau_\varepsilon)^2+\tau_\varepsilon}{(\tau_u+\tau_\varepsilon)^2+\tau_u+\tau_\varepsilon}.$$

   These two conditions help pin down τ_u and τ_ε, respectively.

2. At the micro level, the model implies the there is a positive correlation between firms' belief about the technology shock and the change of the labor share of AI research workers. Denote

$s_{it} = \frac{L_{it}}{L_{it}+N_{it}}$ as the AI workers' share. It follows that

$$s_{it} \approx \bar{s} + \bar{s}(1-\bar{s})(\ell_{it} - n_{it})$$

where $\bar{s} = \frac{L}{L+N}$. Consider the following regression

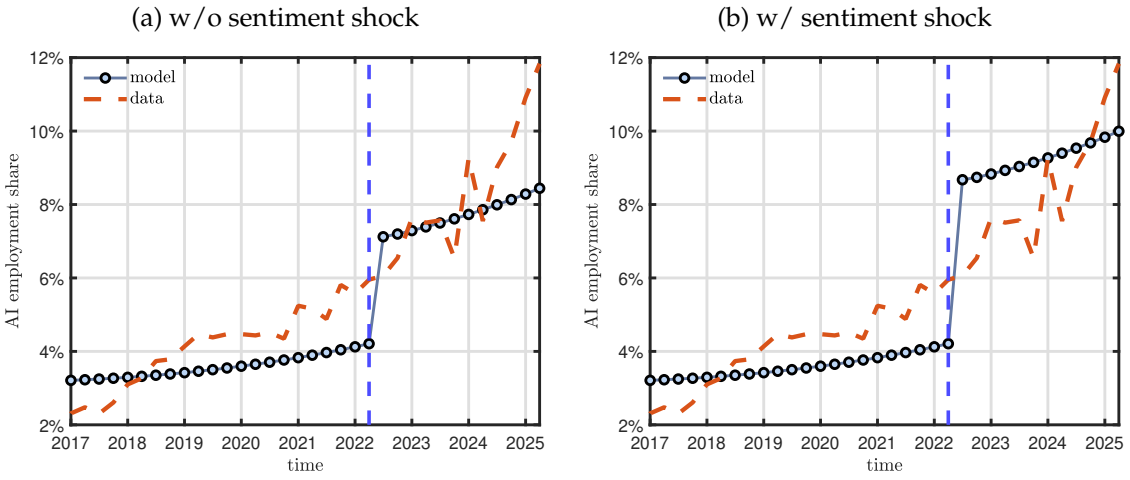
$$s_{it} = \beta_t \mathbb{E}_{it}[a_T] + \text{residual}_{it}$$

In the model, there is a pass-through from beliefs to changes in employment share in every period. We compute the average β_t in the model over time to match the corresponding empirical moment, 0.024. This yields the R&D Frisch elasticity η to be 1.25.

5.2 Results

We first present the baseline results. Similar to the real economy, we assume that in the first 6 years in the model from 2016 to 2022, firms only rely on private signals about the AI technology shock. After 2022, the public signal realizes, which captures the idea that ChatGpt was released later on.

Figure 8: Response to News Shock: Model vs Data



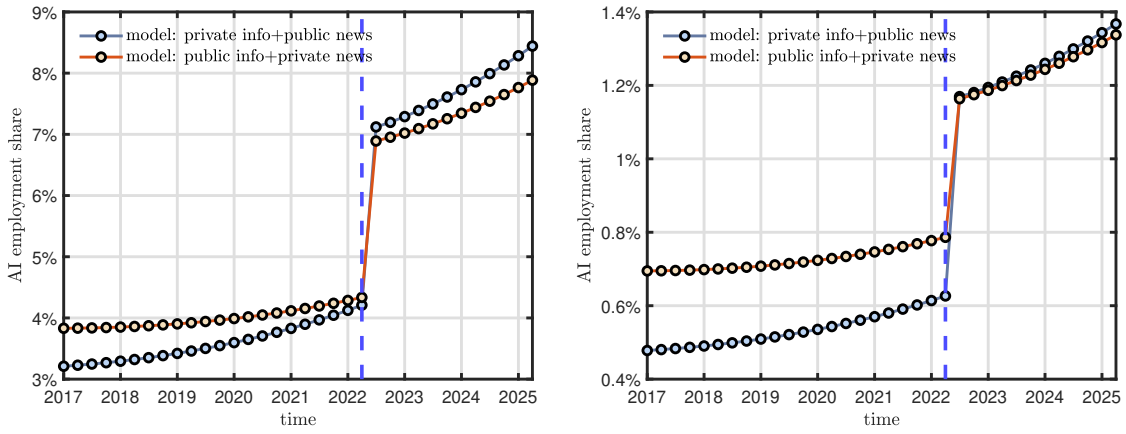
- The left panel of Figure 8 compares the model with the data counterpart. There is a slow increase of the employment share in the beginning, which is largely due to the dispersed information about the potential of the AI shock.

- After the release of ChatGPT, there is a discrete jump of the AI employment share. This is due to two effects: first, more information is available and firms have better idea about the potential of AI; second, the public signal helps firms to confirm that other firms also receive this information. The strategic complementarity induces firms to further increase their investment as when they are sure that other firms will also do so.

- The right panel of Figure 8 shows that with a positive realization of the sentiment shock, the spike of AI investment will further increase.

Figure 9: Comparison with Alternative Information Structure

(a) $T = 50$ (b) $T = 100$



- The left panel of Figure 9 compares the baseline model with a hypothetical scenario: the initial information is public, but later on the news is private. We choose the precision of the public signal to be the same as that of the original private signal, and the precision of the private news to be the same of the original public news. Therefore, the first-order expectations about a_T are the same in the comparison. However, the baseline model shows an initial smaller response followed by a larger jump when the news arrives later, consistent with the intuition developed in Proposition 4.5.

- The right panel shows that supposing the realization of the shock is in 25 years instead of 12 years, the role of incomplete information and the reduction of the higher-order uncertainty both loom larger, as shown in Proposition 4.5.

6. Conclusion

References

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo. 2022. "Artificial Intelligence and Jobs: Evidence from Online Vacancies." *Journal of Labor Economics* 40 (S1):S293–S340.

Acharya, Sushant, Jess Benhabib, and Zhen Huo. 2021. "The anatomy of sentiment-driven fluctuations." *Journal of Economic Theory* 195:105280.

Angeletos, George-Marios and Jennifer La'O. 2010. "Noisy business cycles." *NBER Macroeconomics Annual* 24:319–378.

———. 2013. "Sentiments." *Econometrica* 81 (2):739–779.

Angeletos, George-Marios, Guido Lorenzoni, and Alessandro Pavan. 2022. "Wall Street and Silicon Valley: A Delicate Interaction." *Review of Economic Studies* 90 (3):1041–1083.

Araci, Dogu. 2019. "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models."

Atolia, Manoj and Ryan Chahrour. 2020. "Intersectoral linkages, diverse information, and aggregate dynamics." *Review of Economic Dynamics* 36:270–292.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson. 2024. "Artificial intelligence, firm growth, and product innovation." *Journal of Financial Economics* 151:103745.

Barsky, Robert B. and Eric R. Sims. 2011. "News shocks and business cycles." *Journal of Monetary Economics* 58 (3):273–289.

Baslandze, Salomé, Zachary Edwards, John Graham, Ty McClure, Brent H Meyer, Michael Sparks, Sonya R Waddell, and Daniel Weitz. 2026. "Artificial Intelligence, Productivity, and the Workforce: Evidence from Corporate Executives." Working Paper 34984, National Bureau of Economic Research.

Beaudry, Paul and Franck Portier. 2006. "Stock Prices, News, and Economic Fluctuations." *American Economic Review* 96 (4):1293–1307.

Benhabib, Jess, Pengfei Wang, and Yi Wen. 2015. "Sentiments and Aggregate Demand Fluctuations." *Econometrica* 83:549–585.

Bergemann, Dirk, Tibor Heumann, and Stephen Morris. 2017. "Information and Interaction." Tech. rep., Yale University. Cowles Foundation Discussion Paper No. 2088.

Blanchard, Olivier Jean, Jean-Paul L'Huillier, and Guido Lorenzoni. 2013. "News, Noise, and Fluctuations: An Empirical Exploration." *American Economic Review* 103 (7):3045–70.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at Work." *The Quarterly Journal of Economics* 140 (2):889–942.

Bui, Ha, Zhen Huo, Andrei A Levchenko, and Nitya Pandalai-Nayar. 2022. "Noisy global value chains." Tech. rep., National Bureau of Economic Research.

Bybee, Leland, Bryan T Kelly, Asaf Manela, and Dacheng Xiu. 2023. "Business news and business cycles." Forthcoming, *Journal of Finance*.

Chahrour, Ryan, Kirstoffer Nimark, and Stefan Pitschner. 2021. "Sectoral Media Focus and Aggregate Fluctuations." *American Economic Review* 111 (12):3872–3922.

Faria-e Castro, Miguel and Serdar Ozkan. 2026. "Can AI Optimism Raise Inflation? What a Standard Macro Model Says." *St. Louis Fed On the Economy* .

Gazzani, Andrea and Filippo Natoli. 2024. "The macroeconomic effects of AI-based innovation." Tech. rep., Working Paper.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature* 57 (3):535–74.

Goertz, Christoph, Christopher Gunn, and Thomas Lubik. 2025. "What Can News Shocks Tell Us About the Effects of AI?" *Federal Reserve Bank of Richmond Economic Brief* (25-16).

Hassan, Tarek A., Stephan Hollander, Aakash Kalyani, Laurence van Lent, Markus Schwedeler, and Ahmed Tahoun. 2025. "Text as Data in Economic Analysis." *Journal of Economic Perspectives* 39 (3):193–220.

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." *The Review of Financial Studies* 36 (12):4919–4964.

Hassan, Tarek A, Stephan Hollander, Laurence Van Lent, and Ahmed Tahoun. 2019. "Firm-Level Political Risk: Measurement and Effects." *The Quarterly Journal of Economics* 134 (4):2135–2202.

Hassan, Tarek A., Jesse Schreger, Markus Schwedeler, and Ahmed Tahoun. 2024. "Sources and Transmission of Country Risk." *Review of Economic Studies* 91 (4):2307–2346.

Hébert, Benjamin M and Jennifer La'O. 2023. "Information Acquisition, Efficiency, and Non-Fundamental Volatility." *Journal of Political Economy* 131 (10):2666–2723.

Huang, Allen H., Hui Wang, and Yi Yang. 2023. "FinBERT: A Large Language Model for Extracting Information from Financial Text." *Contemporary Accounting Research* 40 (2):806–841.

Huo, Zhen and Naoki Takayama. 2015. "Higher Order Beliefs, Confidence, and Business Cycles." Mimeo, University of Minnesota.

Kalyani, Aakash, Nicholas Bloom, Marcela Carvalho, Tarek Hassan, Josh Lerner, and Ahmed Tahoun. 2025. "The Diffusion of New Technologies." *The Quarterly Journal of Economics* 140 (2):1299–1365.

Lian, Chen. 2021. "A theory of narrow thinking." *The Review of Economic Studies* 88 (5):2344–2374.

Lorenzoni, Guido. 2009. "A Theory of Demand Shocks." *American Economic Review* 99 (5):2050–84.

Loughran, Tim and Bill Mcdonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *The Journal of Finance* 66 (1):35–65.

Loughran, Tim and Bill McDonald. 2020. "Textual Analysis in Finance." *Annual Review of Financial Economics* 12 (1):357–375.

Maslej, Nestor, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, Tobi Walsh, Armin Hamrah, Lapo Santarlasci, Julia Betts Lotufo, Alexandra Rome, Andrew Shi, and Sukrut Oak. 2025. "The AI Index 2025 Annual Report." Tech. rep., AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA. URL https://aiindex.stanford.edu/.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." *arXiv preprint arXiv:1301.3781* .

Morris, Stephen and Hyun Song Shin. 2002. "Social Value of Public Information." *American Economic Review* 92 (5):1521–1534.

Nimark, Kristoffer P. 2014. "Man-Bites-Dog Business Cycles." *American Economic Review* 104 (8):2320–67.

Noy, Shakked and Whitney Zhang. 2023. "Experimental evidence on the productivity effects of generative artificial intelligence." *Science* 381 (6654):187–192.

Sautner, Zacharias, Laurence Van Lent, Grigory Vilkov, and Ruishen Zhang. 2023. "Firm-Level Climate Change Exposure." *The Journal of Finance* 78 (3):1449–1498.

Yotzov, Ivan, Jose Maria Barrero, Nicholas Bloom, Philip Bunn, Steven J Davis, Kevin M Foster, Aaron Jalca, Brent H Meyer, Paul Mizen, Michael A Navarrete, Pawel Smietanka, Gregory Thwaites, and Ben Zhe Wang. 2026. "Firm Data on AI." Working Paper 34836, National Bureau of Economic Research.

A. Data Appendix

In this Appendix, we provide details on our data construction process.

A.1 Processing Earnings Calls

We systematically extract sentences related to Artificial Intelligence (AI) using various machine learning approaches. The earnings calls cover all continents and are not restricted to specific industries. Extracting AI-related sentence is a high-precision, low prevalence task, as artificial intelligence i) is a relatively recent phenomenon, ii) makes up only about 0.1% of sentences and thus requires carefully curated training data, and iii) is semantically close to general tech-related language and therefore needs to be distinguished from non-AI "tech talk".

As a first step, we manually download the raw earnings calls from Refinitiv Workspace as $.txt$ files. For preprocessing we follow the approach outlined in Hassan et al. (2025), thus abstaining from otherwise common approaches such as stemming or lemmatization and preserving the sentences in their original form to maintain readability. We remove non-English transcripts and de-duplicate the remaining transcripts.[8] We remove the following sections from the transcripts: headers (date, time, and speaker section), footers (disclaimers, terms and conditions), speaker and section reference blocks that introduce the next spoken block, and any other standard components, so that we are left with spoken text. We split each transcript into sentences using Python's SpaCy library and number each sentence within a transcript. We use speaker names and section names to identify, for each sentence, whether it is part of the presentation or Q&A section, and remove any blocks spoken by analysts. For some earnings calls no speaker names could be identified. In such cases, we kept all text from the presentation section (where only corporate speakers should be present) and removed all blocks from the Q&A section that ended in inquisitive questions (removed with a simple Regex filtering for common question words). From the earnings call metadata and document title, we identify the date and time of the earnings call, and the RIC (firm identifier), which we use to match with firm-level data from Refinitiv Workspace, providing information about the country and industry of the firm.

A.1.1 Extracting AI-terms

The second step involves a binary classification task thus assign a label AI (1) or not-AI (0) to each sentences, added as a separate column to each sentence. By focusing only on explicit mentions in individual sentences, we miss any implicit or purely referential mention of AI (e.g."that", "this" in reference to some AI term in the preceding sentence) and provide a lower bound of total AI mentions by publicly listed firms.

Our definition of artificial intelligence is narrow, although including areas of applications. We thus include terminology around large language models, generative artificial intelligence, machine learning (ML), neural networks, and certain popular models or products, as well as some applications of autonomous products and vehicles, natural language processing (NLP), while excluding more general phenomena that are highly correlated but not necessarily integrated with AI. Such not explicitly extracted areas include robotics, automation, augmented or virtual reality (AR/VR), the development of chips, computing power, or other related products.

[8] For duplicated transcripts (downloaded multiple times), we remove the duplicated downloads. In all other cases, we remove the sections that were repeated multiple times during a day (e.g. multiple earnings calls within one day which carry the same presentation) to avoid double counting of sentences. Additionally, when aggregating to the firm-quarter level, we integrate all transcripts during a calendar year-quarter into one observation. Therefore the total number of firm-quarter observations is lower than the number of transcripts.

Regular Expressions. We start by identifying the most common acronyms (e.g. AI, ML, RNN), names (e.g. OpenAI, ChatGPT, DeepSeek) (both case-sensitive) and phrases (e.g. artificial intelligence, autonomous vehicle) (case-insensitive) to filter for AI-sentences. We expand our dictionary iteratively (also in connection with supervised machine learning), but ultimately decided against a keyword-based algorithm (as used, for example, to filter climate change terms, e.g. Sautner et al. (2023)) as certain terms (most prominently AI) are not uniquely identifying artificial intelligence (for example, Avian Influenza, Autoplan International).

Supervised Machine Learning. We follow Barrios et al. (2024) and fine-tune a FinBERT model (Araci, 2019; Huang, Wang, and Yang, 2023) to classify AI terms. We proceed in three stages using curriculum-style learning. In the first stage, we created a stratified sample of 3,000 sentences selected from the regular expressions, 1,500 sentences from the period after 2016, and 2,500 entirely random sentences. The first stage discriminates easily between easy hits and misses, but lacks nuance and grossly overpredicts AI-sentences. Thus, in a second stage we added new training data, selected iteratively from pre-2016 Regex matches and multiple second-stage validation sets, amounting to about 5,000 more sentences that provide a more nuanced overview of what sentences contain AI. We retrain the first stage model and assess performance. In both stages we use Optuna grid search to select the optimal model parameters, thugh in the second stage we freeze non-critical parameters. As the FinBERT model still shows ambiguity after retraining for certain sentence patterns, we add a third stage using a logistic classifier. The classifier takes the probabilities from the first and second models, year, sentence length, regex or high-confidence match, and industry as inputs and is trained on a validation set of 2,750 sentences to decide the weighting of each regressor. We apply the meta-classifier to all sentences which received a score between 0.3 in the first two stages and 0.9 in the second stage (since we did not identify high-confidence false positives), label all sentences above 0.9 as 1 (below 0.3 as 0), and label each sentence receiving a score above 0.5 by the meta-classifier as 1.

Table A1 details the earnings calls sample by year, split by all calls and calls with sentences labeled as AI, and broken into US and non-US. Table A2 summarizes the revenues (assets), employees, and other variables in the financial records of the firms in the earnings calls sample, split by AI and non-AI labeled earnings calls. Firms mentioning AI are among the larger firms within industry, and the gap between AI-mentioning and non-AI mentioning firms widens after the ChatGPT announcement.

Table A1: Earnings calls coverage by year and region (US vs Non-US).

	US				Non-US			
	RICs	Files	Files (AI)	RICs (AI)	RICs	Files	Files (AI)	RICs (AI)
2001	247	336	3	3	44	54	1	1
2002	2,126	5,491	28	20	392	907	5	5
2003	2,603	9,593	53	32	974	2,808	6	5
2004	2,777	11,017	56	39	1,091	3,790	7	5
2005	3,038	11,607	46	36	1,337	3,866	11	9
2006	3,186	11,911	42	28	1,586	5,195	13	10
2007	3,326	12,051	38	33	1,798	5,136	10	9
2008	3,436	12,460	48	34	1,975	6,375	24	19
2009	3,307	12,113	33	25	1,906	6,332	18	12
2010	3,241	11,279	45	37	2,014	6,371	32	20
2011	3,234	11,695	71	52	2,193	6,575	30	23
2012	3,231	11,636	78	57	2,487	7,076	40	25
2013	2,906	10,394	104	77	2,075	6,565	42	31
2014	3,188	11,325	168	121	2,323	6,868	63	49
2015	3,262	11,430	235	150	2,290	6,523	118	84
2016	3,079	10,230	414	265	2,362	6,829	258	163
2017	3,169	11,399	946	519	3,126	8,372	597	392
2018	3,166	11,828	1,273	651	3,634	11,136	1,159	692
2019	3,166	11,830	1,388	737	3,838	12,055	1,273	773
2020	3,194	11,923	1,317	702	4,119	12,699	1,247	780
2021	3,475	12,591	1,696	886	4,121	10,869	1,065	761
2022	3,481	13,012	1,714	903	4,525	13,366	1,347	864
2023	3,345	12,296	2,662	1,222	4,756	13,892	2,313	1,305
2024	3,203	11,652	3,473	1,471	4,977	14,197	3,120	1,693
2025	3,247	11,680	4,558	1,881	5,447	14,807	4,246	2,369
Total	**7,720**	**272,779**	**20,489**	**3,122**	**9,314**	**192,663**	**17,045**	**3,924**

Table A2: Industry distributions of firm characteristics, stratified by AI mentions and period.

			Pre-ChatGPT		Post-ChatGPT	
			AI	Non-AI	AI	Non-AI
Communication Services	Firms	n	353	291	460	91
	Revenue / Assets	median	0.13	0.13	0.13	0.14
		SD	0.12	0.12	0.15	0.14
	Employees (FTE)	median	4,988	2,475	4,904	608
		SD	61,080	9,041	49,086	10,854
	ROA (%)	median	1.29	1.51	1.67	1.72
		SD	5.83	3.53	11.99	8.99
	ROE (%)	median	6.87	4.84	7.01	4.39
		SD	38.04	56.66	40.74	15.71
Consumer Discretionary	Firms	n	457	781	679	596
	Revenue / Assets	median	0.24	0.25	0.24	0.25
		SD	0.19	0.19	0.17	0.19
	Employees (FTE)	median	16,101	3,962	6,790	4,251
		SD	115,457	32,350	172,320	11,300
	ROA (%)	median	2.02	1.54	2.23	1.93
		SD	4.36	4.55	6.50	5.30
	ROE (%)	median	9.37	8.62	9.75	6.97
		SD	20.50	21.56	24.77	29.78
Financials	Firms	n	494	931	832	603
	Revenue / Assets	median	0.07	0.04	0.06	0.03
		SD	0.09	0.13	0.12	0.08
	Employees (FTE)	median	13,904	2,309	9,093	2,314
		SD	60,691	31,103	60,317	15,569
	ROA (%)	median	1.07	1.11	1.14	1.10
		SD	4.08	3.95	6.20	4.39
	ROE (%)	median	11.08	9.68	12.28	10.44
		SD	19.15	11.19	18.02	23.41
Health Care	Firms	n	425	1,107	618	785
	Revenue / Assets	median	0.12	0.09	0.14	0.11
		SD	0.14	27.60	0.16	47.46
	Employees (FTE)	median	1,169	154	902	202
		SD	42,170	10,271	29,272	9,746
	ROA (%)	median	1.63	-2.38	1.19	-4.50
		SD	11.45	15.89	43.33	32.00
	ROE (%)	median	5.36	-2.74	3.15	-5.25
		SD	65.89	4378.43	49.20	58.55
Industrials	Firms	n	554	939	887	755
	Revenue / Assets	median	0.20	0.20	0.21	0.20
		SD	0.19	0.14	0.15	0.16
	Employees (FTE)	median	16,170	4,800	14,700	4,274
		SD	75,904	23,745	58,590	19,240
	ROA (%)	median	1.76	1.41	1.94	1.56
		SD	13.03	5.09	8.48	5.91
	ROE (%)	median	10.41	7.42	10.78	7.40
		SD	17.77	33.38	140.81	37.18
Information Technology	Firms	n	871	378	1,024	154
	Revenue / Assets	median	0.17	0.18	0.17	0.19
		SD	0.12	0.12	0.12	0.13
	Employees (FTE)	median	2,755	758	2,326	688
		SD	51,002	14,965	62,224	7,075
	ROA (%)	median	1.69	1.33	2.07	1.27
		SD	10.90	11.08	10.14	11.73
	ROE (%)	median	6.12	4.57	8.20	4.53
		SD	30.31	18.95	31.28	35.18

A.1.2 Classifying Sentiment

To classify sentiment, we rely on two standard sources in the finance literature. First, we apply the Loughran-McDonald dictionary (Loughran and Mcdonald, 2011; Loughran and McDonald, 2020). The dictionary scans each sentence for whether it contains words signaling positive or negative sentiment. We classify a sentence as positive, if it contains more positive than negative words. Sentences containing neither positive nor negative words are labeled as neutral. Beyond the positive and negative sentiment categories, we also use the uncertainty category which follows the same logic as classifying sentiment. As a second sentiment measure we rely on FinBERT, a LLM trained on financial text to classify sentences into positive or negative sentiment. We additionally record the certainty of the FinBERT model for the assigned sentiment category. Sentiment is added as a separate column per category (encoding positive, negative and uncertainty) to each sentence in the dataset.

A.1.3 Extracting Properties of AI-sentences

In a fourth step, we extract certain properties of AI sentences. For this, we first apply FinBERT-FLS[9] to classify sentences as specific, unspecific or not forward-looking. We then categorize each sentence according to action-, performance- or efficiency-mentions, and extract time labels for all sentences. For time labels, we start with an exogenous classifier based on date or relative time mentions in transcript, which can be compared directly to the date of the earnings call and do not require further interpretation (about 10% of sentences). The chosen time labels are past, present, short-run, long-run. For action, efficiency, and performance, we group sentences into positive and negative mentions. Each sentence receives a separate column per category.

Unsupervised Machine Learning. Owing to the granularity of our classification, we proceed using OpenAI's ChatGPT API to classify the sentences according to the categories we specified. We formulate a separate prompt for time labels (some sentences have been classified ex ante), and a second multi-task prompt for the remaining categories which is fed to all sentences. We use zero-shot classification, and find the best performance using GPT-4o[10].

Each sentence receives a separate column per category. The labels are as follows:

1. Time: past (more than a quarter in the past), present (between last and next quarter), short-term (between next quarter and 3 years), long-term (more than three years in the future, or indefinite)
2. Efficiency: 1 (positive impact on efficiency), 0 (no mention of efficiency), -1 (explicit no or negative effect on efficiency)
3. Performance: 1 (positive impact on performance), 0 (no mention of performance), -1 (explicit no or negative effect on performance)
4. Action: 1 (positive action), 0 (no mention of action), -1 (explicit no or negative action)
5. Relation: none, customers (users), partners (suppliers, collaborators), competition (rivals, strategic position), market (trends, developments, impact)

Time considers the time horizon considered in relation to AI in each AI-related sentence and can range from the past (if a firm mentions that it launched products in the past or has been working on something for years), present (current operations, products, or market situation) to the long-term future (e.g., anticipated impact of AI). For performance, efficiency, and action in each only positive,

[9] FLS stands for forward-looking sentences

[10] We use GPT-4o-2024-11-27, with temperature 0.0, logprobs set to true. We could have also used ChatGPT-5 but it does not allow the extraction of logprobs

negative, or non-mentions are considered. If a firm mentions specifically that AI does not affect them or that they cannot act, this is perceived as a negative instance. Furthermore, performance and efficiency are distinct in that performance focuses exclusively on financial or business metrics (e.g., revenues, costs, profits) though these may be implicit (e.g., demand, growth), whereas efficiency concentrates on operational improvements and value gains (e.g., customer performance gains, process efficiency, product quality improvements, AI speeding up processes), which are not directly linked to company performance. The category relation captures external relationships mentioned in the sentence. Customers may be explicitly named or generically stated (clients, certain interest groups), which applies to partners and competition as well. Partners are understood vertically and horizontally. They refer both suppliers or providers of the concerned company (which includes the firm itself using products by someone else), as well as to joint developers or partners. We grouped vertical and horizontal partners because i) both classes were relatively rare and ii) the decision boundary between both was fuzzy, as most relationships are bidirectional. Competition may refer both to mentioning other firms without confirming a certain (customer/partner) relationship, or by them referring to their own position in the market (being ahead or behind, catching up, leadership, etc.). Market captures trends and global developments surrounding AI that are not firm-specific. Such developments include how AI reshapes certain sectors or areas, which features or tools will be important, or how total market size develops (or how companies estimate it). However, market does not refer to company-only statements (e.g., firm-specific growth or profit opportunities, or impact of AI on the company). Last, all categories are nonexclusive (i.e. a sentence may refer to both performance, action, efficiency, relation, and time), although the labels within a category are mutually exclusive.

The structure of our prompts can be summarized as follows:[11]

- We begin each prompt with instruction of the task. This description is split in two parts. The first parts instructs ChatGPT to act as a classifier for financial text. The second part concerns the mode (either central sentence-only, or with context), and states the conditions for considering context.

- For time we add a line towards the beginning that refers to the *date* column in the data, and for relation we add a reference to the *RIC* (Reuters Information Code, firm identifier) column to provide the LLM with sensible reference points when determining the distance of dates (for those not captured by the exogenous classifier, i.e., in utilized context) and presence of an external relationship[12].

- In each prompt, we specify the exact label space and a short definition of category to prime the LLM on what features it should look for. That definition amounts to one to two lines. We then provide definitions of the labels within the category, each of about one to four lines depending on the granularity and attention to ambiguity required. Further, in each prompt, we also give a definition of the null label such that the LLM is forced to provide some class for each sentence and does not respond with empty output.

- After label definitions, we add further rules for label assignment, such as treatment of conflicting positive and negative mentions, or occurrences of different labels within a category for each sentence.

- As a last component for the few-shot classification, we add examples for the different labels

[11] The exact prompts are available upon request

[12] This feature is relevant for cases where firm names may be used self-referentially, e.g., an NVIDIA executive stating *Enterprises are also racing to deploy generative AI, driving strong consumption of NVIDIA-powered instances in the cloud as well as demand for on-premise infrastructure.*, which could otherwise be misinterpreted as a statement about competition

(which are representative of common sentences or represent certain complicated cases that were systematically misclassified in previous runs)

- An exception is the *action* category. Here we found best performance by only providing a short definition of action and very limited definitions of the ternary labels, and removed any examples or decision rules.

We follow a hierarchical architecture in which we first subject all sentences alone to ChatGPT (without context), and in a second step pass all sentences with their preceding and succeeding sentences. We instruct ChatGPT to consider context for classification only if it is directly relevant to the central sentence and accordingly administer the appropriate label for the entire triplet.[13] The reason is that when passing the prompt to the entire triplet of sentences, ChatGPT had challenges to correctly identifying relevant context patterns and suffered from context leakage, where labels were sometimes applied wrongly. We therefore keep the results from the first stage for two reasons. First, as an indication where context altered the label, and second as a fallback option when the context pass accidentally downgrades a nonzero label.[14] The approach prevents information from being duplicated or noisy signals being introduced by irrelevant adjacent sentences. Hence, even when two neighboring AI sentences share the same triplet, labels of both sentences are only applied if they apply to exactly the same topic.

Hardware. We carry out the entire classification pipeline using Python 3.13 on Windows. Specifically, the scripts are run locally on a DELL Precision 7960 Tower Workstation at The University of Texas at Austin with a Intel(R) Xeon(R) w9-3595X processor on a NVIDIA RTX 6000 Ada Generation 48 GPU with 48 GB RAM. We keep the system constant for all processing steps.

Running generative large language models (LLMs) which hold multiple billions of parameters is computationally expensive, that even with a powerful GPU computing power is too constrained. For the robustness checks we hence make use of the Texas Advanced Computing Cluster's (TACC) Lonestar6 computer, which hosts 84 A100 GPU nodes with the addition of 3 NVIDIA A100 GPUs. Each A100 GPU has a peak performance of 9.7 TFlops in double precision and 312 TFlops in FP16 precision using the Tensor Cores. Additionally, there are 4 H100 GPU nodes that support 2 NVIDIA H100 GPUs. Each H100 GPU has a peak performance of 26 TFlops in double precision and 1513 TFlops in FP16 precision using the Tensor cores. We pass the serialized script to Lonestar6 and let the machine handle allocation of GPU nodes and computing resources and therefore run all jobs are on an A100 GPU node on a single core split into 4 shards to distribute workload.

A.1.4 Validation

To validate our results, we plot the total AI mentions extracted per quarter over time and compare them to findings of other papers. Given the evolution of the field of machine learning, we expect nearly no mentions in earnings calls before 2016, and a strong increase after the ChatGPT release, which are both visible in Figure A1. As the number of sentences per quarter changes only very little over time, the same pattern holds when viewing the share of AI sentences per total sentences each quarter.[15] Comparing the count of AI sentences to those covered by the selected keywords in Hassan et al. (2025) also shows strong alignment of our method. Their method, based on keyword extraction

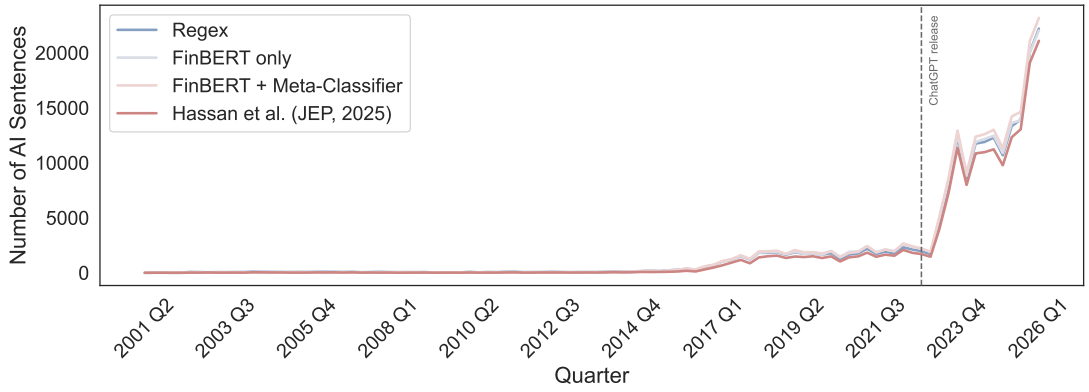
[13] Relevant context is defined by elliptical sentences, named entity or noun repetition, implicit mentions or reference to the next or previous sentence, or logical continuation (e.g., indicated by connector phrases).

[14] The reasoning behind this is that if a non-zero label is detected in the first stage, this should remain valid or may be superseded only by another non-zero label.

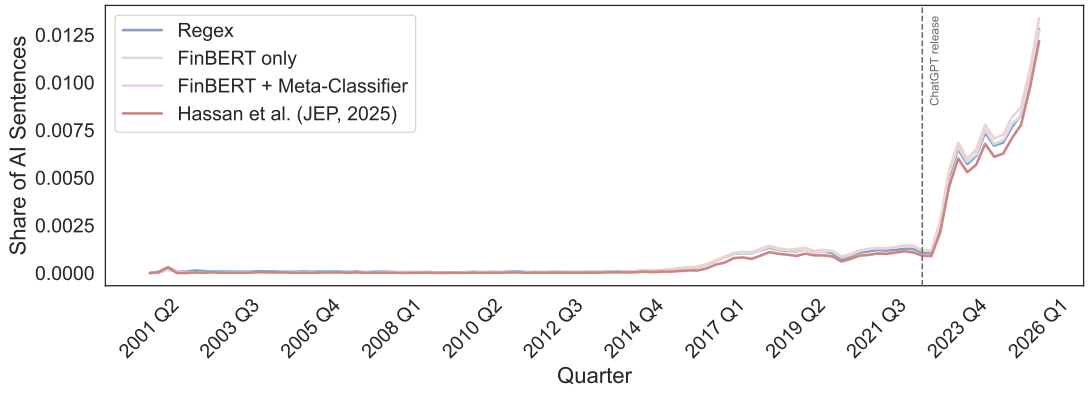
[15] The only exception is in 2001, where there is relatively little data and some AI mentions, primarily from Time Warner about their movie *A.I.*, leading to a bump in the time series when looking at the share per sentence

and pure lexical analysis expectedly yields slightly higher counts in the pre-2016 period because their measure cannot differentiate between different uses of the word *AI*, and slightly lower counts after 2016 as their selected keywords do not cover the entire breadth of AI-relevant content. In general, however, both measures are quite close and follow the exact same trends, which makes us confident in our algorithm.

Another check is looking at the distribution of sentences across industries over time. We only look at the post-2016 period as before that time the pot is uninformative given the high volatility and sparsity of AI-sentences. Figure A2 shows another feature of the data which may be expected: AI sentences occur mostly in the Information Technology (IT) sector, and to a lesser degree in Communication Services and Industrials, whereas only very few mentions occur in distant sectors such as Energy, Materials, or Utilities. AS the IT sector is the primary sector concerned with developing AI, we would expected the most mentions being concentrated there. Similarly, in communications, finance, health care, and industrials (mostly manufacturing), applications of AI have become increasingly frequent which explains their exposure, whereas in less technology-exposed industries the share of mentions is much lower. Over the past eight years, this allocation between sectors has remained largely constant, suggesting the most growth in AI was driven by importance of the technology in general, and not by shifting the sectoral composition of AI exposure.



(a) Count of AI Sentences



(b) AI Sentences per Total Sentences

Figure A1: Quarterly AI Mentions Over Time

As a last validation step, we print word clouds of the most frequent unigrams and bigrams in the entire sample, as well as before and after ChatGPT. In Figure A3 one can see that after lemmatizing and removing stopwords, words commonly associated with artificial intelligence such as AI, data,

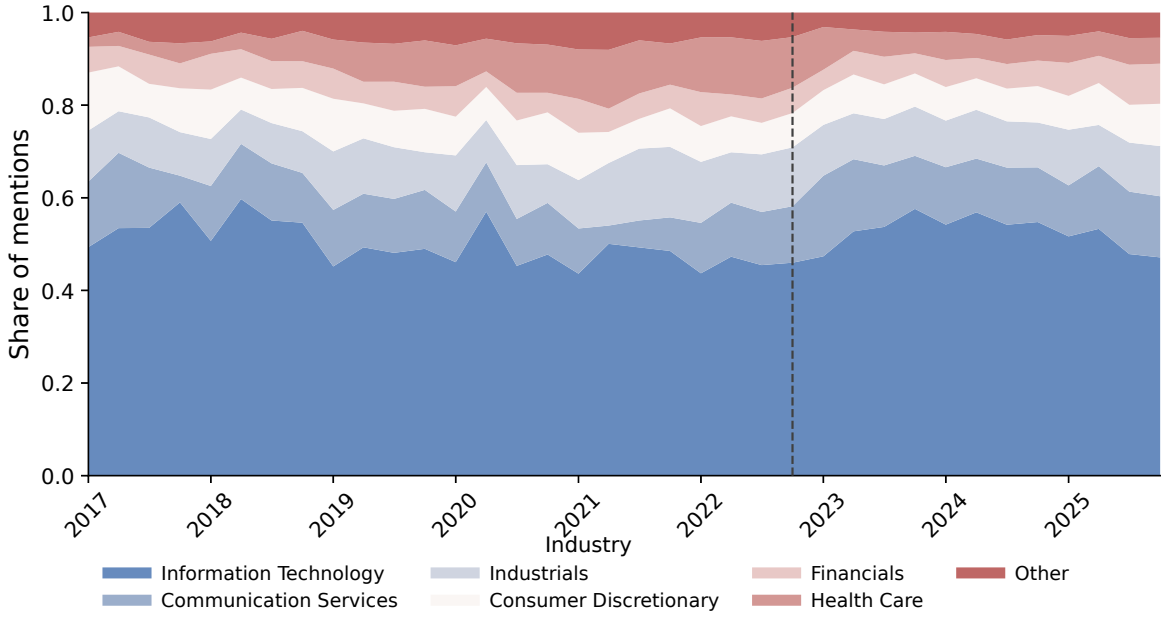
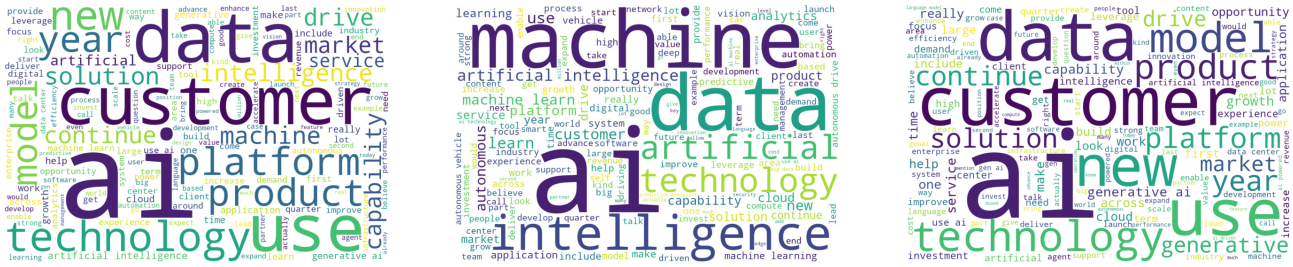


Figure A2: Sectoral Composition of AI Mentions Over Time

intelligence, technology, machine, or platform are most frequent. Before ChatGPT the focus seems more on the technical aspect whereas after ChatGPT commercialization and application of products became more prominent. Overall, the word clouds suggest that our sentences are properly capturing talk about AI and do not pick up noisy signals.



(a) Full data (b) Pre-ChatGPT (c) Post-ChatGPT

Figure A3: Word Cloud of Most Frequent Words in AI-Sentences

Notes: Wordclouds include lemmatized unigrams and bigrams. English stopwords have been removed (NLTK) as well as common colloquial words. Pre-ChatGPT period is Q1 2017–Q3 2022, Post-ChatGPT is Q4 2022–Q4 2025

A.1.5 Robustness of ChatGPT Classification

Additionally to the concern of stability of labels between the S1 and context stages, we address concerns of model and prompt stability by repeating our classification (time, performance, action, efficiency, relation) using different models and prompt architectures following the approach adopted in Clayton et al. (2025). As ChatGPT is a closed and proprietary model and not freely accessible at scale, we use our main classification as a benchmark and compare it against the classification by other state-of-the-art large language models. Although ChatGPT may be assumed to be superior in terms of performance, the lack of transparency and accessibility and present obvious concerns about

replicability. This approach also allows us to evaluate how much of our classification depends on the model and the final prompts chosen.

We carry out the classification using Meta's `Llama-3.1-8B-Instruct` and Alibaba's `Qwen-2.5-7B-Instruct`. All models are open-source and accessible via HuggingFace. Additionally, we specify alternative (system) prompts to be fed to the large language model to assess the reliability of its output (and stability of classification) across different semantic or sentence structures (while not changing the underlying scope and meaning of the prompt). We apply prompt perturbation in the following three ways:

1. Small perturbation: change single sentences or words from the canonical prompt
2. Medium perturbation: change wording and language but keep structure similar
3. Large perturbation: reduce prompt content to minimally informative instruction

Given that our prompts are i) more targeted at a single class and thus leave less room for fundamental changes, we choose $K = l + 3$ (split l small, 2 medium, 1 large, with l being the number of available labels inside a class) instead of the $K = 20$ as used by Clayton et al. (2025).[16] Small perturbations are carried out for $l - 1$ labels (excluding the label none) by changing the description of a single label (without altering its meaning), and additionally by reducing examples to hold only a single one per label.[17] We apply medium perturbations in two ways. First, we change the wording and language of the labels slightly, and alter the examples (as well as their composition); and second we keep the prompt as in the canonical form but strip all examples. For reference, we are only exchanging the content-related part of the prompt (we thus keep structural elements such as the consideration of neighboring sentences and label choice constant; the prompt content below may thus be inserted into both the S1 (central sentence only) and context (CTX, central sentence plus preceding and succeeding sentence) version of the respective prompt.

There are three main differences to their approach. First, our approach differs in that we selected ChatGPT as our baseline model. Since running the same robustness analysis for the ChatGPT-produced labels is cost-prohibitive at scale, we do not assess consistency within-ChatGPT but only compare the ChatGPT labels to labels produced by the other models.[18] Second, since our data is on the sentence-level and only covering a small subset of the totality of earnings calls, we carry out the robustness checks for all 207,241 sentences (whereas Clayton et al. (2025) only rely on a subset of documents). Thus, despite the smaller number of applied perturbations, our methodology is still representative for the whole corpus.[19] Third, given our structure of feeding each sentence to ChatGPT twice and determining the final label afterwards, we follow the same structure for the reliability analysis. Hence, each sentence is passed to each LLM with a S1-only and context prompt version (where the content part is exactly the same) and consolidated afterwards following similar rules. Accordingly, we carry out the robustness analysis for each the S1, context, and final label.

For each sentence i, semantic field $d \in \{\text{time}, \text{performance}, \text{efficiency}, \text{action}, \text{relation}\}$, and label type $\ell \in \{\text{S1}, \text{CTX}, \text{Final}\}$, let $Y_{id}(p_k, m)$ denote the categorical label returned by model m under prompt variant p_k. Because all models are queried at zero temperature, the mapping $(p_k, m) \mapsto Y_{id}(p_k, m)$ is

[16] As action, performance, and efficiency are ternary labels, $l = 3$ for these three categories. For time and relation each five labels exist, thus $l = 5$ for these categories.

[17] The only exception is the "action" prompt where instead of removing examples we introduce new ones as the canonical prompt does not carry examples

[18] Internal note: It would still be possible to run the prompt perturbation through ChatGPT with a subset of sentences to get an overview of how whether consistency also holds there. I would wait with that until everything else is finalized or referees/comments push for that step.

[19] For the time-head 192,758 only are evaluated because 14,483 sentences receive a time label based on an exogenous classifier.

deterministic. We describe the observed label as

$$Y_{id}(p_k, m) = \eta_{id} + \varepsilon_{idkm},$$

where η_{id} captures the semantic signal for sentence i in field d, and ε_{idkm} captures instability due to changes in prompt specification or model architecture.

We consider two forms of reliability. The first measures the robustness of a given model m to prompt perturbations within that model as reported in Clayton et al. (2025). The second measures the robustness of the baseline prompt to changes in model architecture which assesses how much the classification of the baseline prompt hinges on choosing a specific model. Considering within-model prompt reliability, for each model m, the semantic signal is anchored on that model's own baseline prompt p_1. The signal variance for field d and label type ℓ is defined as $\hat{\sigma}^2_{\eta,d,\ell}(m) = \mathrm{Var}_i\big(Y_{id,\ell}(p_1, m)\big)$, capturing the cross-sentence variation in the model's baseline labeling function. To measure the noise introduced by prompt perturbations, we compute for each sentence i the empirical variance across the available prompt variants $p_k \in \mathcal{K}_{d,m}$: $\widehat{\mathrm{Var}}_{k\in\mathcal{K}_{d,m}}\big(Y_{id,\ell}(p_k, m)\big)$, and average this within-sentence variance across all sentences with at least two observed prompt labels: $\hat{\sigma}^2_{\varepsilon,d,\ell}(m) = \frac{1}{N_{d,m}} \sum_{i=1}^{N_{d,m}} \widehat{\mathrm{Var}}_{k\in\mathcal{K}_{d,m}}\big(Y_{id,\ell}(p_k, m)\big)$. The within-model prompt reliability is then

$$\widehat{R}^{\text{prompt}}_{d,\ell}(m) = \frac{\hat{\sigma}^2_{\eta,d,\ell}(m)}{\hat{\sigma}^2_{\eta,d,\ell}(m) + \hat{\sigma}^2_{\varepsilon,d,\ell}(m)}.$$

$\widehat{R}^{\text{prompt}}_{d,\ell}(m)$ thus measures the extent to which variation in $Y_{id,\ell}(p_k, m)$ reflects true semantic heterogeneity across sentences rather than sensitivity to prompt wording within model m.

To evaluate how robust the baseline prompt p_1 is to changes in model architecture, we treat the ChatGPT baseline labels $Y_{id,\ell}(p_1, m_{\text{ChatGPT}})$ as the reference labeling function. The signal variance is therefore $\hat{\sigma}^2_{\eta,d,\ell} = \mathrm{Var}_i\big(Y_{id,\ell}(p_1, m_{\text{ChatGPT}})\big)$. For noise variance, we hold the prompt fixed at p_1 and compute, for each sentence i, the variance across baseline labels returned by the different models $m \in \mathcal{M}$: $\widehat{\mathrm{Var}}_{m\in\mathcal{M}}\big(Y_{id,\ell}(p_1, m)\big)$, and then average these within-sentence variances: $\hat{\sigma}_{\varepsilon,d,\ell^{2,\text{model}}} = \frac{1}{N_d} \sum_{i=1}^{N_d} \widehat{\mathrm{Var}}_{m\in\mathcal{M}}\big(Y_{id,\ell}(p_1, m)\big)$. The cross-model reliability for the baseline prompt is therefore

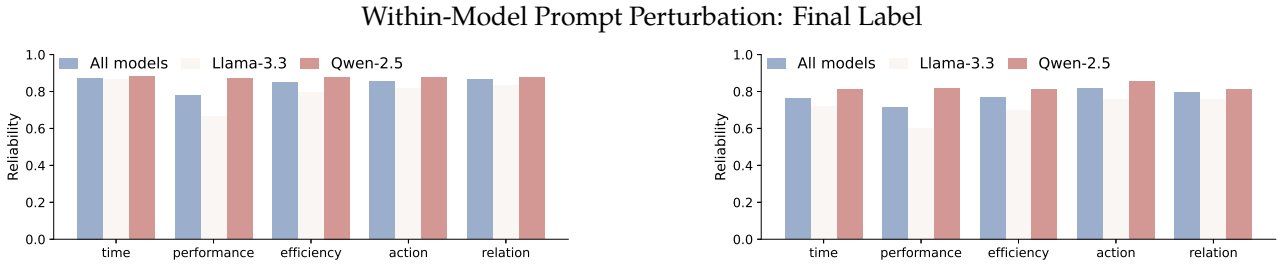
$$\widehat{R}^{\text{model}}_{d,\ell} = \frac{\hat{\sigma}^2_{\eta,d,\ell}}{\hat{\sigma}^2_{\eta,d,\ell} + \hat{\sigma}^{2,\text{model}}_{\varepsilon,d,\ell}}.$$

Therefore, $\widehat{R}^{\text{model}}_{d,\ell}$ quantifies the stability of the baseline prompt to model choice: a value close to one indicates that the different models produce baseline labels that vary little relative to the semantic variation across sentences, whereas a value close to zero indicates substantial model-induced noise. Since the labels are categorical, all computations operate on a unified integer encoding of the observed categories for each field d and label type ℓ. This ensures that the variance decomposition captures disagreement across prompt and model perturbations relative to the underlying between-sentence semantic structure.

To account for our differences in methodology, we additionally report Krippendorff's α and Cohen's κ to capture labeling disagreement (treating different prompt instructions and LLMs as individual labelers) for categorical data, as well as cross-model agreement on our baseline ChatGPT results.

Evaluation. Prompt perturbations do not seem to affect classification too much. Figure A4 shows that for small perturbations within-model, reliability scores are upwards of 0.8 for all categories other than performance, which sits slightly lower around 0.65 in some cases. This suggests that changing

the prompts slightly within a model does not obscure classification significantly and thus results are unlikely to be driven by wording choices inside the prompts. Only performance reacts slightly more, which is likely due to the missing differentiation with efficiency when slightly adapting the prompt (thus not only focusing on financial performance). Similarly, in the case of all perturbations the level of the reliability scores is slightly lower (which is expected given that more information is removed and details are different in the prompt). However, reliability scores are still above 0.6 and hovering around 0.8 in all cases, therefore suggesting that even when altering many prompt details (or infusing a minimal prompt with the same meaning), the model captures a similar structure and thus the results are robust to the mere details inside the prompt.



(a) Small Perturbations

(b) All Perturbations

Figure A4: Reliability Score of Prompt Perturbations

Notes: Within-Model Prompt Perturbation take baseline prompt inside the model as true variance and use perturbations as source of measurement error. Small perturbations refer to perturbations that only change a small portion of the prompt, all perturbations include medium and strong perturbation. All models compares across Llama-3.3 and Qwen-2.5.

The robustness statistics compiled in Table A3 further suggest that the model broadly supplies reliable outputs. First, the raw agreement (i.e. sentences for which the chosen label matches) is above 0.8 for the strong ChatGPT perturbation in all cases, with the time label being lower around 0.6 in all cases, given that the time label leaves the most space for ambiguity. This is seconded by high α, κ, and V scores for all models, showing a high degree of association and labeler agreement such that one may presume that by significantly altering the prompt (while reserving its meaning), a high degree of information is retained and label assignment is not primarily driven by the prompt itself.

When comparing instead the results for the Llama-3.3 and Qwen-2.5 models using the baseline prompts, significant differences emerge. Although neither one comes close to the ChatGPT perturbation quality, Qwen performs much better than Llama, with higher α, κ, V and Agreement scores across all stages and classes. Lower performance is expected since the training data and process matters (i.e. models are unlikely to report the exact same results), and ChatGPT being more powerful than either Qwen or Llama, given that it includes about 200 billion parameters. Since Qwen is more recall- and Llama more precision-oriented, higher agreement with Qwen is the consequence since it is less likely to report a missing or zero label than Llama. However, even within models there are differences between lasses. For Llama it stands out that for the S1 and final label for performance and efficiency, Krippendorff's α is negative, suggestive systematic labeler disagreement, which is seconded by low Cohen's κ. For CTX values are slightly higher yet still lower than for the other classes. Similarly, time and relation show slightly lower agreement across all statistics since there are more categories and thus a higher degree of uncertainty in contrast to the ternary labels of action, efficiency, and performance. While this pattern clearly holds for Qwen, it is less evident for Llama, which seems to do relatively well on time labels but shows issues around the assignment of performance and efficiency labels.

Limitations. The robustness checks conducted here only provide an estimate of internal consistency. Assessing how well the classification holds up under different models and prompts signals how much measurement varies across these different approaches, but does not capture alternative sources of measurement error that concern the overall measure (external validity).

First, selection is a probable issue. Given that firms voluntarily choose whether to hold a call (in contrast to quarterly earnings disclosures (Forms 10-K and 10-Q) there is no obligation to hold an earnings call), and have some discretion about which questions from analysts they allow and how they answer them, strategic motivation plays a key role. Managers further have an incentive to overplay the impact of AI (especially for producers) as for them the technology is a key segment and markets react strongly to their projected future prospects. Such strategic reporting choices constitute a source of measurement error that we cannot address by prompt- or model-perturbations.

Second, with AI being a significant driver of technological innovation that has evolved into a prominent trend over the past years, executives have incentives to jump onto the bandwagon and exhibit novelty and recency bias. With AI becoming more widespread, they are more likely to comment on new technologies and the most recent events. It might also be that new breakthroughs or initial adoptions are discussed, but firms do not mention AI again (as users) when there is no novel development (or no tangible impact) in the recent past. Thus, for users of AI, it is possible that despite ongoing use and adoption of AI, the observed mentions of AI in earnings calls are single-peaked declining after an initial spike despite the company retaining (or even increasing) their use of AI.

Third, we cannot directly assess the factual accuracy of statements in earnings calls. This means we cannot externally verify whether e.g., executives citing performance gains are describing them correctly, or whether the cited implementations or customer applications deliver exactly the described value. Additionally, we cannot evaluate whether forward-looking statements conform to ex post materialization or whether executives made up forecasts or expectations ad hoc. We are aware of the strategic element in earnings calls disclosures and how executives respond being observed by competitors and analysts for the statements they make, and that they may be framed by analyst questions (see e.g., Hollander, Pronk, and Roelofsen (2010); Cao et al. (2023); Allee, Do, and Sterin (2021)). The data construction takes no stance as to whether companies are able to correctly infer how AI develops or how it will impact their business, and solely focus on whether LLMs are capable of detecting observed AI-related sentences together with their stated features. The theoretical model in Section 4 provides a model of firm expectations about the technology and their evolution over time.

Label	Head	Model	α	κ	V	Agree
		ChatGPT (strong perturbation)	0.414	0.449	0.597	0.612
	time	Llama-3.3	0.279	0.295	0.321	0.650
		Qwen-2.5	0.395	0.399	0.429	0.667
		ChatGPT (strong perturbation)	0.747	0.749	0.789	0.941
	performance	Llama-3.3	-0.374	0.062	0.339	0.302
		Qwen-2.5	0.586	0.590	0.610	0.865
		ChatGPT (strong perturbation)	0.780	0.780	0.706	0.922
S1	efficiency	Llama-3.3	-0.285	0.082	0.231	0.358
		Qwen-2.5	0.703	0.703	0.626	0.887
		ChatGPT (strong perturbation)	0.814	0.815	0.799	0.912
	action	Llama-3.3	0.364	0.401	0.341	0.732
		Qwen-2.5	0.693	0.693	0.564	0.850
		ChatGPT (strong perturbation)	0.828	0.828	0.814	0.874
	relation	Llama-3.3	0.161	0.245	0.397	0.418
		Qwen-2.5	0.632	0.635	0.642	0.739
		ChatGPT (strong perturbation)	0.418	0.450	0.611	0.628
	time	Llama-3.3	0.251	0.271	0.331	0.699
		Qwen-2.5	0.210	0.226	0.322	0.632
		ChatGPT (strong perturbation)	0.734	0.736	0.779	0.933
	performance	Llama-3.3	0.056	0.199	0.404	0.547
		Qwen-2.5	0.443	0.452	0.468	0.799
		ChatGPT (strong perturbation)	0.787	0.787	0.604	0.907
CTX	efficiency	Llama-3.3	0.235	0.299	0.302	0.615
		Qwen-2.5	0.517	0.520	0.412	0.796
		ChatGPT (strong perturbation)	0.793	0.795	0.659	0.903
	action	Llama-3.3	0.403	0.427	0.341	0.740
		Qwen-2.5	0.528	0.529	0.381	0.768
		ChatGPT (strong perturbation)	0.793	0.794	0.776	0.858
	relation	Llama-3.3	0.215	0.265	0.362	0.439
		Qwen-2.5	0.524	0.527	0.540	0.675
		ChatGPT (strong perturbation)	0.477	0.499	0.642	0.667
	time	Llama-3.3	0.228	0.243	0.311	0.661
		Qwen-2.5	0.229	0.243	0.330	0.631
		ChatGPT (strong perturbation)	0.735	0.737	0.769	0.923
	performance	Llama-3.3	-0.333	0.069	0.380	0.323
		Qwen-2.5	0.480	0.491	0.539	0.792
		ChatGPT (strong perturbation)	0.782	0.783	0.636	0.902
Final	efficiency	Llama-3.3	-0.168	0.104	0.273	0.442
		Qwen-2.5	0.606	0.606	0.505	0.821
		ChatGPT (strong perturbation)	0.796	0.798	0.761	0.911
	action	Llama-3.3	0.334	0.373	0.338	0.750
		Qwen-2.5	0.633	0.633	0.510	0.832
		ChatGPT (strong perturbation)	0.805	0.805	0.792	0.855
	relation	Llama-3.3	0.156	0.226	0.342	0.402
		Qwen-2.5	0.544	0.548	0.556	0.670

Table A3: Robustness Statistics of Model Perturbations

Notes: ChatGPT classifications undertaken using GPT-4o-2024-11-20. Llama-3.3 refers to the 8B version, Qwen-2.5 to the 7B version. Llama 3.3 and Qwen-2.5 use the same prompt as the baseline ChatGPT classification. ChatGPT (strong perturbation) uses the minimal prompts specified under strong perturbation.

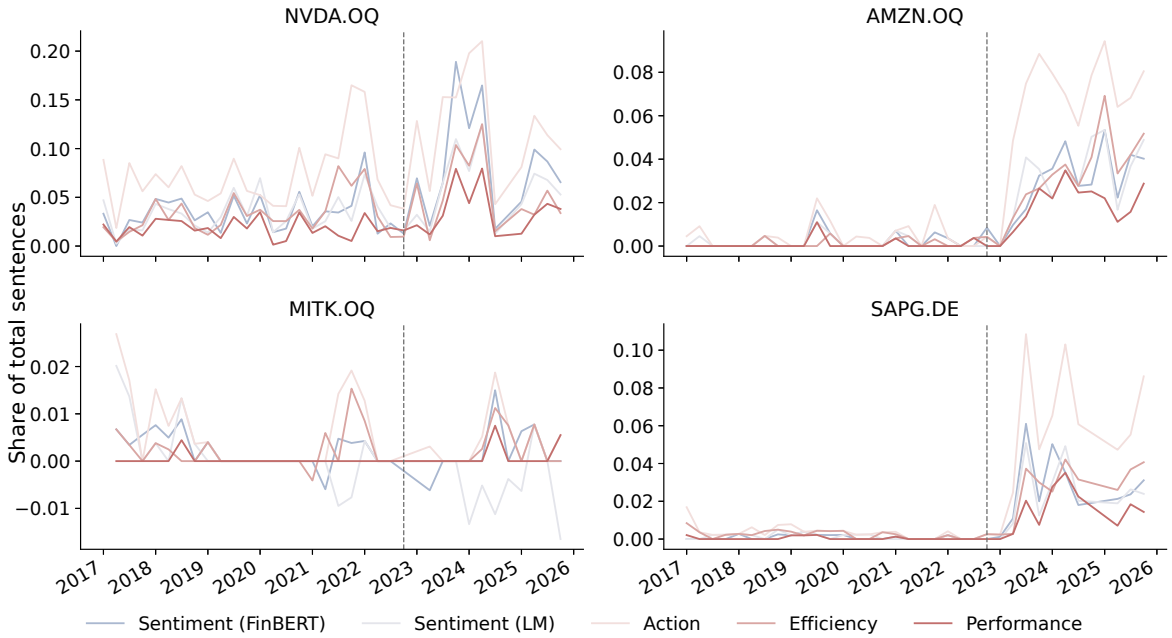
A.1.6 Examples

Table A4 provides a list of selected examples. Due to an error in the earnings call transcript, the first "sentence" is actually two sentences. ChatGPT understands that developers using GitHub copilot will be more productive and therefore gives efficiency a label of 1. Additionally, the action label is 1, presumably because the sentence implies that the integration of GitHub copilot will be completed by the end of the year, and doing so necessitates an action. ChatGPT assigns the time label "short-term", presumably because the sentence states that all developers will use copilot by the end of the year.

Sentence	Firm	Sentiment (FinBERT)	Performance	Efficiency	Action	Time Label
And already today, we use it to optimize our customer interactions, and we make it a daily companion for our employees.100% of Proximus developers will be using GitHub copilot by the end of the year.	Proximus NV 2024	0	0	1	1	short-term
A lot of people worry about the effect of artificial intelligence, the sort of collapse of the middle class job, doctors being out-diagnosed by artificial intelligence, Google Health and matters like that.	Wolters Kluver NV 2019	-1	0	0	0	none
Now with respect to the artist community and what we're hearing, we've heard a number of ours expressing their concerns about AI being employed to create content that's training on their work without consent.	Universal Music Group 2023	-1	0	-1	0	present
At Google I/O, we also highlighted how AI is improving Google Maps, including enhancing the experience with Assistant and AR features.	Alphabet Inc 2018	1	0	1	1	present
So I think the kind of bottom line is I think we're positioned extremely well for the eventual rollout of autonomous vehicles.	Lyft Inc 2020	1	0	0	1	long-term
Under the leadership of Barbara Lavernos, who's in charge of R&I and technology, we plan to invest on technology, on data and on AI to become the leading beauty tech company.	L'Oreal SA 2023	1	0	0	1	short-term
Secondly of course, we're also investing into connectivity and electric and autonomous driving in trucks, which has – which was a burden also in the first quarter.	Mercedes-Benz AG 2019	0	-1	-1	1	present
Non-GAAP gross margin in Q4 was 78.8%, slightly lower than Q4 of last year, primarily due to strategic investments in AI, partially offset by efficiency gains.	Zoom Communications 2025	-1	-1	1	1	present
Key wins in the quarter included Columbia University, using high-end Quadro GPUs for AI, and Siemens, using them for CT and ultrasound solutions.	NVIDIA 2018	1	1	0	1	present
Going forward, we see AI as a significant PC demand driver as Microsoft and other large software providers incorporate generative AI into their offerings	Advanced Micro Devices 2023	1	1	0	1	short-term
All up, the number of Azure AI customers continues to grow and average spend continues to increase.	Microsoft Corp 2024	1	1	0	0	present
But it just – it took another meaningful step forward with generative AI, but still it's complicated to actually figure out which models you want to work, you want to use and how you actually want to employ them and trying to make sure you have the right results, trying to make sure you get safe results, trying to make sure you end up with a cost structure and a customer experience that you want.	Amazon.com 2023	0	0	-1	0	present
As compared to ERNIE 3.0, 3.5 is capable of producing safer, more responsible and more creative responses, making significant improvements in question answering, reasoning and coding.	Baidu Inc 2023	1	0	1	0	present

Table A4: Selected Examples of Labeled Sentences

Figure A5: Sentence Shares of Earnings Calls Variables Over Time



Notes: The figure shows total firm-level quarterly action, efficiency, and performance mentions (all scaled by total sentences within the period), over time for Q1 2017–Q4 2025 for NVIDIA, Amazon.com, Mitek Systems and SAP. Quarters with zero AI mentions are coded as 0. Dashed line indicates ChatGPT release in Q4 2022.

A.2 Processing Lightcast Data

We use Lightcast (formerly Burning Glass Technologies) data on vacancies for US firms for the period 2010-2025, which contains online vacancies posted on more than 220,000 publicly accessible online job portals. As hiring for AI might occur at any stage or level, we do not filter vacancies further (i.e. we keep part-time jobs, internships, and other non-full-time arrangements). Lightcast already deduplicates job listings and only registers recurrent listings every 60 days, which we do not restrict further.

For each company, we count the number of total job listings and the number of total job listings that are flagged by the Lightcast Sector "Artificial Intelligence", and compute the number of AI job listings at the firm-level by quarter. Since lightcast only provides the start and end date of a vacancy, we define the stock of all active vacancies during a quarter as the union of vacancies that have a start date before the end of the quarter, and an end date after the beginning of the quarter.

Whereas the Refinitiv earnings calls transcripts are at the level of the publicly traded (parent) company, job postings are more granular at the level of the actual employer (e.g. Pixar Animation Studios belongs to Disney but has independent job postings). For that reason, we aggregate job postings to the parent firm using Wharton Research Data Services (WRDS) Company Subsidiary Data derived from SEC filings (Exhibit 21). Based on that aggregation, we compute the AI vacancy share os the fraction of AI job postings of total job postings within a period.

Given that Lightcast job listings do not contain a universal firm identifier, we use their cleaned company names and normalize company names from the earnings call and WRDS data by stripping firm endings (adapting the algorithm by Autor et al. (2020)). We additionally allow for fuzzy token matching to match the remaining firms based on minor spelling or word differences. For the remaining unmatched firms in the earnings call data we manually check for alternative firm names and specify these special cases.[20]

[20]We are left with about 200 unmatched firms in the earnings call data which fall into two broad cases: i) investment

Table A5 provides summary statistics by quarter on the job listings data.

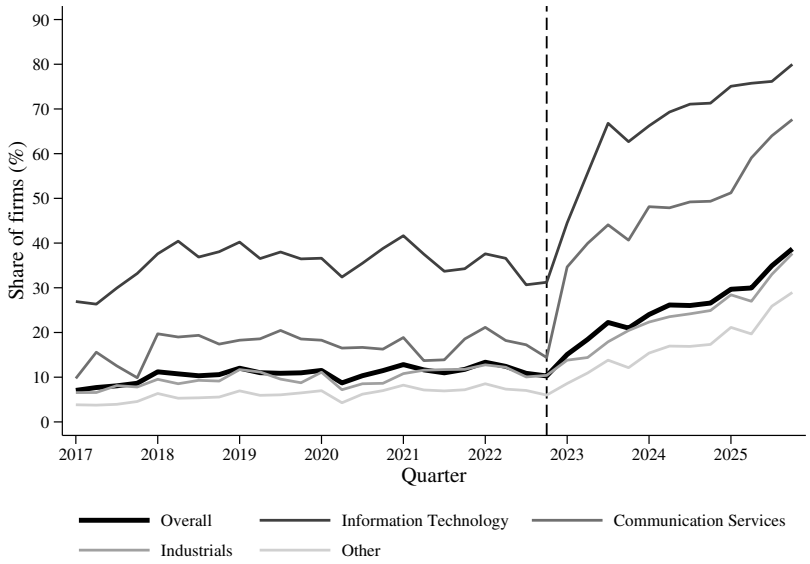
companies and REITs, ii) small or recent firms that have not posted any vacancies.

Table A5: Job listings by quarter: distinct firms, totals, AI jobs, and AI share.

Quarter	Distinct RICs	Total Jobs	AI Jobs	AI Share
2017Q1	2,768	1,610,083	15,993	0.99%
2017Q2	2,856	1,709,168	17,466	1.02%
2017Q3	2,847	1,804,403	20,200	1.12%
2017Q4	2,874	1,679,768	21,047	1.25%
2018Q1	2,948	1,666,378	27,956	1.68%
2018Q2	3,336	1,901,742	26,885	1.41%
2018Q3	3,465	2,289,552	34,826	1.52%
2018Q4	3,458	2,386,116	40,155	1.68%
2019Q1	3,414	2,055,957	38,233	1.86%
2019Q2	3,481	2,059,840	37,098	1.80%
2019Q3	3,536	2,027,764	32,523	1.60%
2019Q4	3,499	1,997,155	34,377	1.72%
2020Q1	3,494	2,083,247	36,052	1.73%
2020Q2	3,389	1,511,636	29,077	1.92%
2020Q3	3,449	1,642,807	29,228	1.78%
2020Q4	3,488	1,991,611	33,261	1.67%
2021Q1	3,502	1,907,321	39,816	2.09%
2021Q2	3,641	2,362,009	50,089	2.12%
2021Q3	3,588	2,561,632	51,096	1.99%
2021Q4	3,625	2,723,507	56,397	2.07%
2022Q1	3,624	2,666,592	63,247	2.37%
2022Q2	3,677	2,810,959	71,255	2.53%
2022Q3	3,692	2,619,157	64,060	2.45%
2022Q4	3,609	2,367,900	56,827	2.40%
2023Q1	3,544	1,906,461	37,562	1.97%
2023Q2	3,583	2,014,888	36,954	1.83%
2023Q3	3,557	2,029,989	37,367	1.84%
2023Q4	3,393	1,552,833	32,225	2.08%
2024Q1	3,429	1,557,652	37,574	2.41%
2024Q2	3,367	1,563,617	43,388	2.77%
2024Q3	3,406	1,777,290	49,820	2.80%
2024Q4	3,363	1,731,940	57,423	3.32%
2025Q1	3,300	1,681,230	53,081	3.16%
2025Q2	3,281	1,735,546	59,983	3.46%
2025Q3	3,222	1,861,648	74,201	3.99%
2025Q4	3,224	1,683,414	91,490	5.43%
Mean	3,387	1,987,023	42,729	2.16%

Notes: Job listings only for publicly listed US firms (aggregated at the parent level), as matched from Lightcast data to Refinitiv earnings call data. Job postings refer to vacancies posted during the reference period. AI jobs classified by Lightcast sector Artificial Intelligence

Figure A6: Quarterly Firm AI Mentions by Industry



Notes: The figure shows the fraction of firms mentioning AI in an earnings call within a quarter. Data is aggregated quarterly. See Appendix A for details on the AI-mentions classification. Industry follows 2-digit GICS business classification, remaining industries are grouped into Other.

B. Empirical Appendix

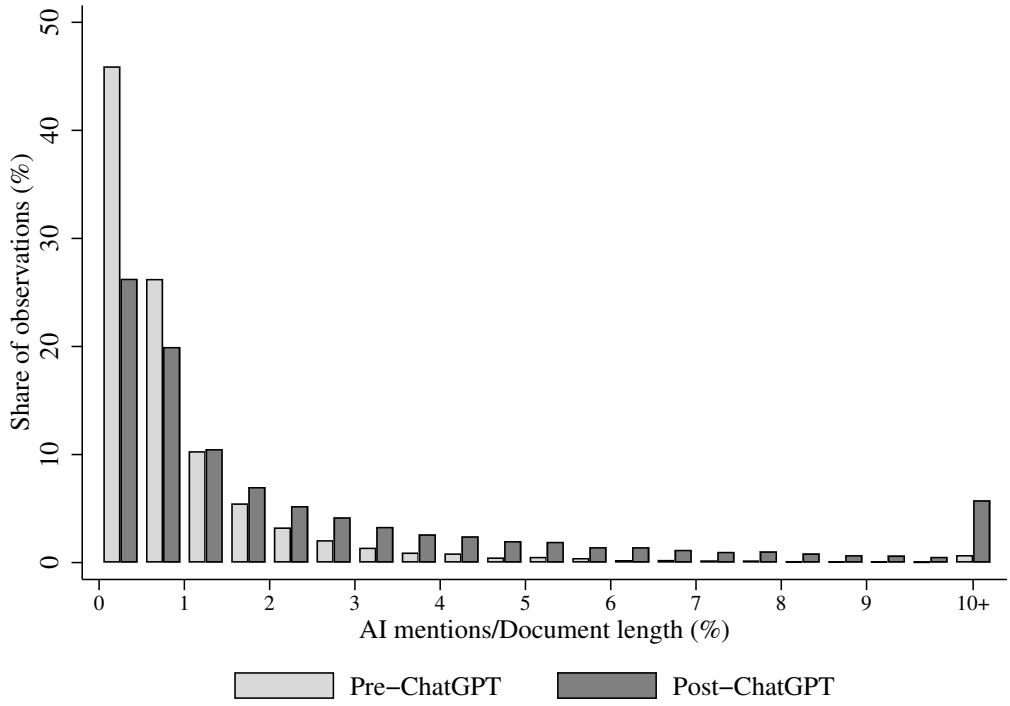
B.1 Additional Evidence of Heterogeneity

In this section we document additional patterns in the firm-level data illustrating heterogeneity in AI-related talk. We show that there is heterogeneity in the share of firms that mention AI. Conditional on mentioning AI atleast once, there is also substantial heterogeneity in the number of mentions of AI per earnings call. Both of these extensive margins of AI-mentions show a shift following the ChatGPT announcement. Finally, we document heterogeneity in net positive AI-related sentiment, action, efficiency and performance mentions of firms. The intensive margin heterogeneity does not noticeably change following ChatGPT.

Heterogeneity in share of firms mentioning AI. Figure 3 illustrated substantial heterogeneity in the share of AI-sentences, across industry and over time, with a surge in AI sentences following ChatGPT. Figure A6 corroborates that these patterns also hold at the firm-level. It shows the fraction of firms that mention AI at least once in an earnings call, by quarter. Prior to the ChatGPT announcement, the share of firms mentioning AI was relatively stable, at around 10% per quarter. The information technology sector showed the highest share of mentions, around 35-40% per quarter. Following the ChatGPT announcement, there is a prominent extensive margin effect. The share of firms mentioning AI sharply increases in all industries, to over 20% on average and over 70% in the information technology sector. These shares continue to trend upwards following this event, to the end of our data in Q4-2025.

Heterogeneity in mentions of AI per earnings call. While Figure A6 illustrated the extensive margin variation in whether or not firms mention AI in earnings calls, Figure A7 illustrates within-earnings call heterogeneity in the number of AI mentions. Conditional on AI mentions ≥ 1 within an earnings call, the for the vast majority of firms (nearly 80%) in the pre-ChatGPT period, AI sentences are less

Figure A7: Distribution of AI-Document Share



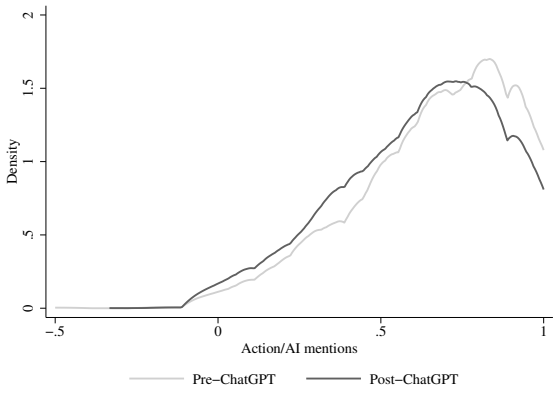
Notes: The figure shows plots the distribution of AI-related sentences relative to the total number of sentences over firm-quarter observations, conditional on having at least one AI-related sentence. Pre-ChatGPT period runs from 2017Q1–2022Q3, post-ChatGPT period from 2022Q4 onwards. The binwidth is set to 0.005, pre- and post-ChatGPT bars for same bin are shown next to each other.

than 0.5% of document length. However, there is substantial heterogeneity, and for a small share of firms, mentions are over 5% of document length. The distribution shifts to the right after the ChatGPT announcement, with more than 50% of firms' AI mentions being greater than 1% of document length.

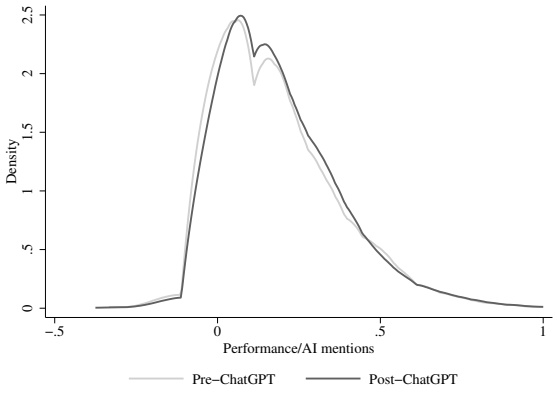
Heterogeneity in AI-associated variables. Finally Figure A8 conditions on a sample of firm-quarters with a minimum of 5 AI mentions in an earnings call, and plots the kernel densities of the net action/efficiency/performance and sentiment labels in this sample. There is again clear evidence of heterogeneity at the intensive margin in all variables, with most of the distribution in the net positive range for all variables (consistent with the evidence in Figure 5). In contrast to the share of firms mentioning AI or the number of AI mentions per document, however, there is no evident change in the beliefs of AI-related sentiment, performance or efficiency within the group of firms with more than 5 mentions of AI, following the ChatGPT announcement.

Correlations of AI-associated variables within earnings calls. Supplementing the evidence on heterogeneity of AI mentions and within-earnings call heterogeneity in AI-associated sentiment, action, efficiency and performance, Table A6 provides the pairwise correlation between all measures of AI-related variables, within a sample of firm-earnings calls with a minimum of one AImention. There is a strong positive correlation between all variables, suggesting that conditional on talking about AI, firms that have stronger net positive sentimnt about AI also speak positively about actions, performance and efficiency impacts.

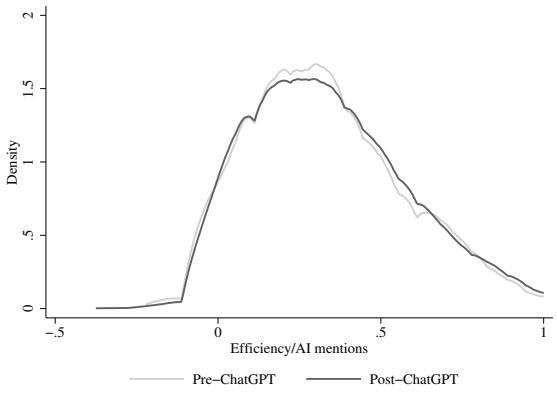
Figure A8: Heterogeneity in AI-Associated Variables



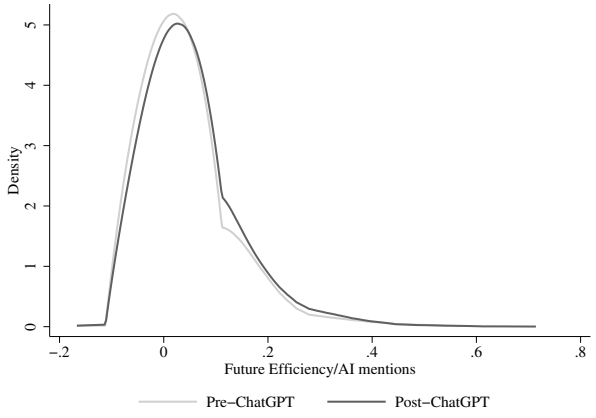
(a) Action



(b) Performance



(c) Efficiency



(d) Future Efficiency

Notes: Kernel density function (KDF) plots for action, performance, efficiency, and future efficiency. Future efficiency is a subset of efficiency and includes only sentences which have a time label of short-term or long-term. Data are aggregated to the firm-quarter level for the period 2017-2025. Pre-ChatGPT is form Q1 2017 to Q3 2022, post-ChatGPT from Q4 2022 onwards. See Appendix A for details on the classification. KDF using Epanechnikov estimation wit bandwith of 0.05. Only considering firm-quarter observations with at least 5 AI sentences.

Table A6: Cross-sectional correlation of AI-related labels.

	Action	Sentiment (LM)	Sentiment (FinBERT)	Efficiency	Performance
Action	1.000				
Sentiment (LM)	0.843	1.000			
Sentiment (FinBERT)	0.895	0.860	1.000		
Efficiency	0.850	0.771	0.762	1.000	
Performance	0.768	0.713	0.844	0.601	1.000

Notes: Data for Q1 2017–Q4 2025. Only documents with AI sentences are considered. Correlations are on firm-quarter level based on the exposure per document length as shown in 2.1.

B.2 The Similarity Measure

This section defines the AI-talk similarity measure used in Section 3 to assess whether firms' discussions of AI become more or less similar over time.

Definition. For a given quarter t, let there be N_t documents (defined as the firm-quarter observations of AI-sentences) indexed by $f \in \{1, \ldots, N_t\}$. Define:

- $\mathcal{A}_t \subset \{1, \ldots, N_t\}$: set of documents with at least one AI sentence,
- $\mathcal{Z}_t = \{1, \ldots, N_t\} \setminus \mathcal{A}_t$: set of documents with zero AI sentences.

For each document $f \in \mathcal{A}_t$, let $\{\mathbf{v}_{fi}\}_{f=1}^{n_i}$ denote the sentence embeddings of its AI sentences. The document-level embedding (normalized to unit length) is $\mathbf{m}_f = \frac{\frac{1}{n_f}\sum_{s=1}^{n_f}\mathbf{v}_{fi}}{\left\|\frac{1}{n_f}\sum_{i=1}^{n_f}\mathbf{v}_{fi}\right\|}$.

Define pairwise cosine similarity as $c_{fj} = \frac{\mathbf{m}_f^\top \mathbf{m}_j}{\|\mathbf{m}_f\|\|\mathbf{m}_j\|} = \mathbf{m}_f^\top \mathbf{m}_j$, since document embeddings are of unit length

Weighting. Let n_i denote the number of AI sentences in document f. For documents without AI content, set their effective sentence count to 1 (such that they are not omitted when weighting). Define the effective count

$$n_f^{\text{eff}} = \begin{cases} n_f & \text{if } f \in \mathcal{A}_t, \\ 1 & \text{if } f \in \mathcal{Z}_t. \end{cases}$$

In the weighted specification, each observed pair (f, j) receives weight

$$w_{fj} = \frac{n_f + n_j}{2},$$

where n_f and n_j are the numbers of AI sentences in documents f and j.

For pairs involving a non-reporting document, it follows that for a zero-AI document f and an AI-reporting document j the corresponding pair weight is $w_{fj} = \frac{n_j}{2}$, and for two zero-AI documents f, j, $w_{fj} = 0$.

Random-Draw Imputation for Non-Reporting Documents. Documents without AI sentences do not have observed document embeddings in this construction, so similarities involving such documents are imputed. For each quarter t, define two quarter-specific empirical distributions of observed AI–AI similarities:

$$\mathcal{D}_t^{\text{within}} \qquad \text{and} \qquad \mathcal{D}_t^{\text{cross}}.$$

- $\mathcal{D}_t^{\text{within}}$ is the distribution of similarities among low-AI documents in quarter t,
- $\mathcal{D}_t^{\text{cross}}$ is the distribution of similarities between low-AI and high-AI documents in quarter t.

A draw regime then truncates these distributions. For example:

$$\mathcal{D}_t(r) = \begin{cases} \{c_{fjt} : c_{fjt} \le q_{0.25}\} & \text{if } r = \text{p25}, \\ \{c_{fjt} : c_{fjt} \le q_{0.50}\} & \text{if } r = \text{median}, \\ \{c_{fjt}\} & \text{if } r = \text{full}. \end{cases}$$

For each simulation draw b, the imputed similarities are generated as

$$d_{t,b}^{\text{within}} \sim \mathcal{D}_t^{\text{within}}(r), \qquad d_{t,b}^{\text{cross}} \sim \mathcal{D}_t^{\text{cross}}(r).$$

Thus, all unobserved similarities involving non-reporters are treated as random draws from quarter-specific empirical AI-similarity distributions. For every pair (f, j) with $f < j$, define the pairwise similarity

$$\tilde{c}_{fj,t} = \begin{cases} \mathbf{m}_{ft}^\top \mathbf{m}_{jt} & \text{if } f, j \in \mathcal{A}_t, \\ d_{fjt}^{\text{within}} & \text{if } f, j \in \mathcal{Z}_t, \\ d_{fjt}^{\text{cross}} & \text{if } f \in \mathcal{Z}_t,\ j \in \mathcal{A}_t, \end{cases}$$

where

$$d_{fjt}^{\text{within}} \sim \mathcal{D}_t^{\text{within}}, \qquad d_{fjt}^{\text{cross}} \sim \mathcal{D}_t^{\text{cross}}.$$

Similarity Measure. Each pair (f, j) receives weight

$$w_{fjt} = \begin{cases} \frac{n_{ft}+n_{jt}}{2} & \text{if } f, j \in \mathcal{A}_t, \\ \frac{n_{jt}}{2} & \text{if } f, j \in \mathcal{Z}_t, \\ 0 & \text{if } f \in \mathcal{Z}_t,\ j \in \mathcal{A}_t. \end{cases}$$

The quarter-level similarity measure is then $\bar{c}_t^{\text{weighted}} = \dfrac{\sum_{f<j} w_{fjt}\, \tilde{c}_{fjt}}{\sum_{f<j} w_{fjt}}$.

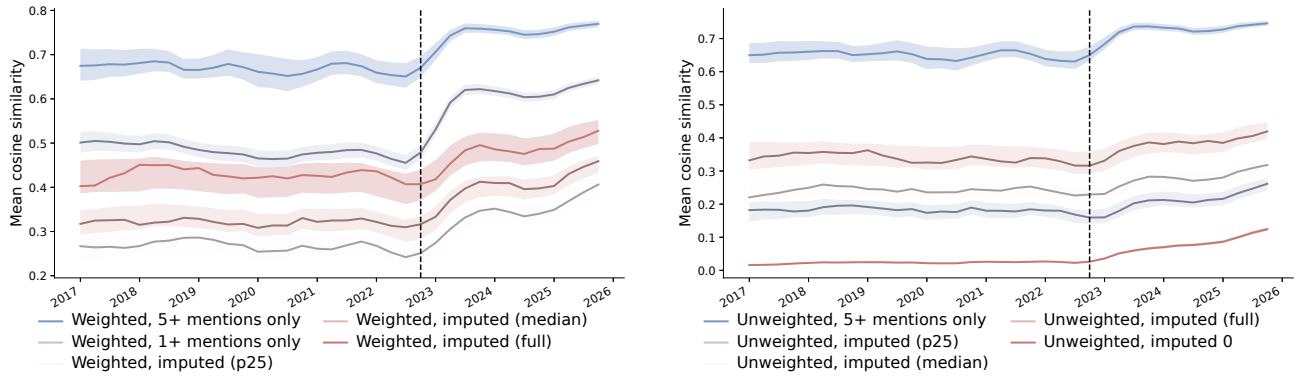
The unweighted version assigns equal weight to all pairs, so that $\bar{c}_t^{\text{unweighted}} = \frac{1}{\binom{N_t}{2}} \sum_{f<j} \tilde{c}_{fjt}$.

Standard errors are obtained by document-cluster bootstrapping with 500 repetitions.[21]

Figure A9 illustrates the similarity measure over time for different samples, weighting schemes, and with different assumptions on the distribution from which to impute similarities for firms with no AI-talk. Under all assmptions, there is a noticeable increase in AI talk similarity following the ChatGPT announcement.

[21] For an overview of recent advances in using word embeddings for document similarity, see Seegmiller, Papanikolaou, and Schmidt (2023)

Figure A9: Robustness of AI talk similarity



(a) Alternative Weighted Similarity Measures

(b) Unweighted Similarity Measures

Notes: The figure displays semantic similarity of AI-related discussion at the *document* (firm-quarter) levels using `all-mpnet-base-v2`, constructed as discussed in the text. Document-level represents the average embedding of all AI-sentences inside an earnings call. All embeddings are L2-normalized and the sample contains AI-flagged sentences only. Weighting follows the approach described in B.2. Under unweighted similarity all documents receive the same weight. *5+ mentions* only considers firm-quarter observations with at least 5 mention, analogously for *1+ mentions*. These similarity measures do not consider documents with fewer mentions. Under the *Imputed 0* measure, all pairwise combination that include any documents without AI sentences always receive a similarity of 0. Under the *Imputed (p25, median, full)* regime, similarities of documents without AI sentences randomly drawn from different percentiles of the distribution of observed similarities. Vertical marker indicates the ChatGPT release (Q4 2022). 95% confidence intervals are obtained via a transcript-cluster bootstrap per quarter. Mean and CI are smoothed with a centered three-quarter MA.

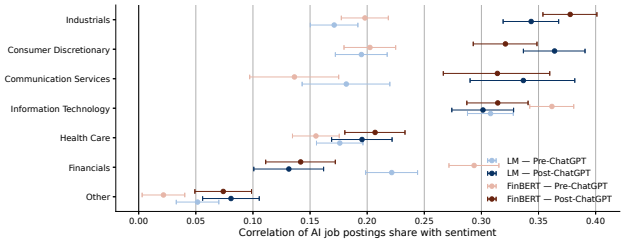
B.3 Additional evidence: AI job listings and AI-associated sentiment/action mentions.

Table 1 shows estimates of equation 3.1 that net action measures and AI job vacancy shares are both positively correlated with firm beliefs about AI efficiency and performance, conditional on firm and time fixed effects. Panel A of Figure A10 shows the unconditional correlation, by broad industry, of AI-related sentiment and AI job vacancy shares. There is a positive correlation before and after ChatGPT across all industries. However, without conditioning on firm and time fixed effects as in equation 3.2, there is no clearly visible shift in the correlation after ChatGPT, with increases for some industries and decreases for some industries. Panel B of the Figure shows that there is a positive correlation between the broader net action variable and AI job vacancy shares for all industries. To assess whether there is a pattern in the industries that see increasing correlations following ChatGPT, we estimate

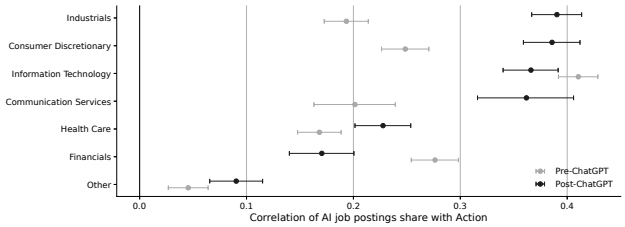
$$y_{ft} = \alpha_f + \lambda_t + \beta \left(\text{Exposed}_{i(f)} \times \text{Post}_t \right) + \varepsilon_{ft} \tag{B.1}$$

where y_{ft} is the AI vacancy share of firm i at time t, α_f are firm fixed effects, λ_t are year fixed effects, $\text{Exposed}_{i(f)}$ indicates whether firm f belongs to a exposed sector/industry i, and Post_t is an indicator equal to one for $t \geq$ Q4 2022. To define exposed industries, we take three approaches. First, we assume firms operate in Information Technology or Communication Services are exposed. Second, we consider as treated firms belong to GICS business sub-sectors with higher expected exposure to AI. Finally, we consider as treated firms operating in NAICS4 industries whose AI Industry Exposure Index (AIIE)—constructed by Felten, Raj, and Seamans (2021) from occupational AI task exposure in 2019—is strictly positive (above mean AI exposure). In all cases, firms operating in more exposed industries significantly increase AI vacancy shares following the ChatGPT release. Figure A11 estimates the event-study version of equation B.1, and illustrates that there were no pre-trends for more-AI exposed industries prior to ChatGPT, but these industries differentially increased

Figure A10: Correlations of AI Job Listings and Firm Outcomes



(a) Sentiment and Job listings in AI



(b) Firm mentions of Action and Job listings in AI

Notes: Pre-ChatGPT contains observations between Q1 2017 and Q3 2022, post-ChatGPT contains observations from Q4 2022. Coverage is restricted to public US firms only. Industry follows 2-digit GICS business classification with the following industries grouped into *Other*: Real Estate, Utilities, Materials, Consumer Staples, Energy. See Appendix A for details on the classification. AI job share is computed from Lightcast data dividing the number of AI jobs by the number of total jobs listed by the firm. Job listings are aggregated quarterly to match the frequency of earnings calls. We aggregate job listings at the parent company level and apply substring matching to combine the data with earnings call information. Panel a) The figure shows the cross-sectional Pearson correlation (with 95% confidence interval) between sentiment and AI job share on the document level. Sentiment density is computed as the number of positive sentences minus the number of negative sentences divided by the number of total sentences in the earnings call. We compute sentiment using two methods: LM follows the lexical approach from the Loughran-McDonald Dictionary (Loughran and Mcdonald, 2011) whereas FinBERT is a pre-trained large language model to detect sentiment in financial text. Panel b) The figure shows the cross-sectional Pearson correlation (with 95% confidence interval) between action-classified AI sentences and AI job share on the document level. The share of action AI sentences is computed as the number of AI sentences classified as containing content related to firm actions towards AI divided by the total number of sentences.

their AI job vacancy shares after the ChatGPT announcement.

In Table A8 we show the positive correlation exists even when further conditioning on firm beliefs that AI will impact future performance or efficiency. The elasticities decrease by about one-third for beliefs about future efficiency (0.275 vs 0.790 using net action as the outcome, and 0.025 vs 0.07 using AI job vacancy shares) and by about one-half for beliefs about future performance (0.238 vs 0.488 using net action and 0.022 vs 0.044 using AI job vacancy shares), suggesting the positive response associated with beliefs about future efficiency and performance is more muted than when the beliefs are about current efficiency or performance. Table A9 shows further that the association between beliefs about future efficiency and performance and AI-actions or AI-job vacancy shares strengthens after the ChatGPT announcement. Similar to 2, for columns (3) and (4) the main effect becomes negative, suggesting that the AI job vacancy share begins to increase only after the ChatGPT announcement, for firms that had positive beliefs about future AI impacts on efficiency and performance.

Table A7: AI Vacancies Share Before and After ChatGPT

	(1) Industry	(2) GICS	(3) AIIE
Exposed × Post	0.037	0.026	0.013
	(0.005)	(0.003)	(0.002)
Firm FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	81,852	81,852	56,272

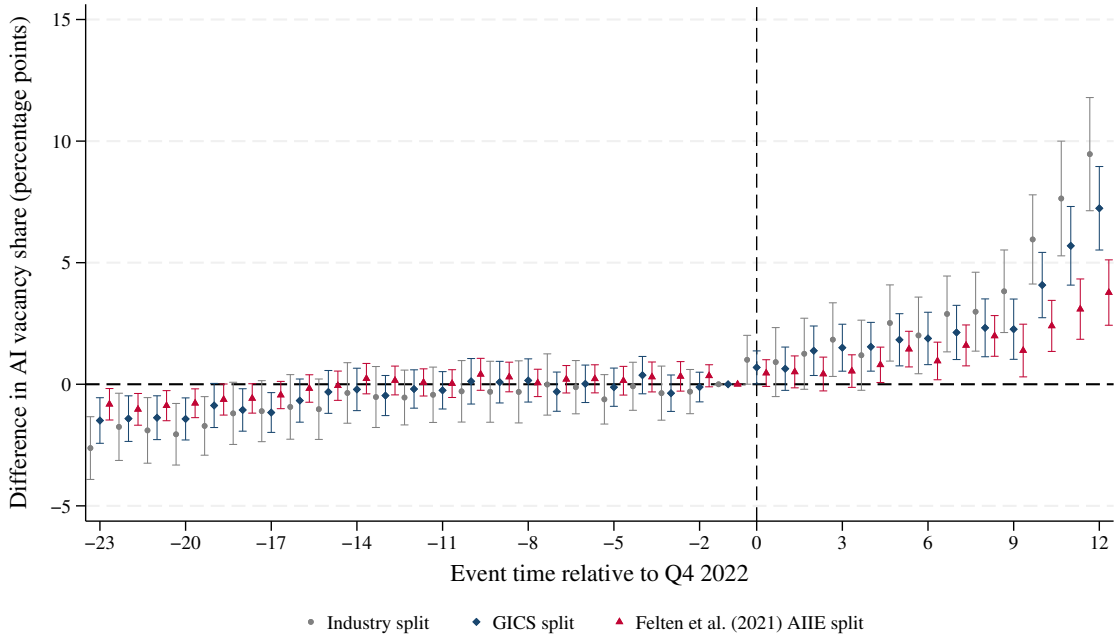
Notes: The outcome variable is the firm-level AI vacancy share, computed each quarter as the ratio of AI-related job vacancies to total job listings. All regressions are estimated at the firm–quarter level. Treatment status is defined at the sector or industry level. Standard errors are reported in parentheses. For the *Industry split*, treated firms operate in Information Technology or Communication Services; all other firms are untreated. For the *GICS split*, treated firms belong to GICS business sub-sectors with higher expected exposure to AI. For the *AIIE split*, treated firms operate in NAICS4 industries whose AI Industry Exposure Index (AIIE)—constructed by Felten, Raj, and Seamans (2021) from occupational AI task exposure in 2019—is strictly positive (above mean AI exposure). The estimating equation is (B.1).The pre-ChatGPT period comprises observations from Q1 2017–Q3 2022. The sample is restricted to matched publicly listed U.S. firms.

Table A8: Actions and Beliefs about the Future

	Net Action		AI Vacancy Share	
	(1)	(2)	(3)	(4)
Net Efficiency (Future)	0.275		0.025	
	(0.008)		(0.004)	
Net Performance (Future)		0.238		0.022
		(0.011)		(0.004)
R^2	0.619	0.623	0.515	0.515
Observations	210,251	210,251	79,795	79,795
Firm FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

Notes: The table shows estimates of specification (3.1), where the belief measures condition on future time labels (near or distant future). The AI Vacancy share is constructed from Lightcast data and only available for US firms. For the definitions of the net efficiency and performance measures, see Section 2, specifically, equation (2.4). Standard errors clustered at the firm level and reported in parentheses. The sample period runs from 2017q1 to 2025q4.

Figure A11: Event Study: AI Vacancies Before and After ChatGPT



Notes: The AI vacancy share is computed at the firm–quarter level as the ratio of AI job listings to total job listings. AI job listings are the subset of vacancies assigned by Lightcast to the "Artificial Intelligence" occupational category. Means are then computed separately for treated and untreated firms, and the point estimates represent the change in the treated-untreated difference in a given year relative to the Q3 2022 difference (normalized to 0 in t-1). Treatment groups are defined at the sector or industry level. For the *Industry split*, treated firms operate in Information Technology or Communication Services. For the *GICS split*, treated firms belong to GICS business sub-sectors with higher expected exposure to AI. For the *AIIE split*, treated firms operate in NAICS4 industries whose AI Industry Exposure Index (AIIE)—developed by Felten, Raj, and Seamans (2021) based on occupational task exposure to AI in 2019—is strictly positive (above mean exposure). The estimating equation is: $y_{ft} = \alpha_f + \lambda_t + \sum_{k \neq 2022q3} \delta_k \left(\text{Exposed}_{i(f)} \times \mathbf{1}\{t = k\} \right) + \varepsilon_{ft}$, corresponding to the DiD estimating equation (B.1). The sample is restricted to matched publicly listed U.S. firms. Each coefficient δ_k measures the change in the treated-untreated mean difference in quarter k relative to the baseline period Q3 2022. 95% confidence intervals are reported.

Table A9: Actions and Beliefs about the Future pre and post ChatGPT

	Net Action		AI Vacancy Share	
	(1)	(2)	(3)	(4)
Net Efficiency (Future)	0.117		-0.010	
	(0.012)		(0.008)	
Post × Net Efficiency (Future)	0.185		0.043	
	(0.017)		(0.009)	
Net Performance (Future)		0.110		-0.010
		(0.011)		(0.014)
Post × Net Performance (Future)		0.143		0.037
		(0.017)		(0.014)
R^2	0.628	0.629	0.516	0.516
Observations	210,251	210,251	79,795	79,795
Firm FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

Notes: The table shows estimates of specification (3.2), where the indicator variable post_t is defined as $\text{post}_t = 1\{t \geq 2022q4\}$ and the belief measures condition on future time labels (near or distant future). The AI Vacancy share is constructed from Lightcast data and only available for US firms. For the definitions of the net efficiency and performance measures, see Section 2, specifically, equation (2.4). Standard errors clustered at the firm level and reported in parentheses. The sample period runs from 2017q1 to 2025q4.

C. Proofs

D. Quantification Appendix

References

Allee, Kristian D., Chuong Do, and Mikhail Sterin. 2021. "Product Market Competition, Disclosure Framing, and Casting in Earnings Conference Calls." *Journal of Accounting and Economics* 72 (1):101405.

Araci, Dogu. 2019. "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models."

Autor, David, David Dorn, Gordon H. Hanson, Gary Pisano, and Pian Shu. 2020. "Foreign Competition and Domestic Innovation: Evidence from US Patents." *American Economic Review: Insights* 2 (3):357–374.

Barrios, John Manuel, John L. Campbell, Ryan G. Johnson, and Y. Christine Liu. 2024. "Signals or Smoke? The Determinants and Informativeness of Corporate Artificial Intelligence (AI) Disclosures."

Cao, Sean, Wei Jiang, Baozhong Yang, and Alan L. Zhang. 2023. "How to Talk When a Machine is Listening: Corporate Disclosure in the Age of AI." *The Review of Financial Studies* 36 (9):3603–3642.

Clayton, Christopher, Antonio Coppola, Jesse Schreger, and Matteo Maggiori. 2025. "Geoconomic Pressure." Working Paper 34020, National Bureau of Economic Research.

Felten, Edward, Manav Raj, and Robert Seamans. 2021. "Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses." *Strategic Management Journal* 42 (12):2195–2217.

Hassan, Tarek A., Stephan Hollander, Aakash Kalyani, Laurence van Lent, Markus Schwedeler, and Ahmed Tahoun. 2025. "Text as Data in Economic Analysis." *Journal of Economic Perspectives* 39 (3):193–220.

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):531–563.

Huang, Allen H., Hui Wang, and Yi Yang. 2023. "FinBERT: A Large Language Model for Extracting Information from Financial Text." *Contemporary Accounting Research* 40 (2):806–841.

Loughran, Tim and Bill Mcdonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *The Journal of Finance* 66 (1):35–65.

Loughran, Tim and Bill McDonald. 2020. "Textual Analysis in Finance." *Annual Review of Financial Economics* 12 (1):357–375.

Sautner, Zacharias, Laurence Van Lent, Grigory Vilkov, and Ruishen Zhang. 2023. "Firm-Level Climate Change Exposure." *The Journal of Finance* 78 (3):1449–1498.

Seegmiller, Bryan, Dimitris Papanikolaou, and Lawrence DW Schmidt. 2023. "Measuring Document Similarity with Weighted Averages of Word Embeddings." *Explorations in Economic History* 87:101494.