Norwegian School of Economics
Bergen, Fall 2023

AI and Firm Values: How the Release of ChatGPT Affected Expectations Toward AI-exposed Firms

A Difference-in-Differences Approach

Espen Møller Pettersen and Fredrik Øverland Hatling

Supervisor: Maximilian Rohrer

Master thesis, Economics and Business Administration

Major: Financial Economics & Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

We would like to express our deep gratitude to our supervisor, Maximilian Rohrer, whose expertise and guidance have been invaluable to the successful completion of this thesis. We also want to thank the Library staff at Norwegian School of Economics for providing us with the resources necessary to conduct this study.

Norwegian School of Economics

Bergen, December 2023

Espen Møller Pettersen

Fredrik Øverland Hatling

Abstract

The release of ChatGPT constituted a major technology shock to AI, offering new information about AI advancement and its future potential. This paper studies the impact of the release on analyst stock recommendations and earnings forecasts for AI-exposed firms, which are argued to have a competitive advantage in capitalizing on such advancements. We measure AI-exposure with a combination of natural language processing techniques applied to earnings call transcripts. Using a difference-in-differences model, we find evidence of recommendations for AI-exposed firms moving towards "sell". Findings also indicate a positive effect on earnings expectations likely driven by reduced cost forecasts. However, recommendations indicate that the earnings effect is not strong enough to justify an upgrade. Thus, we attribute "sell" recommendations to stock prices moving beyond what analysts consider fundamental values.

Keywords – ChatGPT, artificial intelligence, natural language processing, analyst expectations, difference-in-differences

Contents

List of Figures

List of Tables

1 Introduction

The 21st century has seen remarkable advancements in artificial intelligence (AI), which have transformed industries, redefined how we interact with computers and revolutionized business operations (Press, 2019). Of the many breakthroughs within AI, one that has received significant attention is the emergence of generative AI models powered by deep learning and neural networks. At the forefront stands ChatGPT, developed by OpenAI and released to the public on November 30$^{\text{th}}$, 2022 (OpenAI, 2023). The advent of ChatGPT has been called "the iPhone moment of AI" owing to its wide applicability and disruptive potential, and by January 2023, the model had already become the fastest-growing consumer application in history (Hu, 2023; Viviani, 2023).

Recent developments in generative AI, highlighted by the release of ChatGPT, are widely seen as a major technology shock with large potential impacts on firm values (Eisfeldt et al., 2023). However, ChatGPT launched only a year ago, and its full disruptive potential might not yet be realized. To fully understand the impact of the ChatGPT release, it is therefore interesting to study how it affected expectations for the *future* of AI companies. This paper is premised on the belief that the ChatGPT release represented a paradigm shift in AI technology, providing signals of AI's increased importance and power in the future. As such, the release is hypothesized to trigger a positive shift in expectations toward firms positioned to capitalize on such advancements. This group of firms is referred to as "AI-exposed", defined as companies that develop and/or extensively utilize AI in business processes and product offerings. Through studying analyst stock recommendations and earnings forecasts (collectively referred to as analyst expectations), our paper provides an understanding of how the ChatGPT release impacted firm values for AI-exposed companies.

Before studying analyst expectations, we measure firm-level AI-exposure. The exposure measure indicates how related earnings call discussions are to AI. In earnings calls, executives and investors discuss important topics such as recent results and their implications for the future (Price et al., 2012). Therefore, if AI constitutes a core part of a company, it is likely discussed in these events. We capture AI-related discussions by counting bigrams in earnings call transcripts. The bigrams, which are combinations of two consecutive words such as “artificial intelligence” or “neural network”, are weighted by their relevance to AI. We apply a combination of two weighting techniques. First, we utilize a pre-trained natural language processing (NLP) model to assess the similarity between each bigram and the bigram “artificial intelligence”. Second, we weigh each bigram by how related it is to a corpus consisting of AI text relative to a corpus of earnings call text. The resulting firm-level exposure score is found to be a good proxy for AI-exposure in accordance with our definition, capturing companies widely known for utilizing AI and a wide spectrum of AI integration across industries.

To test our hypothesis of how the ChatGPT release affected analyst expectations toward AI-exposed firms, we utilize a difference-in-differences (DiD) model. The DiD framework is ideal for our purpose because it enables us to compare firms with high AI-exposure to firms with low exposure over time. Firms above a certain exposure threshold presumably possess advanced AI infrastructure and expertise, making them advantageously positioned to capitalize on future AI developments. Following this argument, we define the treated group as the top 5% of firms ranked by AI-exposed exposure (149 firms). As a preliminary control group, we define the bottom 50% of AI-exposed firms (1,481 firms). Omitting semi-exposed companies ensures a significant exposure gap between groups. To improve the validity of our models, we also construct a control group based on nearest-neighbor propensity score matching.

We begin by analyzing stock recommendations where we find no significant effects in the first few months after the ChatGPT release. When segmenting the post-treatment period into months, we find a significant downgrade in recommendations after 4-6 months. We argue that these effects are likely driven by an increase in share prices and that analysts considered stocks to be overvalued compared to fundamentals.

Next, we analyze components of firm value. We study earnings forecasts and subsequently decompose the earnings effect into revenues and costs by studying sales and gross margin forecasts. Forecast data for different horizons is used, specifically 1-5 years ahead. In addition, we study the effect on long-term growth forecasts, which is the expected average growth over the coming business cycle. In our earnings analysis, we find some evidence of improved analyst expectations consistent with our hypothesis. However, the effects are not as convincing as anticipated, where the only conclusive finding is an increase on the three-year horizon. Further, we analyze sales and gross margin forecasts, attempting to decompose the observed increase in earnings forecasts. Our findings suggest a positive sales reaction on the one-year horizon, yet statistical significance is weak (10%). The results indicate a downward adjustment in sales forecasts on longer horizons, particularly for long-term growth. The findings for gross margin forecasts are inconclusive. From jointly analyzing the findings, we argue that the most likely explanation for the increase in earnings forecasts is a decrease in projected cost levels. However, the results are not strong enough to make any firm conclusions.

In sum, our findings suggest that earnings expectations toward AI-exposed firms improved slightly following the release of ChatGPT. Stock recommendations were downgraded following the release, suggesting that the increase in forecasted earnings was not sufficiently large to upgrade recommendations, justify increases in share prices, or both. Our findings suggest that the release of ChatGPT was incorporated into analyst expectations as positive news for AI-exposed firms, although the effect is not as strong as anticipated. In addition, the negative reaction to long-term sales forecasts could be an indication that the competitive advantage of AI-exposed firms is expected to dwindle over time.

Our paper mainly relates to two strands of literature. First, there is a large body of literature concerning the effect of AI on firm values. Among the effects studied is AI's potential to accelerate the pace of innovation and improve efficiency and productivity (Cockburn et al., 2018; Hang & Chen, 2022). Further, ChatGPT is only a year old, yet the literature concerning its impact is growing. For instance, Eisfeldt et al. (2023) find that companies exposed to generative AI earned daily excess returns of 0.4% following the release of ChatGPT and that the disruptive potential across industries is wide. Second, our paper relates to the literature on how technological shocks impact expectations toward affected firms. In this field, much of the literature is concerned with incumbent firms that face radical technological change. For instance, Hobijn and Jovanovic (2001) find that stock prices of incumbent firms fell during the IT revolution in the early 1970's in expectation of new firms leveraging the technology more proficiently. In addition, Benner (2010) finds that analysts are more positive towards incumbent firms that preserve and extend existing technology rather than pursuing new technology. If these findings are attributable to AI-exposed firms, it could explain the weak analyst reaction to the ChatGPT release.

However, these papers study settings where new technology can be regarded as a threat to incumbents. In contrast, our study assumes that incumbents (AI-exposed firms) are advantageously positioned to leverage the technology shock represented by the ChatGPT release.

This paper contributes to existing literature in three ways. First, most literature on the effect of AI adoption are either theoretical discussions or empirical analyses of past impacts. Studying analyst forecasts quantifies the expected effects of AI technology in the *future*, which is a valuable approach considering the pace at which AI develops and the uncertainty surrounding its future potential. Second, by analyzing forecasts for different horizons, we can gain a deeper understanding of the timing of the anticipated effects of the ChatGPT release. While for instance stock prices can indicate market expectations, they do not offer a clear distinction between effects occurring in one year compared to five years or if the effect is driven by risk or fundamentals. Third, rather than studying the effects of technological change on a specific industry or firm, our analysis offers insights into the effects on firms from various countries and industries with one thing in common: their exposure to AI technology developments.

2 Hypothesis Development

2.1 Background

We develop our hypotheses on the premise that the release of ChatGPT constituted a major technology shock within the realm of AI, providing new information to investors and the public about AI advancement and its potential applications. In addition, we assume that AI-exposed firms are better positioned to take advantage of the increased potential in AI than their non-AI counterparts. In the following, we explain the foundation for these assumptions.

2.1.1 The ChatGPT Release as a Technology Shock

Prior studies highlight the positive impacts of AI on business performance, yet it has been argued that it must develop into a more "general AI" to truly have a large-scale impact (Furman & Seamans, 2019). This type of technology can be referred to as artificial general intelligence (AGI), which displays cognitive abilities and problem-solving skills comparable to, or even greater than, that of humans (Goertzel, 2014). To date, no true AGI has been developed, yet ChatGPT has been identified as an emerging AGI and a significant step towards general intelligence (Bubeck et al., 2023; Morris et al., 2023). Following these arguments, we argue that the ChatGPT release represented a major shock to AI technology. Specifically, ChatGPT represents advancements to AI that can be applied for general purposes, which opens a new realm of possibilities for utilizing AI in all aspects of a corporation. In addition, we argue that the release signified massive progress, not only for generative AI but for AI in general, strengthening expectations towards the achievements and significance of AI developments in the future.

The significance of the ChatGPT release is underscored by the considerable adoption rates and attention it has attracted. As of November 2023, a year after launch, ChatGPT had over 100 million weekly users and over 92% of the Fortune 500 companies utilized the model to some extent (Kreps, 2023). Figure 2.1 displays Google Trends search statistics of the terms "artificial intelligence" and "generative AI" (Google Trends, 2023a, 2023b), where both trends exhibit significant spikes a few months after the release. Google Trends represents the general interest in a topic. Thus, the graphs illustrate how the attention towards AI escalated following the release of ChatGPT, underscoring its significance as a technology shock to AI.

Figure 2.1: Google trends of "artificial intelligence" and "generative AI"

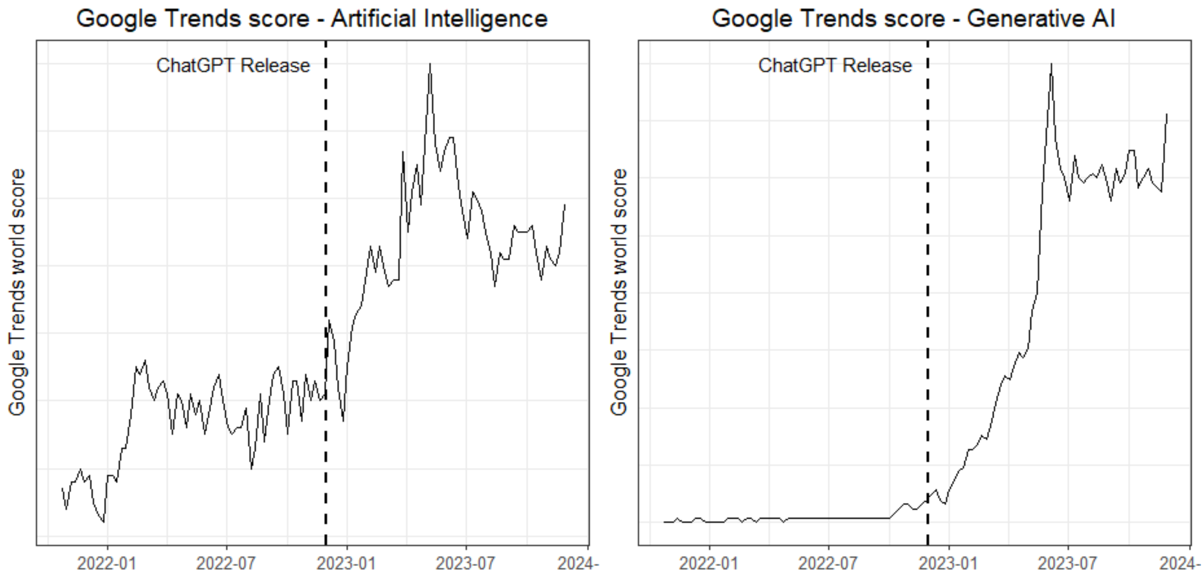

Figure 2.1 displays Google Trends graphs of the terms "Artificial Intelligence" and "Generative AI". The Google Trends score represents the interest in a topic. The graphs illustrate an escalation of attention towards AI following the ChatGPT release.

2.1.2 The Competitive Advantage of AI-Exposed Firms

Further, we argue that AI-exposed firms will stand at the forefront of the new era of AI development and applications. AI-exposed firms, provided they have a significant level of exposure, already possess advanced AI infrastructure and expertise. This enables them to leverage new technologies more proficiently and swiftly than their non-AI counterparts. Further, as suggested by Bughin et al. (2017), the competitive advantage of leading AI adopters is set to grow as there are no shortcuts to developing the digital foundation that AI adoption depends on. Since successful AI integration requires both time and skilled talent, where there is a particular shortage of the latter (Strack et al., 2021), we expect AI-exposed firms to outperform their non-AI counterparts in leveraging new technological developments such as generative AI.

2.2 Hypotheses

Under the assumption that the ChatGPT release constituted a major technology shock to AI and that AI-exposed firms have a competitive advantage in leveraging this shock, we expect analyst expectations toward these firms to improve following the release. Specifically, we expect the shock to signal an improved earnings potential for AI-exposed firms, increasing forecasted cash flows and firm values. In the first part of the analysis, we study analyst stock recommendations. Improved expectations towards the earnings capacity of a firm should lead to an upgrade in recommendations. Hence, we formulate our first hypothesis as follows:

***H1**: Stock recommendations for AI-exposed firms were adjusted towards "buy" following the release of ChatGPT.*

In the second part of the analysis, we study components of firm value to get a concise understanding of what drives analyst recommendations. We start by studying earnings forecasts. Since earnings are determined by revenues and costs, we decompose the earnings effect by studying sales and gross margin forecasts. We base our hypotheses for this part of the analysis on three ideas as to how the ChatGPT release is expected to affect revenues and costs for AI-exposed firms.

First, we expect the release to provide signals of an accelerated pace of innovation. Cockburn et al. (2018) predict that AI advances within general-purpose deep learning could come to be classified as the "*invention of a method of inventing*", with the potential of radically improving innovation capacity and quality. They suggest that advances in deep learning can dramatically reduce marginal search costs in R&D and improve the performance of existing research projects. Although ChatGPT was not released at the time of Cockburn et al. (2018), the infrastructure powering generative AI represents a major leap forward within deep learning (Chui et al., 2023). This suggests that the effects on innovation described could materialize. An improved capacity for innovation can have substantial effects on revenues and costs, for instance through monetization of new products and services, reduced R&D costs, and improved cost-effectiveness of business processes.

Second, studies suggest that AI's effect on productivity will be amplified following the advent of generative AI models. For instance, Dell'Acqua et al. (2023) found that consultants with access to ChatGPT-4 improved performance by as much as 40% and completed tasks 25% quicker than the control group. The potential impact on revenues can be illustrated following the example of consultants. If they can deliver work of higher quality faster, they can charge a higher price per billable hour. In addition, to the extent that AI can alleviate workers of time-consuming and labour-intensive tasks, more effort can be put towards activities that stimulate innovation and sales, such as strategy and research. Other studies have suggested that AI can reduce labor costs by improving the efficiency of the workforce (see for example Yang (2022) or Acemoglu and Restrepo (2020)). However, for this to have a cost-reducing effect, it would necessitate layoffs. Although it is plausible that human capital can be partly exchanged in the long run, we argue that increased productivity will rather scale up the output level, thus having a stronger effect on revenues.

Third, we expect the release of ChatGPT to signal greater potential for AI-driven data analysis. Historically, the main strength of AI models has been processing capacity and performing optimization tasks such as predictive modeling (Chui et al., 2023; Krakowski et al., 2023). Such applications can greatly impact revenues and costs, for instance by optimizing resource allocation, improving supply chain management, and predicting customer behavior. As AI systems become increasingly adept at analyzing vast amounts of data and learning from being exposed to new data, the accuracy in which AI can perform optimization and predictions will only increase.

We hypothesize that these effects will be stronger for longer horizons as the competitive advantage of AI-exposed firms intensifies. The immediate effect of the ChatGPT release was that all firms had a powerful AI tool at their disposal. However, AI-exposed firms are likely to develop and incorporate more advanced solutions than their non-AI counterparts over time. In addition, as Brynjolfsson et al. (2018) argues, the impact of AI technology might not be observed right after implementation. This may be particularly true for adopting advanced technologies such as those typically employed by AI-exposed firms. Further, monetizing AI technology has proven challenging, especially in the short-term (Wixom et al., 2020). Thus, the revenue effect of an accelerated pace of innovation for firms that offer AI services might not materialize quickly.

Based on the above arguments, the ChatGPT release is expected to lead to increased revenues and decreased costs for AI-exposed firms. We also expect the effects to be stronger on longer horizons. Presumably, analysts will incorporate the increased potential into their expectations. Since revenues and costs jointly determine earnings, we formulate our second hypothesis and two sub-hypotheses as follows:

H2: *Earnings forecasts for AI-exposed firms increased following the release of ChatGPT, with stronger effects on longer horizons.*

H2-1: *Sales forecasts for AI-exposed firms increased following the release of ChatGPT, with stronger effects on longer horizons.*

H2-2: *Cost forecasts for AI-exposed firms decreased following the release of ChatGPT, with stronger effects on longer horizons.*

3 Data

3.1 Collecting Data

We obtain data from three main sources: (1) The Institutional Brokers' Estimate System (IBES), (2) SeekingAlpha, and (3) Compustat. Data from IBES and Compustat is provided by Wharton Research Data Services (Wharton, n.d.). From IBES, we obtain data on analyst recommendations and EPS, sales, and gross margin forecasts. IBES compiles data from over 30,000 individual analysts and offers point-level forecasts for different fiscal periods (Dai, 2020). This is crucial to test the hypothesized term structure of the effects of the ChatGPT release. We extract data from the summary statistics databases containing consensus forecasts and recommendations. This means that each forecast is the average of all estimates, recorded monthly.

We obtain earnings call transcripts from SeekingAlpha (SeekingAlpha, n.d.). We use the SeekingAlpha API from RapidAPI (RapidAPI, 2023) to fetch earnings call transcripts from November 2021 through June 2023. Transcripts are obtained for all companies in the SeekingAlpha universe that also appear in our IBES sample. We extract annual fundamentals data and monthly close prices of shares from Compustat. Finally, we use Microsoft 365 (Microsoft, n.d.) to fetch daily currency exchange rates, which is used to convert variables into USD as a common currency.

3.2 Linking and Merging Databases

3.2.1 Linking IBES with SeekingAlpha

First, we extract the entire summary history database from IBES between January 2020 and June 2023. In this dataset, there are 25,574 unique companies. The primary challenge of linking IBES with SeekingAlpha lies in the absence of a uniform company identifier. In IBES, there are four identifiers: IBES ticker, which uniquely identifies a company in IBES; official ticker (OFTIC); company name (CNAME); and CUSIP/SEDOL. The SeekingAlpha API works by applying input to the search function on the SeekingAlpha website, where it is possible to search by two identifiers: ticker and the company name.

These identifiers do not always equal the tickers and company names in IBES. Thus, the challenge is finding the IBES identifier that returns the correct company when used as input in the search function in SeekingAlpha. After testing several IBES identifiers in the API, we conclude that CNAME provides the best match. For 25,574 companies, the API call returned 8,774 distinct SeekingAlpha matches.

To ensure data quality, we conducted a verification process where the SeekingAlpha ticker was cross-referenced with OFTIC from IBES. If there was no ticker match, a subsequent API request was initiated using a combination of CNAME and OFTIC. This provides a stricter requirement for a match because both the ticker and names must be similar. In the final dataset, we retain only companies that either have a ticker match or return a match for the second API call.

197 companies in SeekingAlpha were matched with two or more IBES companies. In these instances, we argue that the likelihood that companies appear in the SeekingAlpha universe increases with the size of the company. Following this logic, the most correct match is the IBES company with the largest analyst following since larger firms attract more attention from analysts. A condition is applied where the IBES company with the highest number of analysts following is chosen as the correct match. This condition is applied to 197 of a total of 8,774 companies, or 2.2% of the sample. If this condition should provide the wrong matches for a few of these 197 companies, the resulting error would be negligible.

Lastly, we conduct extensive manual checks to verify the quality of the matches. In addition, we omit companies that do not have earnings call transcripts in SeekingAlpha prior to the release of ChatGPT. We are left with a matched dataset containing 4,132 companies.

3.2.2 Merging IBES with Compustat

We merge IBES and Compustat so that each monthly IBES observation has the latest recorded fundamentals and price data. For a given month, this is fundamentals data from the previous fiscal year-end and the closing price from the previous month. After merging databases and omitting companies that lack desired data points, the final sample size is displayed in table 3.1.

Table 3.1: Final sample size

	Companies	Earnings calls	EPS Forecasts	Sales Forecasts	GRM Forecats	Recommendations
Full sample	2,959	11,134	392,812	392,916	267,914	56,942
Treated	149	562	20,180	20,710	17,514	2,531
Control	1,481	5,630	192,203	198,540	117,656	11,415

Table 3.1 displays the final sample size after linking and cleaning the data. The treated group consists of the top 5% of firms ranked by AI-exposure. The control group consists of the bottom 50% of AI-exposed firms. Earnings calls are from November 30$^{\text{th}}$, 2021 until November 30$^{\text{th}}$, 2022. Analyst data is from November 30$^{\text{th}}$, 2021 until May 31$^{\text{st}}$, 2023.

3.3 Variables and Data Description

3.3.1 Earnings Calls

Earnings calls are quarterly events associated with earnings releases, arguably some of the most important scheduled events in a firm's calendar (Garcia et al., 2023). Typically, an earnings call begins with prepared statements from management before opening up to questions from analysts, providing an important medium for managers to comment on recent results and emphasize their implications for the future (Kimbrough, 2005; Price et al., 2012). Moreover, the unscripted question-and-answer segment of earnings calls reveals discussions that investors deem important to the prospects of the company. Since these topics would not necessarily be included in official company reports, earnings calls provide candid insights about a company's most important issues, both from a company and investor perspective. Due to the importance of these events, AI is likely to be discussed if it constitutes a core part of a company. Therefore, their transcriptions are an ideal source of textual data to quantify the extent to which a firm utilizes and/or develops AI in accordance with our definition of AI-exposure.

3.3.2 Dependent Variables

Recommendations

In the recommendations analysis, we use $MEANREC_{i,\ t}$ as the dependent variable (hereby $REC_{i,\ t}$), which is the consensus analyst recommendation in a given month. t denotes the month of the recommendation, and i the firm for which the recommendation is made.

Analyst recommendations are reported on a scale from "strong buy" (1) to "sell" (5) and convey an analyst's view of a stock after analyzing available information. If we assume that markets are informationally efficient and that analysts are unbiased with no informational advantage, recommendations should convey a simple price-to-value comparison (Conrad et al., 2006). Basic finance theory suggests that the perceived value of a stock is determined by a company's earnings (cash flow) and its risk profile (discount rate). For the purpose of our study, we only consider the earnings component. If there is a disparity between the price of the stock and its perceived value, the analyst will change the recommendation from neutral (hold) in the direction that corresponds to the analyst's beliefs.

Forecast Variables

For EPS, sales, and gross margin forecasts, we analyze forecast data for horizons from 1 to 5 years ahead (hereafter 1y, 2y . . . , 5y). In the EPS analysis, we normalize estimates across firms by share price. The variable for firm i, forecasted in month t for horizon h, is defined as:

$$EPS\ norm_{i,t,h} = \frac{EPS_{i,t,h}}{P_{i,\ t}} \tag{3.1}$$

Where $P_{i,\ t}$ is the share price of the firm. To prevent normalizing by a variable that is itself affected by treatment, we use the latest share price prior to treatment for all post-treatment observations. Alternatively, the normalized measure can be expressed as:

$$EPS\ norm = \left(\frac{E}{S}\right)/P = \frac{E}{S \times P} \tag{3.2}$$

Or earnings divided by market capitalization, commonly referred to as the percentage earnings yield. Thus, the interpretation of $EPS\ norm_{i,t,h}$ is the forecasted earnings yield based on the current share price. We normalize EPS for two reasons. Firstly, the earnings yield is arguably a better representation of the earnings capacity of a firm than EPS, which is driven by both earnings and shares outstanding. To illustrate the potential confounding impact, consider a 2/1 stock split. If markets are efficient and absent other news, the price is halved. The latter effect is not picked up in EPS, which is then also halved since the denominator doubles. However, the earnings yield remains stable as both earnings and market capitalization are unaffected by the stock split. Secondly, a percentage value produces output with an intuitive interpretation comparable across firms. In contrast, a 1$ increase in EPS is more impactful for a small firm compared to a larger firm.

We use the logarithm of sales forecasts, $\log(SAL_{i,t,h})$, as the dependent variable in the sales analysis. $SAL_{i,t,h}$ is the forecast made in month t for firm i and horizon h. Thus, the coefficients on independent variables can be interpreted as the percentage change in the dependent variable for a one-unit increase in the independent variable. Since the magnitude of forecasts differs greatly with firm size, percentage values are useful for interpretation across firms.

We also study long-term growth forecasts for earnings and sales ($LTG_{i,t,h}$). This is the expected average growth over the coming full business cycle, usually referring to a period between three to five years (Derrien et al., 2021). Since $LTG_{i,t,h}$ is already in percentage terms, we use the forecast as reported in IBES as the dependent variable.

Finally, we use percentage gross margin forecast ($GRM_{i,t,h}$) for analyzing costs. The percentage gross margin is defined as:

$$GRM = \frac{revenue - COGS}{revenue} \tag{3.3}$$

Where $COGS$ is the cost of goods sold. Through a combined analysis of sales and gross margin forecasts, we can gain an understanding of projected cost levels. Since a change in gross margin ($GRM_{i,t,h}$) can happen through a change in either COGS or revenues ($SAL_{i,t,h}$), COGS can be estimated as the residual component of the equation.

3.3.3 Control Variables

Several control variables are used in regressions and propensity score matching. The *intangible asset ratio*, defined as $\frac{Intangible\ assets}{Total\ assets}$, measures the proportion of intangible assets, such as goodwill and intellectual property, to total assets. The *debt to equity ratio*, defined as $\frac{Total\ debt}{Shareholders'\ equity}$, is an indication of risk, financial constraints, and a firm's ability to raise and service debt. We use *employees to total assets*, defined as $\frac{Number\ of\ employees}{Total\ assets}$, as a normalized measure of the labor intensity of a firm. The *market to book ratio*, defined as $\frac{Market\ value\ of\ equity}{Book\ value\ of\ equity}$, is used as a measure of market expectations towards growth. Further, we use the 4-digit *global industry classification* (GICS) and the *country* where the company is headquartered. Lastly, the logarithm of *market capitalization* and *total assets* are used as proxies for firm size.

All variables that are reported in currency have been converted into U.S. dollars for comparability. We do not convert variables prior to calculating ratios, as the levels do not matter for the relative values. We convert EPS and sales forecasts on the day of the forecast publication. For price data, we convert by the exchange rate on the day of the closing price quote, and for fundamentals, we convert at the end of the fiscal year.

3.4 Summary Statistics

Table 3.2 reports summary statistics of all dependent variables. Only observations for the top 5% (treated) and bottom 50% (control) of firms ranked by AI-exposure are included. We observe that there are markedly fewer observations for $EPS_{i,t,h}$, $SAL_{i,t,h}$ and $GRM_{i,t,h}$ for 4y, 5y and LTG horizons, which can have implications for the statistical power of models where these variables are analyzed.

Table 3.3 reports summary statistics of control variables in Panel A as well as country and industry distributions in Panel B. The distributions in Panel B display only the treated group's top 5 countries and industries for the sake of readability. Unsurprisingly, the industry distribution of treated firms leans heavily towards information technology. In addition, the sample consists mostly of U.S. firms for which data availability in the utilized databases is the greatest. Table 3.3 illustrates that there are some observable differences between the groups, particularly for industry classification, market to book, and debt to equity. The implications of such differences will be discussed in the analysis section, where we also construct a matched sample based on propensity score matching.

Table 3.2: Summary statistics of dependent variables

	Control					Treated				
	Obs	Mean	SD	Min	Max	Obs	Mean	SD	Min	Max
EPS										
1y	25759	-0.03	0.52	-3.56	0.33	2536	-0.10	0.54	-3.56	0.33
2y	25563	0.02	0.33	-2.28	0.31	2521	-0.05	0.34	-2.28	0.31
3y	20934	0.07	0.20	-1.29	0.44	2265	0.00	0.21	-1.29	0.44
4y	12034	0.13	0.42	-1.09	2.59	1280	0.07	0.41	-1.09	2.59
5y	7721	0.16	0.40	-0.76	2.60	906	0.06	0.28	-0.76	2.60
LTG	7732	13.26	17.96	-25.10	70.53	1258	17.52	17.52	-25.10	70.53
SAL										
1y	25743	7.02	1.99	-1.62	10.81	2538	6.66	2.52	-1.62	10.81
2y	25553	7.16	1.85	0.34	10.86	2523	6.91	2.26	0.34	10.86
3y	21052	7.47	1.76	1.36	11.01	2295	7.29	2.07	1.36	11.01
4y	12536	7.73	1.78	1.89	11.31	1556	7.66	2.08	1.89	11.31
5y	8728	7.89	1.76	2.72	11.52	1190	7.92	2.12	2.72	11.52
LTG	5197	10.90	11.27	-2.05	54.50	1094	18.30	11.44	-2.05	54.50
GRM										
1y	18447	38.89	22.47	1.45	98.67	2390	62.59	21.57	1.45	98.67
2y	18259	40.17	21.86	7.70	96.30	2385	64.17	20.50	7.70	96.30
3y	13836	42.00	21.87	9.46	96.30	2077	66.39	19.18	9.46	96.30
4y	6911	46.03	22.53	12.00	100.00	1157	70.76	17.64	12.00	100.00
5y	4012	49.82	22.91	12.75	100.00	825	71.67	17.85	13.00	100.00
REC										
REC	25745	2.23	0.49	1.00	5.00	2531	2.14	0.38	1.00	4.00

Table 3.2 reports summary statistics for dependent variables, grouped by treatment status. The dependent variable for 1y - 5y EPS is $EPS\ norm_{i,t,h} = \frac{EPS_{i,t,h}}{P_{i,\ t}}$, which is EPS forecasts scaled by share price (earnings yield). i, t, and h denote the firm, the month of the forecast, and the horizon, respectively. For SAL, GRM, REC, and LTG, the dependent variables are $\log(SAL_{i,t,h})$, $GRM_{i,t,h}$, $LTG_{i,t,h}$ and $REC_{i,\ t}$, respectively. The sample consists of data from November 30$^{\text{th}}$, 2021 until May 31$^{\text{st}}$, 2023. All variables except $REC_{i,\ t}$ are winsorized at the 95% level.

Table 3.3: Summary statistics of control variables

Panel A: Control variables

	Control					Treated				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Intan./Total Assets	26,291	0.20	0.21	0.00	0.73	2,612	0.26	0.22	0.00	0.73
Debt to Equity	26,291	0.99	1.67	−3.73	6.98	2,612	0.67	1.56	−3.73	6.98
Employee Ratio	26,291	0.23	0.28	0.00	1.32	2,612	0.25	0.32	0.00	1.32
Market to Book	26,291	2.86	4.45	−7.43	29.34	2,612	6.08	7.35	−7.43	29.34
log(Total Assets)	26,291	7.66	1.82	3.20	11.43	2,612	7.37	2.15	3.20	11.43
log(Market Cap.)	26,291	7.39	1.94	1.68	12.55	2,612	7.92	2.56	1.68	12.55

Panel B: Country and industry distributions (% of sample)

	Country			Industry	
	Control	Treated		Control	Treated
Information Technology	6.95%	56.38%	USA	75.08%	83.89%
Telecommunication Services	2.70%	15.44%	Israel	0.47%	5.37%
Industrials	19.58%	10.74%	France	1.28%	1.34%
Health Care	10.87%	8.05%	Great Britain	1.82%	1.34%
Consumer Discretionary	15.33%	4.70%	Netherlands	0.47%	1.34%

Table 3.3 displays summary statistics for control variables grouped by treatment status. The sample consists of data from November 30th, 2021 until May 31st, 2023. In Panel A, the employee ratio is defined as $\frac{Number\ of\ employees}{Total\ assets}$ and is a normalized measure of the labor intensity of a firm. All control variables in Panel A have been winsorized at the 95% level. Panel B reports country and industry distributions of firms as percentages of the total sample for each group. For the sake of readability and presentation, we report the 2-digit global industry classification (GICS) in this table. The 4-digit industry code is used in the analysis. The country/industry distributions display only the treated group's top 5 countries and industries.

4 Measuring AI-Exposure

As previously introduced, AI-exposed firms are defined as firms that develop and/or extensively utilize AI in business processes and product offerings. Thus, this analysis aims to identify firms where AI constitutes an integral part of the company.

Measuring AI-exposure is achieved by counting how often AI-related bigrams appear in earnings call transcripts. Therefore, we first need to define AI-related bigrams. Whereas a common strategy is to define a list of words known to be related to a topic, it does not provide much nuance or information about how related the word is to the topic. We apply two different weighting techniques to quantify how related a bigram is to AI. First, we apply a pre-trained NLP model to assess a bigram's similarity to the bigram "artificial intelligence". Next, we weigh each bigram by how frequently it occurs in a corpus consisting of AI text relative to a corpus of earnings call text. As a result, each bigram has a score, or weight, which is applied to the bigram count. For instance, our method finds that "neural network" and "general intelligence" are very important bigrams in the AI context. Thus, if a firm mentions these bigrams in an earnings call, it will add a lot to the final AI-exposure score.

This section details how we measure AI-exposure, from quantifying each bigram's relatedness to AI to calculating AI-exposure for each firm and displaying the results.

4.1 Defining AI Bigrams

4.1.1 Constructing an AI Corpus

The first step of identifying AI-related bigrams involves gathering data for an AI corpus by assembling a diverse range of AI texts. Emphasis is placed on selecting texts that address AI from a broad perspective to ensure a balanced representation of AI branches and avoid industry-specific bias. The corpus primarily includes academic papers encompassing various domains within the application of artificial intelligence. Although academic papers offer a wealth of technical and theoretical insights, they often employ a formal and complex style of writing. Therefore, they might not fully encapsulate the conversational and practical nuances prevalent in earnings call discussions about AI. To address this, we integrate data from non-academic sources such as websites, blogs, and online forums, where language tends to be more accessible and conversational. For the full reference list to the AI corpus, see Appendix A.

We perform a series of standard data preprocessing procedures from the NLP literature to convert the text into bigrams. First, we clean the text for any non-ASCII characters. Next, we remove punctuation and create bigrams by merging two consecutive words. Then, we lemmatize both words in the bigram, which is the process of reducing a word to its base form. For example, "running" and "ran" are both reduced to the lemma "run". This serves the purpose of grouping different words that convey the same meaning so that they can be analyzed as one unit. Lastly, we remove words that do not convey any useful information for the analysis[1]. After data preprocessing, the final AI corpus consists of $\sim 29,000$ unique bigrams.

4.1.2 Weighing Bigrams by Cosine Similarity

Although bigrams are extracted from AI texts, they are not equally related to AI. To address this, we assign a weight to each bigram that represents its similarity to the bigram "artificial intelligence". This weight is the cosine similarity, which is a mathematical expression of how similar two vectors are to each other (Han et al., 2012). However, bigrams are pieces of text and not vectors. Thus, the first step in calculating cosine similarities is to generate word embeddings, which is the numeric representation of text in vector form.

Generating Word Embeddings

For the cosine similarity to accurately represent how similar two bigrams are to each other, it is crucial that the embeddings correctly represent the bigrams both in meaning and context. Therefore, we utilize an NLP model trained on vast amounts of data called BERT, which is proven to yield accurate results on various NLP tasks (Devlin et al., 2018). BERT uses bidirectional transformers, which means that it can understand the context of words in a sentence by looking at the words that come before and after them. This is an important feature for our purpose because it enables the model to understand the relationship between the two words within a bigram and interpret it as one entity rather

[1]We remove all proper nouns and determinants. We also eliminate all bigrams containing non-alphabetic letters, one-character words, words that are not in the English dictionary, and stop-words.

than two separate words. We employ a version of BERT called SciBERT. This model has the same architecture as BERT, but it is trained specifically on scientific texts, making it ideal for analyzing our corpus consisting mostly of academic papers (Beltagy et al., 2019).

The vector representations (embeddings) of all bigrams constitute a vector space. SciBERT creates embeddings so that similar bigrams lie close to each other in the vector space, which is crucial for calculating cosine similarities in the next step.

Cosine Similarities

Cosine similarity measures the similarity between two vectors as $cos(\theta)$, or the cosine of the angle θ between them[2]. Vectors that lie close to each other in the vector space point in a similar direction. Hence, the angle between them is small. Since $cos(0°) = 1$ and $cos(90°) = 0$, the smaller the angle, the larger the cosine similarity. Thus, if a bigram is similar to "artificial intelligence", indicated by its position relative to "artificial intelligence" in the vector space, it will receive a high weight, which is what we want to accomplish.

To further illustrate the logic behind this measure, consider the simplified representation of the embedded forms of "artificial intelligence", "machine intelligence" and "make sure" in Figure 4.1. The embeddings of "artificial intelligence" and "machine intelligence" lie close to each other in the vector space. Hence, the angle θ between them is small, and the similarity is high. On the contrary, the embedding of "make sure" lies far from the embedding of "artificial intelligence". Therefore, the angle θ is greater and the similarity lower.

[2]Given two n-dimensional vectors, A and B, the cosine similarity is mathematically expressed as the dot product of vectors divided by the product of their lengths, or the cosine of the angle θ between them (Han et al., 2012): $\cos(\theta) = \frac{A \cdot B}{||A||||B||}$

Figure 4.1: Illustration of Cosine Similarity

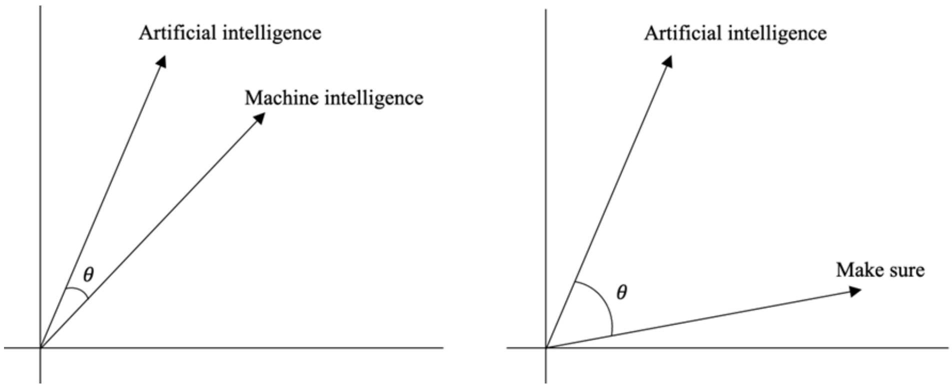

Figure 4.1 is a simplified illustration of embeddings (vector representations of text). The smaller the angle θ between two vectors, the greater the cosine similarity. Thus, "artificial intelligence" is (cosine) similar to "machine intelligence" and not similar to "make sure".

The cosine similarity assigns a higher weight to bigrams closely related to AI. Yet, the difference in the similarity score from the most related bigrams to less related bigrams is not large enough. Therefore, the cosine similarity does not sufficiently distinguish between AI discussions and non-AI discussions. Consider the example of two earnings call text snippets in Table 4.1. The text from Alphabet is clearly a discussion about AI, and the text from Vivos Therapeutics is clearly not (SeekingAlpha, 2022, 2023).

Table 4.1: Earnings call text snippets from Alphabet (clearly about AI) and Vivos Therapeutics (clearly not about AI)

Alphabet (Oct. 2023, Q3):	Vivos Therapeutics (Dec. 2022, Q3):
"This includes our work with the Search Generative Experience, which is our experiment to bring Generative AI capabilities into Search. We have learned a lot from people trying it, and we have added new capabilities, like incorporating videos and images into responses and generating imagery. We've also made it easier to understand and debug generated code."	"Let's move on for a moment now to home sleep test and case starts, which are both key performance metrics. The VivoScore home sleep test we offer to our VIPs from SleepImage are a core advantage because they offer an easy and affordable way for patients to obtain a clinically accurate and diagnostic quality assessment of their breathing and sleep."

Table 4.1 displays text snippets from two earnings call transcripts, one from Alphabet and one from Vivos Therapeutics. The Alphabet text is clearly a discussion about AI, and the Vivos Therapeutics text is not.

Calculating the weighted cosine similarity of both text snippets, they receive scores of 0.426 (Alphabet) and 0.193 (Vivos Therapeutics). Thus, the text about AI has a score that is 121% ($\frac{0.426}{0.193} - 1$) higher than the text not about AI. Considering that the exposure measure aims to capture AI-related discussions, we argue that the difference should be greater. The confounding impact arises due to bigrams such as "core advantage", "quality assessment" and "performance metrics" from the Vivos Therapeutics text being part of our AI corpus. Since regular cosine similarities do not sufficiently weigh down these bigrams, we perform a mathematical transformation to the cosine similarity. Specifically, we apply a combination of squaring, which increases the distance between high and low numbers, and an inverse logarithmic transformation as follows: $\left(-\log\left(1 - cos(\theta)_b{}^2\right)\right)^2$. Figure 4.2 displays the regular and transformed cosine similarity distributions. Clearly, the mathematical transformation yields a greater difference between the most AI-related bigrams and less AI-related bigrams.

With the transformation, the weighted cosine similarities are 0.285 (Alphabet) and 0.069 (Vivos Therapeutics). Thus, the text about AI has a 313% ($\frac{0.285}{0.069} - 1$) higher similarity than the non-AI text, a considerable increase from 121% without the transformation. We argue that the mathematical transformation of similarities is an important step for our measure to capture discussions that are actually related to AI.

Figure 4.2: Distributions of regular and transformed cosine similarities

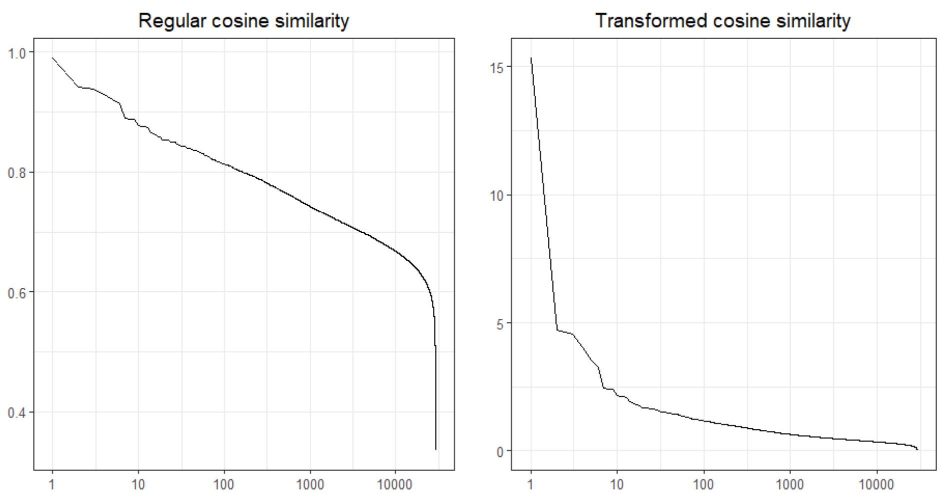

Figure 4.2 illustrates the distributions of the regular cosine similarity (left) and the mathematically transformed cosine similarity (right). The X-axes are logarithmic. The mathematical transformation increases the relative weight of the most AI-related bigrams.

4.1.3 Weighing Bigrams by Relevance to the AI Corpus

Although the transformed cosine similarity weighs down bigrams that are not very related to AI, there are still some bigrams that could create noise in the estimates. Some bigrams are mentioned so frequently in earnings calls that, even though they have low cosine similarities, they matter a lot for the final exposure scores. For instance, consider "make sure" and "machine intelligence", which have transformed cosine similarities of 0.12 and 4.54, respectively. "make sure" occurs 71,143 times in the earnings call sample, giving it a total score of 8,537.16 ($71,143 \times 0.12$), whereas "machine intelligence" occurs only 30 times, giving it a total score of 136.20 (30×4.54). Since "make sure" is such a common phrase, it has a disproportionately large impact on exposure scores compared to its cosine similarity. This creates noise in the estimates if some companies use "make sure" or other frequently occurring bigrams more often than others. In addition, bigrams like "make sure" are presumably more representative of general text than AI text. Therefore, they should not have such a large impact on the final AI-exposure.

To address this, we apply an additional weighting based on how related each bigram is to the AI corpus. The weight is calculated by comparing how often a bigram occurs in the AI corpus relative to a corpus consisting of earnings call text. The earnings call corpus is constructed by randomly selecting 1,000 transcripts that are not in our company sample. We define the relative frequency of bigram b in the AI corpus as $RF_b^{AI} = \frac{n_b^{AI}}{N^{AI}}$, where n_b^{AI} is the number of occurrences of the bigram and N^{AI} is the total number of bigrams in the AI-corpus. Similarly, we define $RF_b^{EC} = \frac{n_b^{EC}}{N^{EC}}$ as the relative frequency of bigram b in the earnings call corpus. The weight of each bigram is calculated as $w_b = \frac{RF_b^{AI}}{RF_b^{EC}}$.

To illustrate the effect of the relative frequency weighting, consider again the bigrams "artificial intelligence" and "make sure". In relative terms, "artificial intelligence" appears more often in the AI corpus than in the earnings call corpus. Therefore, $RF^{AI}_{artificial\ intelligence} > RF^{EC}_{artificial\ intelligence}$ and the bigram will be assigned a high weight. On the contrary, "make sure" is a common term in earnings call transcripts but seldom appears in AI-related text. Hence, $RF^{AI}_{make\ sure} < RF^{EC}_{make\ sure}$ and the bigram will be assigned a low weight.

4.1.4 AI Bigram Results

By multiplying the transformed cosine similarity and the relative frequency weight, the final weight of each bigram can be expressed as follows:

$$BW_b = w_b \times \left(-\log\left(1 - {cos(\theta)_b}^2\right)\right)^2 \tag{4.1}$$

Where $w_b = \frac{RF_b^{AI}}{RF_b^{EC}}$, as outlined in the previous section, and $cos(\theta)_b$ is the cosine similarity of bigram b. BW_b expresses each bigram's relevance to AI. Table 4.2 displays the top 10 and bottom 10 bigrams with their respective bigram weights. The top list contains bigrams that are all highly related to AI, such as "neural network" and "general intelligence". The bottom list contains bigrams that likely occur very frequently in earnings calls but seldomly in AI-related text, indicating that the relative frequency weighting successfully reduced the impact of bigrams that are not representative of AI.

Table 4.2: Top and bottom 10 bigrams

Top 10 Bigrams		Bottom 10 Bigrams	
Bigram	BW_b	Bigram	BW_b
neural network	5,396	additional information	0.013
artificial intelligence	3,073	please note	0.011
machine intelligence	2,696	primarily due	0.011
artificial neural	2,389	make sure	0.009
general intelligence	2,382	second half	0.007
expert system	1,817	go forward	0.006
fuzzy logic	1,492	will like	0.004
unsupervised learning	1,426	little bit	0.002
human intelligence	1,366	next question	0.002
artificial general	1,104	last year	0.001

Table 4.2 displays the top 10 and bottom 10 bigrams ranked by bigram weight. The bigram weight indicates how related each bigram is to AI.

4.2 AI-Exposure Results

The bigram weights are applied in the calculation of firm-level AI-exposure, which proceeds in three steps. First, we combine all transcripts for a given firm between November 30th, 2021, and the release of ChatGPT on November 30th, 2022. Second, in the compiled set of transcripts, we count the number of occurrences of all bigrams that appear in the AI corpus. In this step, each bigram is counted as BW_b to reflect how related it is to AI. We express this weighted count as BC_i, which is the bigram score for firm i. Finally, we divide by the total number of bigrams in the firm's set of transcripts, N_i. Thus, the firm-level AI-exposure measure can be expressed as follows:

$$AI\ exposure_i = \frac{BC_i}{N_i} \tag{4.2}$$

Table 4.3 displays the top 20 companies ranked by AI-exposure. The industry distribution of the most exposed firms is concentrated on Information Technology and Telecommunication Services, which is in line with the general perception of typical AI industries. However, there are deviations from this trend, where Health Care and Industrials both have two companies represented. The diversity of industries validates the capacity of our approach to capture a wide spectrum of AI integration across industries.

Many firms on the list are widely known for utilizing and/or developing AI to a large extent, which indicates the validity of our approach. Some notable examples include Meta Platforms and Alphabet, where both companies have released large language models similar to ChatGPT, namely Llama 2 (Meta, n.d.) and Bard (Pichai, 2023). Alphabet is also the company behind the NLP model BERT, which was utilized to generate the bigram embeddings in this paper. At the top of the list stands Nvidia, one of the companies that developed the very infrastructure that ChatGPT relies on (Nguyen et al., 2023).

Table 4.3: Top 20 companies ranked by AI-exposure

Company Name	AI-Exposure	Industry	Country
NVIDIA Corporation	95.77	Information Technology	USA
C3.ai, Inc.	67.55	Information Technology	USA
LivePerson, Inc.	53.79	Information Technology	USA
Shutterstock, Inc.	53.17	Telecommunication Services	USA
RadNet, Inc.	48.32	Health Care	USA
Genpact Limited	40.87	Industrials	Bermuda
Galmed Pharmaceuticals Ltd.	40.04	Health Care	Israel
Veritone, Inc.	34.73	Information Technology	USA
Alphabet Inc.	34.18	Telecommunication Services	USA
Ambarella, Inc.	32.55	Information Technology	USA
TaskUs, Inc.	31.22	Industrials	USA
Verint Systems Inc.	29.08	Information Technology	USA
Meta Platforms, Inc.	27.37	Telecommunication Services	USA
Cadence Design Systems, Inc.	27.10	Information Technology	USA
Salesforce, Inc.	26.46	Information Technology	USA
Getty Images Holdings, Inc.	26.40	Telecommunication Services	USA
Oracle Corporation	25.34	Information Technology	USA
Pegasystems Inc.	24.56	Information Technology	USA
Taboola.com Ltd.	24.07	Telecommunication Services	USA
Sprinklr, Inc.	23.95	Information Technology	USA

Table 4.3 displays the top 20 companies ranked by AI-exposure. AI-exposed firms refer to companies that develop and/or extensively utilize AI in their business processes and product offerings. AI-exposure is measured by quantifying how related earnings call discussions are to AI.

5 Analysis

5.1 Methodology

To analyze the effect of the ChatGPT release on analyst expectations towards AI-exposed firms, we estimate a difference-in-differences (DiD) model. The DiD framework is ideal for testing our hypotheses because it allows us to compare firms with high AI-exposure to firms with low exposure over time. Through estimating the otherwise unobservable counterfactual, which is what would have happened to AI-exposed firms absent the release of ChatGPT, the method attempts to estimate the causal effect of treatment. This is only a valid approach if the identifying assumptions behind DiD hold, which will be discussed in detail in section 5.2.1.

We define treatment as the release of ChatGPT and the treated group as the top 5% of firms ranked by AI-exposure (149 firms). As previously argued, these firms possess advanced AI infrastructure and expertise, making them advantageously positioned to leverage technological developments within the realm of AI. In addition, selecting a small percentile ensures that all the firms in the treated group have considerable exposure. Further, we select the bottom 50% of AI-exposed firms as a preliminary control group (1,481 firms). Omitting semi-exposed companies ensures that the exposure gap between groups is sufficiently large.

We use two main specifications in our analyses. The first is the standard DiD model, as expressed in equation 5.1.

$$DepVar_{i,t} = \beta_0 + \beta_1(Treat_i \times POST_t) + \beta_2 Treat_i + \beta_3 POST_t + X_{i,t} + a_i + m_t + u_{i,t} \quad (5.1)$$

$DepVar_{i,t}$ is the dependent variable for firm i in month t. $X_{i,t}$ is a vector of controls exogenous to treatment, a_i are firm fixed effects, and m_t are year-month fixed effects. $Treat_i$ is a dummy variable equal to 1 for treated firms and 0 for control firms. Similarly, $POST_t$ is equal to 1 in months after the release of ChatGPT and 0 in months before. The coefficient of interest is β_1, which measures the average change in the dependent variable for the treated group after release, relative to the control group.

The vector $X_{i,t}$ consists of time-variant controls that are exogenous to treatment and likely to correlate with both treatment and outcome trends. *Debt to equity* is likely to affect treatment because firms with a higher proportion of debt are more financially constrained and have less free capital to invest in AI. A high *intangible asset ratio* can suggest a culture of innovation and a strategic focus on technology, which are prerequisites for AI integration and development. We use log(*total assets*) to control for firm size, as larger firms are presumably better positioned to acquire the resources necessary to develop AI, either through talent acquisition or investments. Finally, we control for *employees to total assets*, which is a normalized measure across firms indicating labor intensity. A high labor intensity indicates a broad pool of skill and expertise and a large scale on which AI can be implemented, both of which could make a firm inclined to seek out AI solutions. Since these controls are comprised of balance sheet values and factors that are presumably relatively static, they are not likely to be affected by treatment. This is especially true since we study short-term effects up to six months after the release.

To further account for omitted variable bias, we include a set of fixed effects in our models. First, we include firm fixed effects to control for firm-level differences that are constant throughout our study's time frame. Whereas some factors are observed and can be controlled for, such as industry classification, other factors, such as corporate culture or the regulatory environment of the firm, are difficult to measure. Firm fixed effects control for all constant factors, whether they are observed or unobserved. In addition, we include year-month fixed effects to control for factors that affect all firms equally over time, such as economy-wide macroeconomic shocks. While $POST_t$ controls for such factors, it only considers two time points: before and after release. Including year-month fixed effects allows controlling for these effects with greater granularity, which we argue to be important considering the rapidly changing macroeconomic environment before and after the release of ChatGPT.

We cluster standard errors at the firm level to account for heteroskedasticity and potential serial correlation within firms. Serial correlation can arise due to persistent firm-specific shocks as well as aspects of analyst behavior studied in the literature. For example, analyst herding behavior has been suggested in several studies (See for example Clement and Tse (2005) or Trueman (1994)), which implies that firm-level forecast revisions are correlated with changes in prior consensus forecasts.

In the second specification, expressed in equation 5.2, we expand the standard DiD model to include six post-treatment dummy variables. When interacted with treatment, the coefficients are estimates of the average treatment effect for AI-exposed firms in a specific month after release. This approach is helpful for better understanding the timing of analyst reactions and how the effects of the release unfolded over time.

$$\begin{aligned} DepVar_{i,t} = \beta_0 + \beta_1(Treat_i \times POST1_t) + \beta_2(Treat_i \times POST2_t) + \ldots + \\ \beta_6(Treat_i \times POST6_t) + \ \beta_7 Treat_i + \beta_8 POST1_t + \beta_9 POST2_t \\ + \ldots + \ \beta_{13} POST6_t + X_{i,t} + \alpha_i + m_t + u_{i,t} \end{aligned} \tag{5.2}$$

The post-treatment dummy variables are 1 in the specific month after release denoted by the number in the variable name, and 0 otherwise. For example, $POST1_t = 1$ in December 2022, which is the first month following the release.

5.2 Analyst Recommendations

Table 5.1 illustrates the impact of the ChatGPT release on analyst recommendations for treated firms relative to the control group. We use both the standard DiD model and the model with a segmented post-treatment period, progressively adding fixed effects to illustrate the impact. Adding firm fixed effects has some effect on the estimates, although it is minimal. Adding year-month fixed effects has a negligible impact on the models, which suggests that most of the effects from common shocks are already picked up by the $POST_t$ dummy variables. The coefficient on the common DiD model (3) with all fixed effects is 0.028, statistically significant at the 5% level, indicating that the average effect of being AI-exposed after the release of ChatGPT was an increase in recommendations. Considering the scale goes from 1 to 5, where 1 corresponds to “strong buy” and 5 to “sell”, the effect of the ChatGPT release was a downgrade in recommendations for AI-exposed firms. However, the magnitude is small, of only 2.8% of a full step downgrade. The segmented post-treatment model (3) tells the same story but with more nuance. Specifically, coefficients increase in magnitude and statistical significance with time. This can indicate (1) that analysts took time to react and that the release had a negative impact or (2) that other factors that evolved over time influenced recommendations.

Table 5.1: Analyst recommendations regression results

	Full post-treatment			Segmented post-treatment		
	Recommendations			Recommendations		
	(1)	(2)	(3)	(1)	(2)	(3)
Treat x After	**0.034***	**0.028***	**0.028***			
	(0.015)	(0.014)	(0.014)			
Treat x Dec. 22				0.005	0.006	0.006
				(0.010)	(0.010)	(0.010)
Treat x Jan. 23				0.003	0.003	0.004
				(0.013)	(0.013)	(0.013)
Treat x Feb. 23				0.024	0.016	0.016
				(0.016)	(0.015)	(0.015)
Treat x Mar. 23				**0.040***	**0.032.**	**0.032.**
				(0.019)	(0.018)	(0.018)
Treat x Apr. 23				**0.052***	**0.046***	**0.047***
				(0.022)	(0.020)	(0.020)
Treat x May 23				**0.079****	**0.067****	**0.068****
				(0.024)	(0.023)	(0.023)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE		Yes	Yes		Yes	Yes
Year-month FE			Yes			Yes
Num.Obs.	19 009	19 009	19 009	19 009	19 009	19 009
R2 Adj.	0.045	0.894	0.895	0.045	0.895	0.895

Clustered (firm) standard errors in parathesis
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, .:0.1*

Table 5.1 displays regression output for two types of models, where fixed effects are added progressively from model (1) to model (3). The first type is the standard DiD model with one POST variable. The second type splits the post-treatment period, displaying the average treatment effect in a specific month after release. We control for *debt to equity*, *intangible to total assets*, *employees to total assets* and log(*total assets*). Control variables are winsorized at the 95% level. The sample consists of [t-6, t+6] months of forecast data, where t marks the date of the ChatGPT release (November 30$^{\text{th}}$, 2022).

A negative reaction would indicate that the ChatGPT release contained information that analysts deemed negative for AI-exposed firms. Under the assumption that the event constituted a major technology shock to AI, where AI-exposed firms have a competitive advantage in leveraging the shock, this appears unlikely. Rather, it could be that other factors impacted stock recommendations in the period. As described previously, two important inputs for a stock recommendation are a firm's projected earnings and its current stock price. Therefore, one explanation for why recommendations increasingly moved towards “sell” could be that prices went beyond what analysts considered fundamental value. If this is true, the increasing magnitude of coefficients after release suggests that analysts became firmer in this belief over time.

Figure 5.1 illustrates market capitalization-weighted indices of the treated group versus the control group, both indexed to 0 at the date of the ChatGPT release. Although this is a basic calculation based on monthly share prices from Compustat, it indicates the trajectory of stock prices post-release. The index of AI-exposed firms surpassed the index of the control group in January 2023, and by July it had increased by more than 10% since the release. Compared to the more neutral development of the control group index, the trajectory of stock prices points to the possibility that analysts considered AI-exposed stocks to be overvalued.

Figure 5.1: Stock price development of treated and control groups

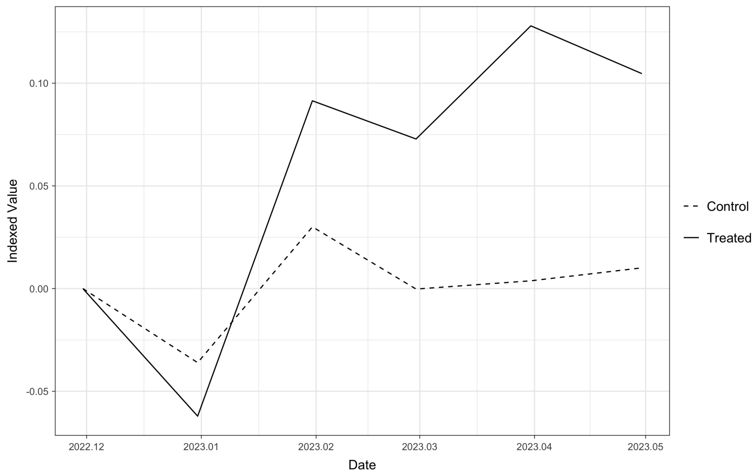

Figure 5.1 illustrates market cap.-weighted indices grouped by treatment status. Both indices are indexed to 0 at the date of the ChatGPT release (November 30$^{\text{th}}$, 2022). Indices are calculated as the sum of market capitalizations for all firms in each group on a monthly basis, converted to USD. The graph is from the date of the ChatGPT release and six months ahead.

5.2.1 Identifying Assumptions

The validity of the estimates depends upon adherence to the identifying assumptions behind a DiD model. The true effect of the release is the actual outcome of analyst expectations towards AI-exposed firms, minus the outcome had the release not happened. The latter is the counterfactual, which is impossible to observe. However, the counterfactual is estimated in a DiD model as the trend of the control group. Therefore, a crucial assumption is that analyst expectations towards treated and control firms would have followed a similar path absent treatment. If trends in the pre-treatment period are parallel, it indicates that this assumption is valid. We check the parallel trends assumption by visually inspecting analyst expectations through time. Figure 5.2 plots the mean of selected dependent variables for the control- and treated group from 12 months before to 6 months after release. Plots for all dependent variables can be found in Appendix B.1.1.

Figure 5.2: Parallel trends visualization

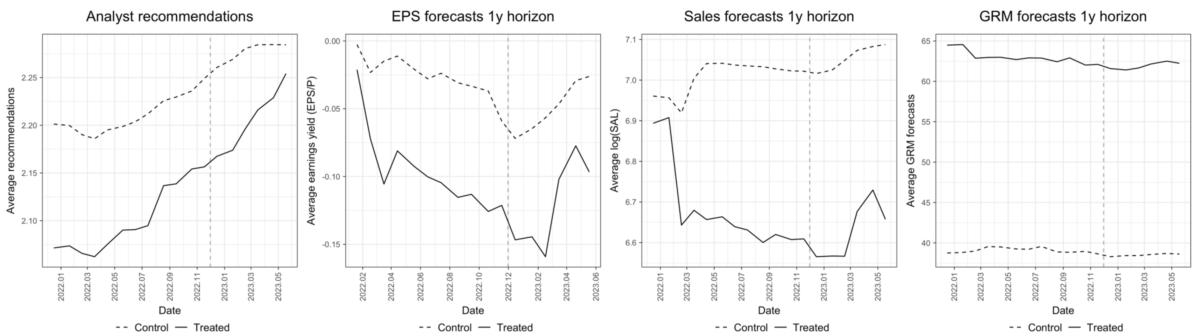

Figure 5.2 illustrates the mean of selected dependent variables, grouped by treatment status. The plots include [t-12, t+6] months of data, where t marks the date of the ChatGPT release (November 30^{th}, 2022). Except for recommendations, all dependent variables have been winsorized at the 95% level.

The recommendations plot exhibits parallel trends to some degree, although the treated group has a higher growth rate than the control group. For EPS forecasts, trends are not parallel in the first months of 2022 but more so in the months leading up to the release. Sales forecasts have a similar development, although trends diverge in the beginning before leveling out, which is a more serious violation of parallelism than what is observed for EPS. The assumption seems mostly fulfilled for gross margins, although there are similar issues as for EPS and sales in the first months of 2022. To avoid these observations from biasing our estimates, we use only data from six months prior to treatment in our analyses. Overall, parallel trends seem only partly fulfilled by visual inspection.

Further, for the parallel trends assumption to hold, there can be no unobserved confounders that correlate with both treatment and the outcome variables. We include fixed effects and control variables to eliminate such effects. However, year-month fixed effects only absorb the effects of shocks that are common to all firms, and there might still be confounding factors that affect the treated and control groups differently. An example of such a factor could be changing interest rates, where the effect will depend on factors such as firm-level risk and the income elasticity of demand for a firm's products.

Lastly, DiD requires homogenous treatment effects, meaning treatment has to affect the treated group similarly. Although the exposure measure is a continuous scale, where for instance Nvidia has a higher score than Meta, we argue that what matters is having a certain level of AI-exposure rather than the absolute level. Firms above a certain exposure threshold are likely to have the required AI capabilities to benefit from the technology shock regardless of the exact exposure level. Additionally, by picking a small percentile, we have ensured that all firms in the treated group are significantly exposed to the effects of the ChatGPT release.

5.2.2 Constructing a Matched Control Group

We construct a matched control group using nearest-neighbor propensity score matching to address the possibility of the identifying assumptions being violated. A propensity score can be understood as a firm's probability of being in the treated group, conditional on observed pre-treatment characteristics. The nearest neighbor algorithm assigns control firms to treated firms to minimize the propensity score difference between them (Austin, 2014). Thus, the groups will be similar on all relevant observed characteristics *except* for AI-exposure. By balancing pre-treatment covariates, the method attempts to mimic the random assignment of a controlled experiment and overcome the selection bias.

We perform 1:3 matching with replacement, meaning that each treated firm is matched with 3 control firms and that one control firm can be matched with several treated firms. We only consider the bottom 50% of AI-exposed firms for matching. As previously argued, this ensures a sufficiently large exposure gap between groups. 1:n matching is thought to improve statistical precision by preventing too many firms from being dropped from the sample. However, setting n too high can lead to bias from lower match quality (Rassen et al., 2012). We perform 1:3 matching to balance these considerations. Since the pool of potential matches is large (1,481), matching quality is likely to be preserved at an appropriate level.

The first step of the matching process is to identify firm-level characteristics that explain the probability of being treated (AI-exposed). This includes *debt to equity*, *intangible to total assets*, and *employees to total assets*, where the assumed correlation with treatment is already explained. In addition, we hypothesize that *market capitalization* could predict treatment, as larger firms are presumably more attractive to high-skilled talent and have more resources to invest in AI development. *Market to book* reflects the market's recognition of growth potential, which is often large for innovative firms with forward-looking strategies that could include the integration of AI. Further, *industry classification* could be an important predictor because firms in certain industries are more inclined towards AI adoption, for example due to competitive pressure. Lastly, countries that are leading in AI research have a greater availability of AI-related resources, such as high-skilled human capital and product components. Therefore, the *country* of a firm's headquarters could also explain the probability of being treated.

We run logit regression to test the effect of these variables on the probability of treatment, expressed in Equation 5.3. Only pre-treatment values are included so potential treatment effects on covariates are not present. For time-varying covariates, we use the latest recorded value before the release of ChatGPT.

$$\begin{aligned} Treat_i = \beta_0 &+ \beta_1 Debt\ to\ equity_i + \beta_2 Intan.\ ratio_i + \beta_3 Employee\ ratio_i \\ &+ \beta_4 \log(Market\ cap.)_i + \beta_5 Market\ to\ book_i + \beta_6 Industry_i \\ &+ \beta_7 Country_i \end{aligned} \tag{5.3}$$

Table 5.2 displays the results from three logit regressions. The first model includes all characteristics except for industry and country, which are progressively added in the second and third models. The coefficients on several characteristics change markedly from model (1) to (2), suggesting that industry has a profound impact on the probability of being treated. Including the country variable in model (3) has a smaller impact, although there are some changes in magnitude and statistical significance. As expected, market to book, market capitalization, and debt to equity are significant predictors of treatment. The signs of coefficients are also in line with assumptions, where for example a one-unit (100%) increase in the market to book ratio is associated with a 0.083 increase in log-odds of being treated. In model (1), a higher intangible asset ratio is associated with an increase in treatment probability as expected. However, in model (2), the sign of the coefficient reverses, and statistical significance is reduced. This implies that the coefficient on intangible assets in model (1) picks up variation that is attributed to industry in model (2), which can be considered logical due to wide disparities in intangible asset levels across industries. It also implies that industry is an important omitted variable in model (1) that must be included in the matching process.

Table 5.2: Logit regressions on firm-level characteristics

	Treated		
	(1)	(2)	(3)
Market to book	**0.145*****	**0.076****	**0.083****
	(0.019)	(0.028)	(0.029)
Intangible to total assets	**1.429*****	**-0.870.**	**-0.922.**
	(0.396)	(0.527)	(0.557)
Log(Market Cap.)	−0.004	**0.130***	**0.157****
	(0.045)	(0.055)	(0.058)
Debt to equity	**-0.393*****	**-0.159.**	**-0.162.**
	(0.074)	(0.091)	(0.095)
Employees to assets	0.083	−0.632	−0.444
	(0.316)	(0.407)	(0.472)
Industry variable		Yes	Yes
Country variable			Yes
Model	Logit	Logit	Logit
Num.Obs.	1630	1630	1630

Clustered (firm) standard errors in parathesis
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, .:0.1*

Table 5.2 displays regression output for three logit models with treatment as the dependent variable and pre-treatment characteristics as independent variables. Model (1) includes all variables that are visible in the table. Models (2) and (3) add industry and country variables, respectively. The coefficients on these variables are removed for the sake of presentation and readability. Control variables that are continuous are winsorized at the 95% level. Variance inflation factors are low in all models, at most 3.71 for the industry variable in model (3). The industry variable is the 4-digit global industry (GICS) code.

We match on all characteristics described above except for employees to total assets since it is not a significant predictor of treatment. We use the MatchIt (Greifer, 2023) package from R to compute propensity scores and assign control firms to treated firms. The resulting matched sample consists of 149 AI-exposed firms and 455 control firms.

Although propensity score matching is a widely used strategy to reduce selection bias, it has several weaknesses. First, it reduces the statistical power of subsequent analyses by omitting a large portion of the sample. We address this by performing 1:3 matching to preserve a decent sample size. Second, the technique relies on the assumption that treatment is exogenous conditional on covariates that affect both treatment and the outcome. Ideally, propensity scores would be calculated based on all confounders, but we can only utilize observable characteristics. Propensity score matching essentially attempts to solve the problem of unobservable confounders by matching on observables. Hence, it cannot completely eliminate bias from unobservable variables. However, with a matched control group that is similar across important covariates, we argue that it is more likely that firms will respond similarly to changes in unobserved variables. Summary statistics for treated and control groups can be found in Appendix B.2, which displays a greater balance in covariates with the matched control group compared to the bottom 50% control group.

Figure 5.3 illustrates trend plots for the matched sample. The EPS plot displays some improvement compared to the unmatched sample, especially for the months leading up to the release. Additionally, the trends for sales forecasts are no longer diverging in the first months of 2022, and recommendations exhibit a greater degree of parallelism compared to the unmatched plot. However, there are still signs of the parallel trends assumption being violated, especially in the sales forecast plot. Plots for all dependent variables can be found in Appendix B.1.2, which gives similar conclusions, especially for longer horizons of EPS and sales forecasts. The violation of parallelism can introduce bias to model estimates that need to be taken into account when interpreting findings. However, parallel trends seem to be fulfilled to a greater extent compared to the unmatched sample. Thus, models with the matched sample should contain the least biased estimates.

Figure 5.3: Parallel trends visualization: matched sample

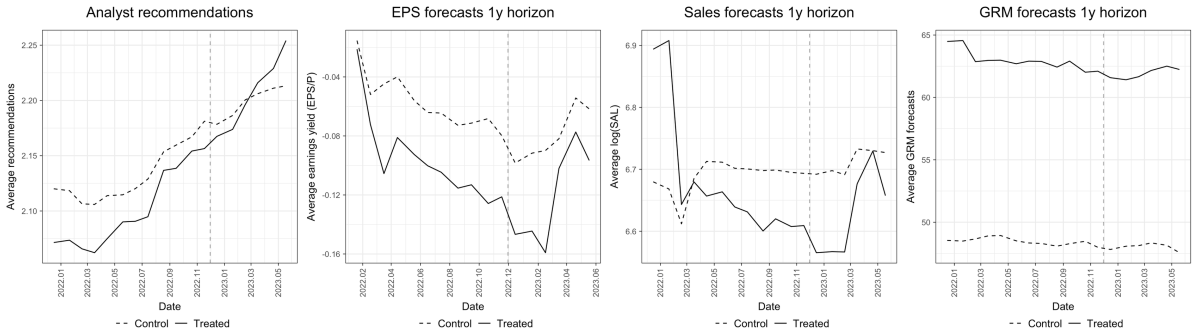

Figure 5.3 illustrates the mean of selected dependent variables for the matched sample, grouped by treatment status. The plots include [t-12, t+6] months of data, where t marks the date of the ChatGPT release (November 30^{th}, 2022). Except for recommendations, all dependent variables have been winsorized at the 95% level.

5.2.3 Recommendations Analysis on Matched Sample

Table 5.3 contains results from the recommendations analysis on both the matched and full sample. The results are similar, although there are some negligible differences in the magnitude and significance of coefficients. Therefore, the interpretation and economic significance of the findings are the same as for the full sample. In the full sample analysis, we argued that the development in analyst recommendations could be explained by share prices increasing beyond what analysts consider fundamental values.

Table 5.3: Analyst recommendations regression output: matched vs. full sample

	Matched sample		Full sample	
	Rec.		Rec.	
	(1)	(2)	(1)	(2)
Treat x After	**0.035***		**0.028.**	
	(0.017)		(0.014)	
Treat x Dec. 22		0.016		0.006
		(0.012)		(0.010)
Treat x Jan. 23		0.014		0.004
		(0.016)		(0.013)
Treat x Feb. 23		0.024		0.015
		(0.018)		(0.015)
Treat x Mar. 23		**0.038.**		**0.031.**
		(0.021)		(0.018)
Treat x Apr. 23		**0.050***		**0.045***
		(0.023)		(0.020)
Treat x May 23		**0.069****		**0.066****
		(0.026)		(0.023)
Year-month FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Controls			Yes	Yes
Num.Obs.	6987	6987	19 009	19 009
R2 Adj.	0.896	0.896	0.895	0.895

Clustered (firm) standard errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, .:0.1*

Table 5.3 displays regression output for the matched and full sample with recommendations as the dependent variable. In the full sample, the bottom 50% of AI-exposed firms constitutes the control group. Models (1) display output from the standard DiD, and models (2) from the DiD where the post-treatment period is segmented into months. In the full sample models, we control for *debt to equity*, *intangible to total assets*, and log(*total assets*). Control variables are winsorized at the 95% level. The sample consists of [t-6, t+6] months of forecast data, where t marks the date of the ChatGPT release (November 30$^{\text{th}}$, 2022).

5.3 Components of Firm Value

Recommendations moving towards "sell" do not necessarily imply a negative earnings reaction. If we assume that prices explain the recommendation results, the reaction toward earnings could still have been positive if the magnitude of price increases was more than proportional to the increase in earnings. In this section, we analyze components of firm value. First, we study earnings forecasts. Then, we decompose earnings expectations into projected revenues and costs by studying sales and gross margin forecasts.

Results for subsequent analyses will be displayed for the matched sample, with full sample tables available in Appendix B.3. Although the results have consistent interpretations, we comment on any notable differences.

5.3.1 EPS Forecasts

Table 5.4 illustrates the findings from analyzing EPS forecasts, scaled by share price in models 1y – 5y and thus interpreted as the forecasted percentage earnings yield. The results show a significant increase in forecasted earnings yield for AI-exposed firms on the three-year horizon, where the average increase ranges from 1.4% to 2.0%, depending on the month. Considering the earnings yield for the S&P 500 was 6.5% in October 2022 (Gilmartin, 2022), a month before release, this is a significant increase in magnitude. The upward adjustment occurred as early as December, evidenced by the coefficient on $Treated \times Dec.$ 22, before leveling out for the rest of the period. This indicates a quick and persistent positive reaction towards earnings for AI-exposed firms. Although coefficients are not statistically significant, there are signs of a similar pattern for other yearly horizons, particularly for two and four years ahead. For most yearly horizons, in almost all months after the release, coefficients are positive. This further indicates a positive earnings reaction. However, one cannot draw conclusions based on non-significant coefficients, as the probability of observing these values is (too) high if the true effect is in fact null. Therefore, the only finding that is convincing from a statistical perspective is that of an upward adjustment in earnings forecasts on the three-year horizon.

The results for long-term growth forecasts are inconclusive. Whereas they show signs of increasing right after the release, they decrease over time relative to the control group. Again, the results are not significant, so there is little evidence of any effect.

In summary, the only conclusive finding from analyzing EPS forecasts is a positive reaction on the three-year horizon. Little evidence supports the hypothesis about stronger earnings effects on longer horizons. Although the increase in magnitude for the three-year horizon is quite large, the earnings reaction towards AI-exposed firms is not as strong as expected when considering all horizons combined.

Table 5.4: EPS forecasts regression output: matched sample

	EPS forecast horizon					
	1y	2y	3y	4y	5y	LTG
Treat x Dec. 22	−0.004	0.011	**0.019***	0.016	−0.007	0.704
	(0.015)	(0.015)	(0.010)	(0.019)	(0.032)	(1.853)
Treat x Jan. 23	−0.003	0.008	**0.020***	0.016	−0.001	−0.206
	(0.015)	(0.015)	(0.010)	(0.020)	(0.032)	(1.290)
Treat x Feb. 23	0.005	0.009	**0.016***	0.023	0.013	−0.685
	(0.015)	(0.014)	(0.008)	(0.022)	(0.037)	(1.431)
Treat x Mar. 23	0.023	0.012	**0.016.**	0.026	0.016	−1.741
	(0.018)	(0.015)	(0.009)	(0.023)	(0.026)	(1.584)
Treat x Apr. 23	−0.001	0.001	**0.014.**	0.012	0.009	−1.451
	(0.020)	(0.013)	(0.008)	(0.025)	(0.031)	(1.585)
Treat x May 23	0.013	0.005	**0.018***	0.025	0.004	−0.531
	(0.018)	(0.012)	(0.008)	(0.022)	(0.030)	(1.774)
Year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Num.Obs.	7045	7017	5840	3296	2394	2296
R2 Adj.	0.910	0.881	0.872	0.815	0.721	0.809

Clustered (firm) standard errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, .:0.1*

Table 5.4 displays regression output from a matched sample DiD model where the post-treatment period is segmented into months, displaying the average treatment effect each month after the release of ChatGPT. The dependent variable for models 1y - 5y is EPS forecasts normalized by share price and has the interpretation of the forecasted percentage earnings yield. The dependent variable for the LTG model is long-term growth forecasts as reported in IBES, which is the expected average growth rate over the coming business cycle. Thus, being AI-exposed in a specific month after release is associated with a $(100 \times \beta_i)$ % change in earnings yield for 1y - 5y models, and a β_i% change in LTG for the LTG model. Dependent variables are winsorized at the 95% level. The sample consists of [t-6, t+6] months of forecast data, where t marks the date of the ChatGPT release (November 30th, 2022).

5.3.2 Decomposing Earnings Into Revenues and Costs

Next, we present findings from analyses of sales and gross margin forecasts. By interpreting the combined findings, we decompose earnings expectations into projected revenues and costs to understand what is driving the adjustment in earnings forecasts for the three-year horizon.

Sales Forecasts

Table 5.5 contains findings from analyzing sales forecasts. For yearly horizons, coefficients are mostly non-significant except for weak evidence of an upward adjustment on the one-year horizon. The increase in one-year sales forecasts range from 6.7% - 8.0% depending on the month. The upward adjustment did not occur before March 2023, perhaps indicating that analysts took time to realize the short-term revenue-enhancing potential for AI-exposed firms. Other yearly horizons show little evidence of moving in either direction. However, coefficients are mostly negative on three-year and longer horizons, giving some indication of a negative sales reaction. This is supported by findings from the full sample, displayed in Appendix B.3.2, that show significant downward adjustments for three-year, four-year, and five-year horizons immediately after the ChatGPT release.

The negative sales reaction is most notable for LTG forecasts, where coefficients are significantly negative for all months following the ChatGPT release. Downward adjustments range from 1.5% - 4.4%, and the magnitude increases for each month following the release. This suggests that analysts considered the ChatGPT release to be a negative sales signal for AI-exposed firms in the long term and that this view became stronger over time. Thus, the findings for longer horizons contradict our hypothesis of positive long-term sales reactions from analysts. It should be noted that there are fewer observations for 4y, 5y, and LTG horizons, as observed in the summary statistics. This could impact statistical power and model validity due to different compositions of control and treated groups. However, the effects on LTG forecasts are highly significant and provide convincing evidence of a negative sales reaction.

To summarize, the results for sales forecasts are not uniform across horizons. They display a positive reaction for the one-year horizon and negative reactions for longer horizons. The negative reaction is especially prevalent in LTG forecasts. The magnitude of the upward adjustment on the one-year horizon (6.7% - 8.0%) is greater than the downward adjustments in LTG forecasts (1.5% - 4.4%). However, the LTG evidence is more convincing from a statistical perspective. Therefore, the findings partly contradict our hypothesis about the revenue-enhancing effects of the ChatGPT release for AI-exposed firms.

Table 5.5: Sales forecasts regression output: matched sample

	Sales forecast horizon					
	1y	2y	3y	4y	5y	LTG
Treat x Dec. 22	−0.005	−0.002	−0.015	−0.024	−0.033	**-1.540****
	(0.013)	(0.013)	(0.011)	(0.019)	(0.023)	(0.519)
Treat x Jan. 23	−0.005	−0.005	−0.017	−0.035	−0.032	**-2.121*****
	(0.014)	(0.014)	(0.013)	(0.023)	(0.026)	(0.628)
Treat x Feb. 23	0.013	0.002	−0.022	−0.030	−0.027	**-2.878*****
	(0.021)	(0.016)	(0.017)	(0.025)	(0.030)	(0.835)
Treat x Mar. 23	**0.077.**	0.011	−0.028	−0.010	−0.006	**-3.482*****
	(0.042)	(0.019)	(0.022)	(0.025)	(0.034)	(1.036)
Treat x Apr. 23	**0.080.**	0.015	−0.017	−0.020	0.033	**-3.721****
	(0.042)	(0.022)	(0.019)	(0.028)	(0.043)	(1.120)
Treat x May 23	**0.067.**	0.008	−0.058	−0.028	0.029	**-4.355*****
	(0.039)	(0.024)	(0.044)	(0.029)	(0.043)	(1.244)
Year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Num.Obs.	7034	6999	5878	3692	2913	1807
R2 Adj.	0.990	0.993	0.994	0.993	0.988	0.904

Clustered (firm) standard errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, .:0.1*

Table 5.5 displays regression output from a matched sample DiD model where the post-treatment period is segmented into months, displaying the average treatment effect each month after the release of ChatGPT. The dependent variable for models 1y-5y is the logarithm of sales forecasts for each horizon. The dependent variable for the LTG model is long-term growth forecasts as reported in IBES, which is the expected average growth rate over the coming business cycle. Thus, being AI-exposed in a specific month after release is associated with a $(100 \times \beta_i)\,\%$ change in forecasts for 1y - 5y models, and a $\beta_i\%$ change in LTG for the LTG model. Dependent variables are winsorized at the 95% level. The sample consists of [t-6, t+6] months of forecast data, where t marks the date of the ChatGPT release (November 30$^{\text{th}}$, 2022).

Gross Margin Forecasts

Table 5.6 contains findings from analyzing gross margin forecasts. The signs of coefficients give a weak indication of a more positive reaction on shorter compared to longer horizons, similar to the development in sales forecasts. However, coefficients are not statistically significant for any horizons. Hence, we cannot conclude that gross margins were adjusted in either direction. Therefore, the findings do not support our hypothesis expecting a decrease in cost forecasts, which would entail increased gross margin forecasts.

Table 5.6: Gross margin forecasts regression output: matched sample

	GRM forecast horizon				
	1y	2y	3y	4y	5y
Treat x Dec. 22	−0.052	0.205	0.178	−0.576	−0.729
	(0.383)	(0.288)	(0.378)	(0.489)	(0.530)
Treat x Jan. 23	−0.004	0.159	0.066	−0.633	−0.903
	(0.420)	(0.313)	(0.389)	(0.503)	(0.577)
Treat x Feb. 23	0.096	0.508	−0.134	−0.555	−0.087
	(0.455)	(0.359)	(0.500)	(0.547)	(0.793)
Treat x Mar. 23	0.156	0.194	0.237	0.414	0.383
	(0.596)	(0.397)	(0.592)	(1.226)	(1.075)
Treat x Apr. 23	−0.007	0.283	−0.227	−0.111	−0.971
	(0.702)	(0.416)	(0.611)	(1.080)	(0.973)
Treat x May 23	0.174	0.239	−0.062	−0.109	−1.056
	(0.738)	(0.515)	(0.570)	(1.129)	(0.960)
Year-month FE	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Num.Obs.	5647	5601	4346	2341	1580
R2 Adj.	0.972	0.981	0.977	0.957	0.982

Clustered (firm) standard errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, .:0.1*

Table 5.6 displays regression output from a matched sample DiD model where the post-treatment period is segmented into months, displaying the average treatment effect each month after the release of ChatGPT. The dependent variable is gross margin forecasts, winsorized at the 95% level. The coefficients can be interpreted as the β_i% change in forecasts from being in the treated group in a specific month after release. The sample consists of [t-6, t+6] months of forecast data, where t marks the date of the ChatGPT release (November 30$^{\text{th}}$, 2022).

What Drives Earnings Forecast Revisions?

Analyzing sales and gross margin forecasts does not clearly indicate what drives the observed increase in earnings forecasts for the three-year horizon. We would expect an increase in sales forecasts, gross margin forecasts, or both. However, the results show neither of these effects. The only convincing result from a statistical perspective is that LTG sales forecasts are adjusted downward.

However, by cautiously interpreting sales LTG forecasts, we can perhaps derive insights into what drives the earnings forecast increase. LTG forecasts indicate the projected growth over the coming full business cycle, usually referring to a period between three to five years (Derrien et al., 2021). The results in the 3y – 5y models show a similar trend as

LTG forecasts of being downward adjusted. Although the coefficients are not statistically significant, if LTG forecasts incorporate sales projections for the three-year horizon, it further indicates a downward adjustment. As mentioned, the full sample findings for sales forecasts also indicate a significant downward adjustment on the three-year horizon. Under the assumption that sales forecasts decreased, the explanation for the increase in earnings forecasts must be a decrease in projected cost levels.

Following the argument above, we would expect to find an increase in gross margin forecasts. Since this effect is not observed, it could be that decreases in other costs drive the increased earnings expectations. The percentage gross margin only incorporates COGS, which are costs directly related to the production of goods and services. Cost items commonly excluded from COGS include depreciation and amortization as well as selling, general and administrative costs (Maverick, 2021). Given that we do not observe any effects on gross margins, a potential explanation could be a decrease in such cost items. For example, it was argued in the hypothesis section that AI integration can decrease the marginal search costs in R&D, leading to a decrease in R&D amortization costs.

In summary, the most likely explanation for the observed increase in earnings forecasts for the three-year horizon is a decrease in costs. However, it is important to note that the above discussions are speculative. Non-significant results can mean that the data supports the null hypothesis and there is no real effect or that there is insufficient evidence to support the hypothesis tested. Either way, the results for sales and gross margin models on the three-year horizon are not strong enough to make firm conclusions.

5.4 Robustness of Findings

To test the robustness of our findings, we perform an additional validity check of the parallel trends assumption. Parallel trends are crucial to inference in DiD models since the design implicitly assumes that the trend of the control group is the counterfactual for the treated group. Since parallel trends appear only partly fulfilled from visual inspection, we perform a formal test to validate the assumption. The robustness test is performed for models with statistically significant findings, specifically the REC, EPS 3y, SAL 1y, and SAL LTG models. The purpose is to understand whether these findings are likely to be driven by differences in AI-exposure or other unobserved factors biasing the estimates.

To perform the test, we specify the models as if the ChatGPT release was on July 31$^{\text{st}}$, 2022 rather than November 30$^{\text{th}}$, 2022. Consistent with previous models, 6 months of forecast data prior to July 31$^{\text{st}}$, 2022 is used as the benchmark. We interact treatment with each month leading up to the actual release. If the treated and control groups behave similarly in months prior to the event, we expect coefficients on $Treat_i \times Month_t$ interactions to be small in magnitude and statistically non-significant. Table 5.7 reports the output from these regressions. Reassuringly, none of the interaction terms with months prior to the ChatGPT release are statistically significant. Further, for all months prior to the actual release, the coefficients on interactions are smaller in magnitude than post-release coefficients. For the SAL 1y model, this does not fully apply. However, the differences in magnitude are not large, and pre-release coefficients are smaller than the coefficients that are statistically significant post-release. In addition, post-ChatGPT results from Table 5.7 are similar in magnitude and significance as when using the actual event date with a different benchmark. This indicates that the findings are robust to using a benchmark consisting of a different sample of forecast data. Overall, the results from the robustness test increase confidence in the validity of the estimates.

Table 5.7: Formal test of the parallel trends assumption for statistically significant findings

	REC	EPS 3y	SAL 1y	SAL LTG
Treat x Aug. 22	0.020	0.007	−0.030	0.308
	(0.019)	(0.008)	(0.027)	(0.947)
Treat x Sep. 22	0.017	0.007	−0.022	−0.133
	(0.021)	(0.009)	(0.028)	(0.978)
Treat x Oct. 22	0.025	−0.005	−0.024	−0.285
	(0.022)	(0.009)	(0.028)	(1.032)
Treat x Nov. 22	0.018	0.011	−0.016	−0.965
	(0.023)	(0.010)	(0.029)	(1.047)
Treat x Dec. 22	0.032	**0.021.**	−0.017	−1.419
	(0.024)	(0.011)	(0.029)	(1.063)
Treat x Jan. 23	0.030	**0.019.**	−0.017	**-1.955.**
	(0.026)	(0.011)	(0.029)	(1.092)
Treat x Feb. 23	0.040	0.014	0.003	**-2.843***
	(0.027)	(0.010)	(0.033)	(1.133)
Treat x Mar. 23	**0.054.**	0.017	**0.068.**	**-3.461****
	(0.029)	(0.011)	(0.039)	(1.226)
Treat x Apr. 23	**0.065***	0.015	**0.071.**	**-3.696****
	(0.030)	(0.010)	(0.041)	(1.318)
Treat x May 23	**0.085***	**0.019***	0.058	**-4.324****
	(0.034)	(0.010)	(0.037)	(1.399)
Year-month FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Num.Obs.	9294	7750	9335	2476
R2 Adj.	0.849	0.832	0.988	0.885

Clustered (firm) standard errors in parathesis
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, .:0.1*

Table 5.7 displays regression output from a formal test of the parallel trends assumption. Treatment is interacted with four months prior to the ChatGPT release (November 30th, 2022), marked by the line in the output. If coefficients on these interactions are small in magnitude and statistical significance, it indicates that the parallel trends assumption is fulfilled. The matched sample is used, and the sample consists of [t-10, t+6] months of forecast data, where t marks the date of the ChatGPT release.

5.5 Discussion

In this section of the analysis, we summarize our main findings and discuss how they relate to the initial hypotheses. From studying analyst recommendations, we find that the ChatGPT release led to a downgrade in recommendations for AI-exposed firms and that the effect became stronger over time. However, the magnitude of the effect is not large, at only 2.8% of a full-step recommendation downgrade. This finding does not support our hypothesis, which states that increased earnings expectations would lead to recommendations being adjusted towards "buy" following the release. From analyzing EPS forecasts, we find evidence of increased earnings expectations, specifically on the three-year horizon. Holding price constant, the increase in earnings expectations should constitute recommendations adjusted towards "buy". Therefore, the effect is likely to be driven by stock prices for AI-exposed firms increasing beyond what analysts consider fundamental value. This is supported by plotting market capitalization-weighted indices of the treated and control group, which indicates that stock prices for AI-exposed firms increased relative to the control group following the release.

The other main hypothesis of this paper was that the release of ChatGPT would lead to increased earnings forecasts for AI-exposed firms, with stronger effects on longer horizons. The earnings effect was expected through changes in forecasts of revenues and costs, where sales were expected to increase and costs to decrease. Testing these hypotheses yields three main findings. First, earnings forecasts were adjusted upwards on the 3y horizon which is consistent with the hypothesis. Although the magnitude of the increase is quite large (1.4% - 2.0% increase in forecasted earnings yield), the earnings reaction is not as strong as expected when considering all horizons combined. We do not find the expected effects on long-term horizons. Second, sales forecast reactions are not uniform across horizons, where one-year forecasts are adjusted upwards, and LTG forecasts are adjusted downwards. The results for the one-year horizon are weak. Thus, the only convincing effect is a downward adjustment in LTG forecasts, which is not in line with the hypothesis. Third, from jointly interpreting all findings, we find that the most likely explanation for the increase in earnings forecasts is a decrease in projected cost levels. In the following, we discuss potential explanations for why results are not as strong, or even opposite, of what was expected.

While we argue that AI-exposed firms have the resources to benefit from the technology shock represented by the ChatGPT release, the competitive advantage might not be as great as expected. Since ChatGPT is a freely accessible tool, it means that non-AI firms can move from zero to basic AI utilization. Thus, the ChatGPT release could be a signal of the future benefits of AI use for all companies. It remains a matter of discussion what analysts perceived as having the greater effect: basic AI use for the control group or being able to utilize new technology more proficiently for the treated group. The results give a weak indication of the latter, but not for longer horizons. Moreover, ChatGPT represents progress toward artificial general intelligence (Morris et al., 2023). This entails that future AI models could have cognitive abilities and problem-solving skills equal to or surpassing that of humans. When (or if) AI models reach that level of intelligence, the effects will likely encompass the entire society and not only AI-exposed firms. If analysts incorporated this in their expectations, it could explain why long-term results are not as strong as expected and why sales LTG forecasts are downward adjusted. However, ChatGPT has been identified as an *emerging* AGI, and the advent of true AGI models is likely far into the future (Bubeck et al., 2023).

Additionally, the hypothesized impacts on earnings might be perceived by analysts as too uncertain to be incorporated into forecasts and recommendations. For instance, while the release could signal improved prospects for innovation, it lacks detailed information about what these innovations entail, their timelines, and their specific impact on firm values. Thus, the uncertain nature of some effects might lead analysts to not consider them in their forecasts.

6 Conclusion

Several papers have analyzed the impact of the ChatGPT release, but to our knowledge, our paper offers the first insights into how analysts responded to this event. By taking the premise that the ChatGPT release represented a major technology shock to AI, we attempt to understand how the event impacted expectations toward firms positioned to leverage such technology advancements. By studying analyst expectations, our paper contributes to understanding the anticipated effects of AI technology in the future and how these effects are expected to unfold over time.

We quantify firm-level AI-exposure from a combination of NLP techniques applied to earnings call transcripts. The exposure measure successfully identifies firms that are widely known for utilizing AI and a wide spectrum of AI integration across industries. Using a difference-in-differences framework, we first estimate the effect of the release on analyst recommendations for AI-exposed firms. Next, we study EPS forecasts before decomposing the earnings effect through a combined study of sales and gross margin forecasts.

Our findings suggest an increase in earnings forecasts on the three-year horizon. From jointly interpreting findings, we argue that reduced cost forecasts likely drive the increase in earnings expectations. Long-term growth forecasts for sales are adjusted downward, suggesting that the competitive advantage of AI-exposed firms might be expected to dwindle over time. Holding price constant, an increase in earnings forecasts should constitute recommendations adjusted towards "buy". However, we find evidence of a downgrade in recommendations for AI-exposed firms. This suggests that the increase in forecasted earnings was not sufficiently large to upgrade recommendations, justify increases in share prices, or both. Since we observe the opposite effect on recommendations than what earnings expectations suggest, we attribute the downgrade to share prices moving beyond what analysts consider fundamental values.

We hypothesized that the release of ChatGPT signaled improved earnings potential for AI-exposed firms through an accelerated pace of innovation, increased productivity, and enhanced potential in data analysis. Although the impact is not as strong as expected, we still observe a positive impact on earnings. Thus, our findings suggest that the predicted effects are expected to materialize to some extent.

References

Acemoglu, D., & Restrepo, P. (2020). Robots and jobs: Evidence from us labor markets. *Journal of political economy*, *128*(6), 2188–2244.

Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in medicine*, *33*(6), 1057–1069.

Beltagy, I., Lo, K., & Cohan, A. (2019). Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*.

Benner, M. J. (2010). Securities analysts and incumbent response to radical technological change: Evidence from digital photography and internet telephony. *Organization Science*, *21*(1), 42–62.

Brynjolfsson, E., Rock, D., & Syverson, C. (2018). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. In *The economics of artificial intelligence: An agenda* (pp. 23–57). University of Chicago Press.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., et al. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

Bughin, J., Hazan, E., Sree Ramaswamy, P., DC, W., Chu, M., et al. (2017). Artificial intelligence the next digital frontier.

Chui, M., Hazan, E., Roberts, R., Singla, A., & Smaje, K. (2023). The economic potential of generative ai.

Clement, M. B., & Tse, S. Y. (2005). Financial analyst characteristics and herding behavior in forecasting. *The Journal of finance*, *60*(1), 307–341.

Cockburn, I. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation: An exploratory analysis. In *The economics of artificial intelligence: An agenda* (pp. 115–146). University of Chicago Press.

Conrad, J., Cornell, B., Landsman, W. R., & Rountree, B. R. (2006). How do analyst recommendations respond to major news? *Journal of Financial and Quantitative Analysis*, *41*(1), 25–49.

Dai, R. (2020). *I/b/e/s on wrds — introduction and research guide*. Wharton University of Pennsylvania. https://wrds-www.wharton.upenn.edu/pages/grid-items/ibes-wrds-101-introduction-and-research-guide/?fbclid=IwAR0QNZXeUrIa4gLcAgJ4k5b3YMtxkf2768mP4rQELg6hbycaUCPA0_8a1nQ

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (24-013).

Derrien, F., Krueger, P., Landier, A., & Yao, T. (2021). Esg news, future cash flows, and firm value. *Swiss finance institute research paper*, (21-84).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Eisfeldt, A. L., Schubert, G., & Zhang, M. B. (2023). *Generative ai and firm values* (tech. rep.). National Bureau of Economic Research.

Furman, J., & Seamans, R. (2019). Ai and the economy. *Innovation policy and the economy*, *19*(1), 161–191.

Garcia, D., Hu, X., & Rohrer, M. (2023). The colour of finance words. *Journal of Financial Economics*, *147*(3), 525–549.

Gilmartin, B. (2022, October 17). *Sp 500 earnings: Sp 500 earnings yield hits high not seen since january 2019*. Seeking Alpha. https://seekingalpha.com/article/4546879-sp-500-earnings-yield-hits-high-not-seen-since-january-2019

Goertzel, B. (2014). Artificial general intelligence: Concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, *5*(1), 1–46.

Google Trends. (2023a, November 16). *Artificial intelligence*. Google. Retrieved November 16, 2023, from https://trends.google.com/trends/explore?date=today%205-y&q=Artificial%20Intelligence&hl=no

Google Trends. (2023b, November 16). *Generative ai*. Google. Retrieved November 16, 2023, from https://trends.google.com/trends/explore?date=today%205-y&q=Generative%20AI&hl=no

Greifer, N. (2023, October 12). *Matchit: Getting started*. https://cran.r-project.org/web/packages/MatchIt/vignettes/MatchIt.html

Han, J., Kamber, M., & Pei, J. (2012). Data mining concepts and techniques third edition. *University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University*.

Hang, H., & Chen, Z. (2022). How to realize the full potentials of artificial intelligence (ai) in digital economy? a literature review. *Journal of Digital Economy*.

Hobijn, B., & Jovanovic, B. (2001). The information-technology revolution and the stock market: Evidence. *American Economic Review*, *91*(5), 1203–1220.

Hu, K. (2023). Chatgpt sets record for fastest-growing user base - analyst note. *Reuters*. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

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

Krakowski, S., Luger, J., & Raisch, S. (2023). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal*, *44*(6), 1425–1452.

Kreps, S. (2023). The openai meltdown will only accelerate the artificial intelligence race. *The Guardian*. https://www.theguardian.com/commentisfree/2023/nov/22/openai-chatgpt-artificial-general-intelligence-race-acceleration

Maverick, J. B. (2021, May 23). *What costs are not counted in gross profit margin?* Investopedia. https://www.investopedia.com/ask/answers/010915/what-costs-are-not-counted-gross-profit-margin.asp

Meta. (n.d.). *Introducing llama 2*. Meta Platforms. https://ai.meta.com/llama/

Microsoft. (n.d.). *Stockhistory function*. Microsoft. https://support.microsoft.com/en-us/office/stockhistory-function-1ac8b5b3-5f62-4d94-8ab8-7504ec7239a8

Morris, M. R., Sohl-dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., & Legg, S. (2023). Levels of agi: Operationalizing progress on the path to agi. *arXiv preprint arXiv:2311.02462*.

Nguyen, J., Lee, J., & Hulseman, I. (2023, June 12). *New video: What runs chatgpt?* Nvidia. https://developer.nvidia.com/blog/new-video-what-runs-chatgpt/

OpenAI. (2023, November 30). *Introducing chatgpt*. OpenAI. https://openai.com/blog/chatgpt

Pichai, S. (2023, February 6). *An important next step on our ai journey*. Google. https://blog.google/technology/ai/bard-google-ai-search-updates/

Press, G. (2019). 114 milestones in the history of artificial intelligence (ai). *Forbes*. https://www.forbes.com/sites/gilpress/2021/05/19/114-milestones-in-the-history-of-artificial-intelligence-ai/?sh=71e5c32c74bf

Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, *36*(4), 992–1011.

RapidAPI. (2023). *What is rapidapi.com?* Rapid API. https://docs.rapidapi.com/docs/what-is-rapidapi

Rassen, J. A., Shelat, A. A., Myers, J., Glynn, R. J., Rothman, K. J., & Schneeweiss, S. (2012). One-to-many propensity score matching in cohort studies. *Pharmacoepidemiology and drug safety*, *21*, 69–80.

SeekingAlpha. (n.d.). *We are seeking alpha*. Seeking Alpha. https://about.seekingalpha.com/

SeekingAlpha. (2022, December 20). *Vivos therapeutics, inc. (vvos) q2 and q3 2022 earnings call transcript*. SeekingAlpha. https://seekingalpha.com/article/4565643-vivos-therapeutics-inc-vvos-q2-and-q3-2022-earnings-call-transcript

SeekingAlpha. (2023, October 24). *Alphabet inc. (goog) q3 2023 earnings call transcript*. Seeking Alpha. https://seekingalpha.com/article/4643107-alphabet-inc-goog-q3-2023-earnings-call-transcript

Strack, R., Carrasco, M., Kolo, P., Nouri, N., Priddis, M., & George, R. (2021). The future of jobs in the era of ai. *Boston Consulting Group*.

Trueman, B. (1994). Analyst forecasts and herding behavior. *The review of financial studies*, *7*(1), 97–124.

Viviani, M. (2023). *How chatgpt and gpt-4 can be used for 3d content generation*. Nvidia. https://resources.nvidia.com/en-us-omniverse-genai/ov-chatgpt

Wharton. (n.d.). *Wharton research data services*. Wharton University of Pennsylvania. Retrieved December 16, 2023, from https://wrds-www.wharton.upenn.edu/

Wixom, B. H., Someh, I. A., Zutavern, A., & Beath, C. M. (2020). Explanation: A new enterprise data monetization capability for ai. *MIT Center for Information Systems Research (CISR), Working Paper*, (443).

Yang, C.-H. (2022). How artificial intelligence technology affects productivity and employment: Firm-level evidence from taiwan. *Research Policy*, *51*(6), 104536.

Appendices

A References for the AI Corpus

Ameen, N., Tarhini, A., Reppel, A., & Anand, A. (2021). Customer experiences in the age of artificial intelligence. *Computers in Human Behavior*, *114*, 106548.

Bengio, Y., Lecun, Y., & Hinton, G. (2021). Deep learning for ai. *Communications of the ACM*, *64*(7), 58–65.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS quarterly*, *45*(3).

Bharadiya, J. P. (2023). A comparative study of business intelligence and artificial intelligence with big data analytics. *American Journal of Artificial Intelligence*, *7*(1), 24.

Borges, A. F., Laurindo, F. J., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*, *57*, 102225.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative ai at work* (tech. rep.). National Bureau of Economic Research.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *Ieee Access*, *8*, 75264–75278.

Chui, M., Hazan, E., Roberts, R., Singla, A., & Smaje, K. (2023). The economic potential of generative ai.

Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A., & De Felice, F. (2020). Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. *Sustainability*, *12*(2), 492.

Cui, M., & Zhang, D. Y. (2021). Artificial intelligence and computational pathology. *Laboratory Investigation*, *101*(4), 412–422.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*. Retrieved October 25, 2023, from https://hbr.org/2018/01/artificial-intelligence-for-the-real-world

DigitalOcean. (n.d.). *How to leverage artificial intelligence to grow your business*. Digital Ocean. https://www.digitalocean.com/resources/article/artificial-intelligence-in-business

EIU. (2023, June 5). *How companies use artificial intelligence*. Economist Intelligence. https://www.eiu.com/n/how-companies-use-artificial-intelligence/

Finlayson, S. G., Subbaswamy, A., Singh, K., Bowers, J., Kupke, A., Zittrain, J., Kohane, I. S., & Saria, S. (2021). The clinician and dataset shift in artificial intelligence. *New England Journal of Medicine*, *385*(3), 283–286.

Frankenfield, J. (2023, September 4). *Artificial intelligence (ai): What it is and how it is used*. Investopedia. https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp

Furman, J., & Seamans, R. (2019). Ai and the economy. *Innovation policy and the economy*, *19*(1), 161–191.

Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., & Spaulding, R. (2021). Attachment and trust in artificial intelligence. *Computers in Human Behavior*, *115*, 106607.

Goertzel, B. (2014). Artificial general intelligence: Concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, *5*(1), 1.

Google. (n.d.). *What is artificial intelligence (ai)?* Google. https://cloud.google.com/learn/what-is-artificial-intelligence

Goralski, M. A., & Tan, T. K. (2020). Artificial intelligence and sustainable development. *The International Journal of Management Education*, *18*(1), 100330.

Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., & Kumar, P. (2021). Artificial intelligence to deep learning: Machine intelligence approach for drug discovery. *Molecular diversity*, *25*, 1315–1360.

Haan, K. (2023). How businesses are using artificial intelligence in 2023. *Forbes*. https://www.forbes.com/advisor/business/software/ai-in-business/

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). Machine learning and artificial intelligence: Definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, *13*, 69–76.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(4), e1312.

Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, *49*, 30–50.

Huynh-The, T., Pham, Q.-V., Pham, X.-Q., Nguyen, T. T., Han, Z., & Kim, D.-S. (2023). Artificial intelligence for the metaverse: A survey. *Engineering Applications of Artificial Intelligence*, *117*, 105581.

IBM. (n.d.). *What is artificial intelligence (ai)?* IBM. https://www.ibm.com/topics/artificial-intelligence

Jha, K., Doshi, A., Patel, P., & Shah, M. (2019). A comprehensive review on automation in agriculture using artificial intelligence. *Artificial Intelligence in Agriculture*, *2*, 1–12.

Jordan, M. I. (2019). Artificial intelligence—the revolution hasn't happened yet. *Harvard Data Science Review*, *1*(1), 1–9.

Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! the challenges and opportunities of artificial intelligence. *Business Horizons*, *63*(1), 37–50.

Kleinings, H. (2023, January 18). *How to get the most out of ai in 2023: 7 applications of artificial intelligence in business*. Levity. https://levity.ai/blog/8-uses-ai-business

Loureiro, S. M. C., Guerreiro, J., & Tussyadiah, I. (2021). Artificial intelligence in business: State of the art and future research agenda. *Journal of business research*, *129*, 911–926.

McKinsey Company. (2023, April 24). *What is ai?* McKinsey Company. https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai

Narang, M. (2023, September 5). *Artificial intelligence (ai) in business: A complete guide*. Knowledge Hut. https://www.knowledgehut.com/blog/business-intelligence-and-visualization/artificial-intelligence-in-business

NI Business Info. (n.d.). *Artificial intelligence in business*. NI Business Info. https://www.nibusinessinfo.co.uk/content/examples-artificial-intelligence-use-business

Rajagopalan, R. (n.d.). *10 examples of artificial intelligence in business*. University of San Diego. https://onlinedegrees.sandiego.edu/artificial-intelligence-business/

SAS. (n.d.). *Artificial intelligence what it is and why it matters*. SAS. https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html

Schwendicke, F. a., Samek, W., & Krois, J. (2020). Artificial intelligence in dentistry: Chances and challenges. *Journal of dental research*, *99*(7), 769–774.

Shapiro, S. C. (1992). Encyclopedia of artificial intelligence second edition. *New Jersey: A Wiley Interscience Publication*.

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., et al. (2022). Artificial intelligence and life in 2030: The one hundred year study on artificial intelligence. *arXiv preprint arXiv:2211.06318*.

Uzialko, A. (2023, October 23). *How artificial intelligence will transform businesses*. Business News Daily. https://www.businessnewsdaily.com/9402-artificial-intelligence-business-trends.html

Verganti, R., Vendraminelli, L., & Iansiti, M. (2020). Innovation and design in the age of artificial intelligence. *Journal of Product Innovation Management*, *37*(3), 212–227.

Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37.

Watters, A. (2021, June 4). *Using ai in business: Examples of artificial intelligence application in business*. CompTIA. https://connect.comptia.org/blog/using-ai-in-business

Wetzstein, G., Ozcan, A., Gigan, S., Fan, S., Englund, D., Soljačić, M., Denz, C., Miller, D. A., & Psaltis, D. (2020). Inference in artificial intelligence with deep optics and photonics. *Nature*, *588*(7836), 39–47.

Wikipedia Contributors. (n.d.). *Artificial intelligence*. Wikipedia. Retrieved October 25, 2023, from https://en.wikipedia.org/wiki/Artificial_intelligence

Wipo. (n.d.). *Impact of artificial intelligence and business concerns*. World Intellectual Property Organization. https://www.wipo.int/tech_trends/en/artificial_intelligence/ask_the_experts/techtrends_ai_firth.html

Zapanta, R. (2023, February 28). *The impact of ai on business*. Micro Sourcing. https://www.microsourcing.com/learn/blog/the-impact-of-ai-on-business/

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, *23*, 100224.

B Analysis Appendix

B.1 Parallel Trends Plots

B.1.1 Full Sample

Figure B.1: Complete parallel trends visualization: full sample

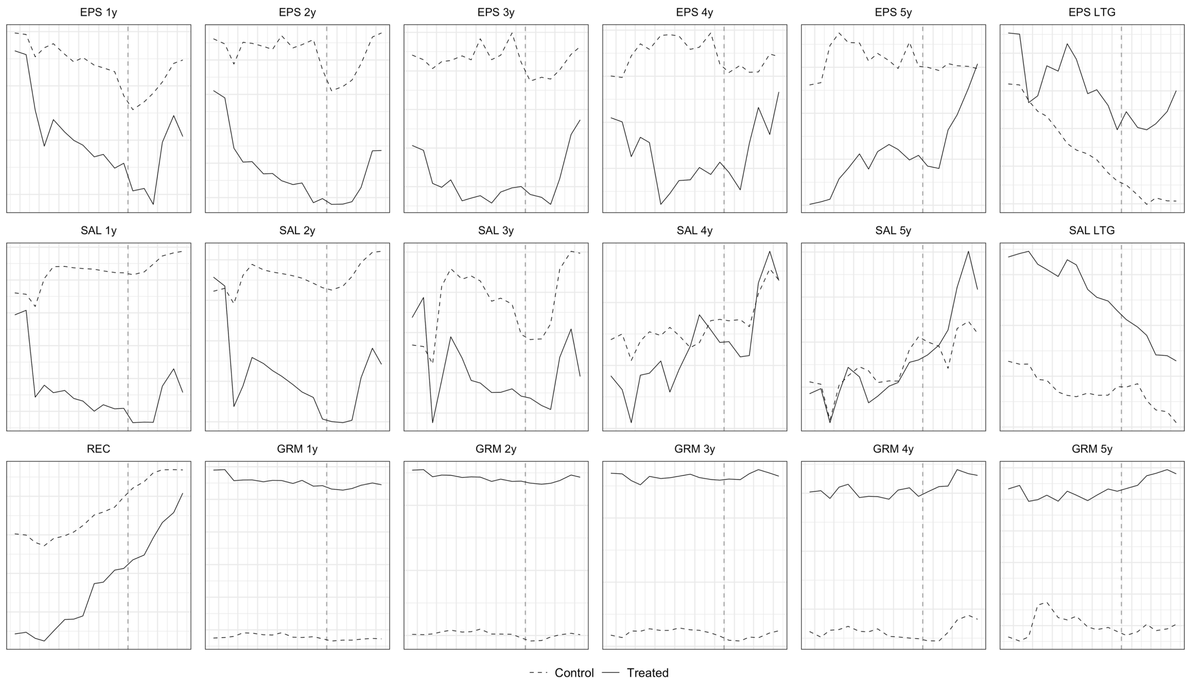

Figure B.1 displays the mean of all dependent variables for the full sample, grouped by treatment status. The treated group is the top 5% of firms ranked by AI-exposure (149 firms), and the control group is the bottom 50% of AI-exposed firms (1,481 firms). The plots include [t-12, t+6] months of data, where t marks the date of the ChatGPT release (November 30$^{\text{th}}$, 2022). Except for recommendations, all dependent variables have been winsorized at the 95% level.

As noted in the analysis section, there are some signs of the parallel trends assumption being violated, especially for longer horizons of EPS and sales forecasts. Thus, a formal test is conducted to verify the assumption for models with significant output (EPS 3y, SAL 1y, SAL LTG, and REC). This test is performed only for the matched sample and can be found in the analysis Section 5.4.

B.1.2 Matched Sample

Figure B.2: Complete parallel trends visualization: matched sample

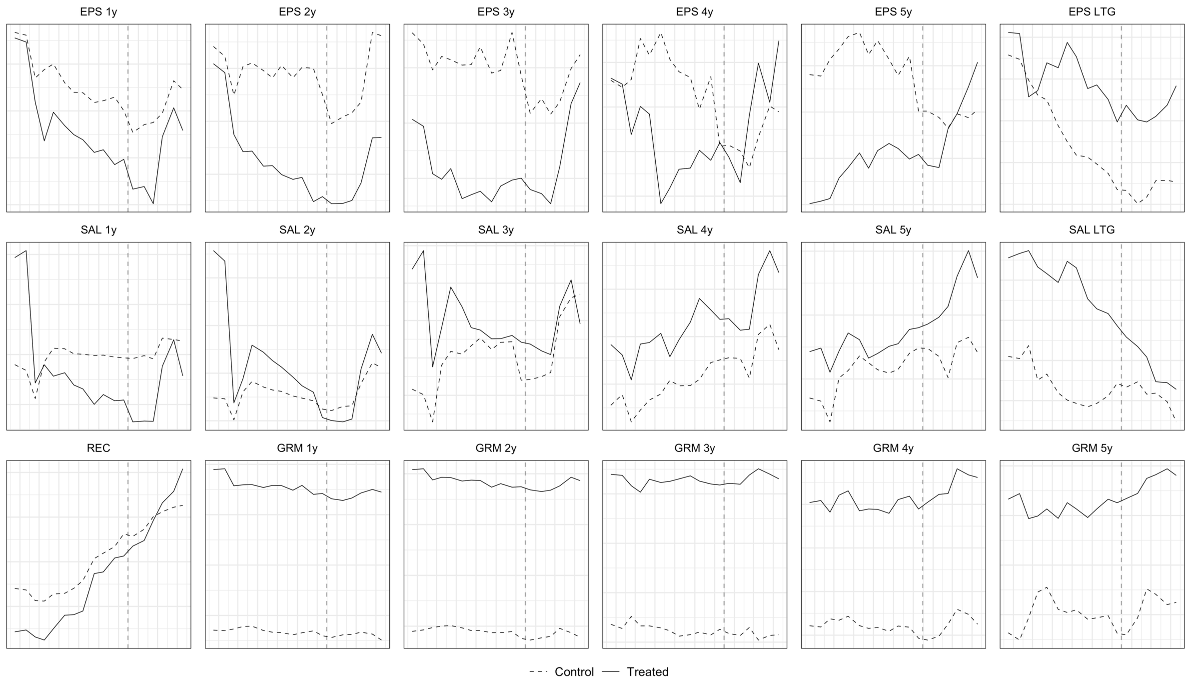

Figure B.1 displays the mean of all dependent variables for the matched sample, grouped by treatment status. The treated group is the top 5% of firms ranked by AI-exposure (149 firms), and the control group is the bottom 50% of AI-exposed firms (1,481 firms). The plots include [t-12, t+6] months of data, where t marks the date of the ChatGPT release (November 30$^{\text{th}}$, 2022). Except for recommendations, all dependent variables have been winsorized at the 95% level.

As noted in the analysis section, there are some signs of the parallel trends assumption being violated, especially for longer horizons of EPS and sales forecasts. Thus, a formal test is conducted to verify the assumption for models with significant output (EPS 3y, SAL 1y, SAL LTG, and REC). This test can be found in the analysis Section 5.4

B.2 Summary Statistics Matched Sample

Table B.1: Summary statistics: full versus matched sample

Panel A: Pre-treatment covariates: full sample

	Control					Treated				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Market to Book	1,481	2.65	4.15	−7.43	29.34	149	5.17	6.26	−7.43	29.34
Intan./Total Assets	1,481	0.20	0.21	0.00	0.73	149	0.26	0.23	0.00	0.73
log(Market Cap.)	1,481	7.26	2.02	1.68	12.55	149	7.71	2.62	1.68	12.55
Debt to Equity	1,481	0.98	1.66	−3.73	6.98	149	0.69	1.51	−3.73	6.98

Panel B: Pre-treatment covariates: matched sample

	Control					Treated				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Market to Book	447	3.44	5.54	−7.43	29.34	149	5.17	6.26	−7.43	29.34
Intan./ Total Assets	447	0.24	0.23	0.00	0.73	149	0.26	0.23	0.00	0.73
log(Market Cap.)	447	7.35	2.02	1.68	12.17	149	7.71	2.62	1.68	12.55
Dept to Equity	447	0.69	1.68	−3.73	6.98	149	0.69	1.51	−3.73	6.98

Panel C: Country and industry distributions (% of sample): full sample

	Country			Industry	
	Control	Treated		Control	Treated
Information Technology	6.95%	56.38%	USA	75.08%	83.89%
Telecommunication Services	2.70%	15.44%	Israel	0.47%	5.37%
Industrials	19.58%	10.74%	France	1.28%	1.34%
Health Care	10.87%	8.05%	Great Britain	1.82%	1.34%
Consumer Discretionary	15.33%	4.70%	Netherlands	0.47%	1.34%

Panel D: Country and industry distributions (% of sample): matched sample

	Country			Industry	
	Control	Treated		Control	Treated
Information Technology	21.25%	56.38%	USA	86.58%	83.89%
Telecommunication Services	8.05%	15.44%	Israel	1.34%	5.37%
Industrials	14.77%	10.74%	France	0.89%	1.34%
Health Care	25.73%	8.05%	Great Britain	1.57%	1.34%
Consumer Discretionary	12.75%	4.70%	Netherlands	0.45%	1.34%

Table B.1 displays summary statistics for control variables grouped by treatment status. Panel A and C are for the full sample, and Panel B and D are for the matched sample. The full sample refers to using the bottom 50% of firms ranked by AI-exposure as the control group. All continuous variables are winsorized at the 95% level. For the sake of readability and presentation, we report the 2-digit global industry classification (GICS) in this table. In the matching process, the 4-digit code is used. For time-varying covariates, the latest value prior to the release of ChatGPT is used. The country/industry distributions display only the treated group's top 5 countries and industries.

From studying the means of covariates between treated and control groups, it is clear that the matching process provides a better balance of pre-treatment covariates than using the full sample. This is especially true for debt to equity and industry, both of which are significant predictors of treatment in the logit model used in the matching process.

B.3 Regression output: Full Sample

B.3.1 EPS Forecasts

Table B.2: EPS forecasts regression output: full sample

	EPS forecast horizon					
	1y	2y	3y	4y	5y	LTG
Treat x Dec. 22	0.006	0.014	**0.019***	0.005	−0.018	0.334
	(0.013)	(0.013)	(0.008)	(0.012)	(0.024)	(1.782)
Treat x Jan. 23	0.006	0.014	**0.021****	−0.002	−0.017	−0.508
	(0.012)	(0.013)	(0.008)	(0.013)	(0.025)	(1.181)
Treat x Feb. 23	0.013	0.013	**0.014***	0.008	−0.010	−0.224
	(0.012)	(0.012)	(0.006)	(0.016)	(0.026)	(1.294)
Treat x Mar. 23	**0.032***	0.012	**0.015***	0.013	0.006	0.387
	(0.014)	(0.011)	(0.006)	(0.016)	(0.017)	(1.448)
Treat x Apr. 23	0.022	0.011	**0.016****	0.002	−0.004	1.159
	(0.015)	(0.011)	(0.006)	(0.017)	(0.023)	(1.439)
Treat x May 23	**0.030***	0.013	**0.018****	0.011	−0.003	2.168
	(0.014)	(0.010)	(0.007)	(0.016)	(0.022)	(1.609)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Num.Obs.	19 033	18 904	15 564	9013	6025	6047
R2 Adj.	0.907	0.916	0.853	0.845	0.842	0.835

Clustered (firm) standard errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, .:0.1*

Table B.2 displays regression output from the full sample DiD model where the post-treatment period is segmented into months, displaying the average treatment effect each month after the release of ChatGPT. The full sample refers to using the bottom 50% of firms ranked by AI-exposure as the control group. The results are similar to the findings for the matched sample and are therefore not commented on in the analysis section. The exceptions are some weak indications of an upward adjustment on the 1y horizon, but we have more confidence in the validity of matched sample models.

The dependent variable for models 1y - 5y is EPS forecasts normalized by share price and has the interpretation of forecasted percentage earnings yield. The dependent variable for the LTG model is long-term growth forecasts as reported in IBES. Thus, being AI-exposed in a specific month after release is associated with a $(100 \times \beta_i)\,\%$ change in earnings yield for 1y - 5y models, and a $\beta_i\%$ change in LTG for the LTG model. Dependent variables are winsorized at the 95% level. The sample consists of [t-6, t+6] months of forecast data, where t marks the date of the ChatGPT release (November 30$^{\text{th}}$, 2022).

B.3.2 Sales Forecasts

Table B.3: Sales forecasts regression output: full sample

	Sales forecast horizon					
	1y	2y	3y	4y	5y	LTG
Treat x Dec. 22	−0.005	−0.014	**-0.024****	**-0.034***	**-0.042***	**-1.866*****
	(0.012)	(0.009)	(0.009)	(0.015)	(0.017)	(0.455)
Treat x Jan. 23	−0.005	**-0.018.**	**-0.032****	**-0.046****	**-0.047****	**-2.402*****
	(0.013)	(0.010)	(0.010)	(0.016)	(0.018)	(0.570)
Treat x Feb. 23	0.024	0.009	**-0.023.**	−0.026	−0.023	**-2.289****
	(0.015)	(0.012)	(0.013)	(0.020)	(0.023)	(0.847)
Treat x Mar. 23	**0.103***	0.024	−0.008	0.004	−0.008	**-2.317***
	(0.040)	(0.016)	(0.019)	(0.021)	(0.023)	(0.906)
Treat x Apr. 23	**0.088***	**0.030.**	0.000	−0.004	0.017	**-2.639****
	(0.038)	(0.017)	(0.017)	(0.021)	(0.030)	(0.959)
Treat x May 23	**0.074***	0.022	−0.047	−0.006	0.019	**-2.731****
	(0.035)	(0.018)	(0.041)	(0.020)	(0.030)	(0.994)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Num.Obs.	19 027	18 894	15 666	9545	6926	4209
R2 Adj.	0.992	0.995	0.994	0.994	0.991	0.859

Clustered (firm) standard errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, .:0.1*

Table B.3 displays regression output from the full sample DiD model where the post-treatment period is segmented into months, displaying the average treatment effect each month after the release of ChatGPT. As is commented on in the analysis section, the full sample results differ from the matched sample results in that there are additional signs of downward adjustments on longer yearly horizons. The results for the 1y and LTG horizons are consistent with the findings from the matched sample model.

The dependent variable for models 1y - 5y is the logarithm of sales forecasts for each horizon. The dependent variable for the LTG model is long-term growth forecasts, as reported in IBES, which is the expected average growth rate over the coming business cycle. Thus, being AI-exposed in a specific month after release is associated with a $(100 \times \beta_i)\,\%$ change in forecasts for 1y - 5y models, and a $\beta_i\%$ change in LTG for the LTG model. Dependent variables are winsorized at the 95% level. The sample consists of [t-6, t+6] months of forecast data, where t marks the date of the ChatGPT release (November 30$^{\text{th}}$, 2022).

B.3.3 Gross Margin Forecasts

Table B.4: Gross margin forecasts regression output: full sample

	Gross margin forecast horizon				
	1y	2y	3y	4y	5y
Treat x Dec. 22	0.061	**0.376.**	**0.618.**	−0.059	−0.365
	(0.290)	(0.213)	(0.345)	(0.442)	(0.541)
Treat x Jan. 23	−0.008	0.288	0.467	−0.108	−0.491
	(0.331)	(0.230)	(0.352)	(0.461)	(0.560)
Treat x Feb. 23	0.042	0.239	0.387	−0.195	−0.081
	(0.368)	(0.266)	(0.413)	(0.505)	(0.713)
Treat x Mar. 23	0.128	0.181	0.694	0.656	0.052
	(0.530)	(0.341)	(0.534)	(1.061)	(0.931)
Treat x Apr. 23	−0.046	0.233	0.121	−0.141	**-1.406.**
	(0.643)	(0.378)	(0.586)	(0.824)	(0.849)
Treat x May 23	0.143	0.178	−0.073	−0.132	**-1.328.**
	(0.678)	(0.486)	(0.538)	(0.810)	(0.763)
Firm FE	Yes	Yes	Yes	Yes	Yes
Year-month FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Num.Obs.	14 007	13 855	10 584	5340	3178
R2 Adj.	0.973	0.977	0.975	0.974	0.984

Clustered (firm) standard errors in parathesis
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, .:0.1*

Table B.4 displays regression output from a full sample DiD model where the post-treatment period is segmented into months, displaying the average treatment effect each month after the release of ChatGPT. There are some weak indications of an upward adjustment on the 2y and 3y horizons and a downward adjustment on the 5y horizon. However, statistical significance is weak, and we have more confidence in the full sample models. Therefore, these results are not commented on in the analysis section.

The dependent variable is gross margin forecasts, winsorized at the 95% level. The coefficients can be interpreted as the β_i% change in forecasts from being in the treated group in a specific month after release. The sample consists of [t-6, t+6] months of forecast data, where t marks the date of the ChatGPT release (November 30$^{\text{th}}$, 2022).