Quantifying Generative AI's Influence on European Firm Valuation

Kevin Selling 25314, Viktor Sjöström 25472

January 8, 2024

Abstract

This paper employs a quantitative approach to evaluate the influence of generative AI workforce exposure on firm value in light of recent AI advancements. We focus on public large-cap companies headquartered in the European Union, analyzing their workforce exposure to these AI technologies. Utilizing an industry-neutral long-short portfolio, our research contrasts the firms with the highest generative AI exposure against those with the lowest. Findings indicate a notable impact on firm valuation up until one month after the release of ChatGPT-3. Moreover, the effects following the release of ChatGPT-4 appear to be greater in comparison to a similar release period, while less definitive due to robustness concerns.

Keywords: Artificial Intelligence, AI, Generative AI, ChatGPT, Workforce Exposure, Corporate Valuation, Labor Inputs, Productivity Improvement, Large Language Models

Tutor: Milda Tylaite, Assistant Professor at the Department of Accounting

Acknowledgements: We would like to thank our tutor Milda Tylaite for her superb guidance throughout the process of writing this thesis. Furthermore, we would like to thank our fellow students who provided valuable feedback during our supervision sessions.

Contents

1 Introduction and Research Question

The rapid advancement of Artificial Intelligence (AI) has recently gained significant attention, sparking new ideas about the future of work. The public releases of Large Language Models (LLMs), such as ChatGPT and Bard, have not only extended the scope of AI capabilities but also opened up new opportunities for firms to leverage these innovations. In academic literature, we find evidence supporting the notion that these models can significantly impact businesses by improving the quality and quantity of output when applied appropriately (Noy and Zhang, 2023; Brynjolfsson et al., 2023; Dell' et al., 2023). While we know more about how these technologies might impact various businesses, we know less about the actual value impact on firms. One study exploring this topic is *Generative AI and Firm Value*, a working paper from the National Bureau of Economic Research (NBER) by Eisfeldt et al. (2023). This United States (U.S.)-based paper explores the concept of labor exposure to generative AI and how the composition of the workforce influences firm valuation after the release of ChatGPT-3. The paper found that firms with higher exposure to generative AI realized daily excess returns 0.4% higher than those with lower exposure. The effect was seen for the following two weeks after the release of ChatGPT-3. Furthermore, the effect was not only significant within industries but also across different business sectors suggesting a widespread effect. Inspired by the emerging role of generative AI in shaping firm valuation, this thesis aims to investigate the impact of labor exposure to generative AI within the context of the European public large-cap market. Our approach involves constructing and analyzing industry-neutral quintile portfolios of 500 large-cap firms headquartered within the European Union (EU). To assess the impact, we investigate how the returns of these distinct portfolios fluctuate before and after the releases of ChatGPT-3 and ChatGPT-4.

1.0.1 AI and Generative AI

Artificial Intelligence (AI) can be described as a general-purpose technology, meaning it can be used for many different purposes. Essentially, it is a collection of tools that enable computers to reason, learn and act in a similar way to humans (Andrew Ng, 2023). The most common branch within AI is

machine learning (ML), which is a program or system that trains a model on input data. The trained model can make useful predictions based on new data points and gives the computer the ability to learn without explicit programming (Dr. Gwendolyn Stripling, 2023).

Recent advancements within the field have given significant attention to a particular type of AI - generative AI. Generative AI, exemplified by Large Language Models (LLMs), such as ChatGPT, utilizes a form of machine learning to process and generate data. These models are capable of understanding and generating human-like output in reaction to a given prompt. The models start by finding a suitable answer in a particular dataset. Thereafter, it formulates the answer by predicting subsequent words in a sequence by treating individual words and sentences as distinct data points. For example, in the sentence “I am currently writing a...”, the model could predict that an appropriate next word would be "thesis." or "book.", based on patterns in its training data. These systems work in the same way when generating pictures by predicting pixels, or when generating code by predicting the next character (Andrew Ng, 2023).

It is important to note generative AI is not a new concept. For instance, Google Translate, released in 2006, uses generative AI techniques to transform text from one language to another. Another example is the assistant “Siri”, introduced by Apple in 2011, which uses generative capabilities in order to understand and generate human-like responses (Mirella Lapata, 2023). However, during the last year, generative AI has made significant improvements and has gained considerable attention with the release of ChatGPT-3 on November 30, 2022.

ChatGPT showcased remarkable capabilities within generative AI, as it could be used as a tool to generate high quality text, images, code, generate ideas, identify business opportunities and solve problems in a way that was not possible before. Within 2 months, ChatGPT reached 100 million users. To put this in perspective, it took 78 months for Google Translate, 55 months for Spotify and 30 months for Instagram to reach the same user base. (Mirella Lapata, 2023)

The landscape of generative AI was further revolutionized with the release of the enhanced version, ChatGPT-4, on March 14, 2023. This model showed further improvement in its capabilities in understanding and interacting with more complex problems. At the release, OpenAI claimed that ChatGPT-4 could beat 90% of humans on the SAT, and get top marks in Law and Medical exams (Mirella Lapata, 2023). Furthermore, ChatGPT-4 significantly reduces hallucinations relative to ChatGPT-3, and scores 40% higher on factuality evaluations (OpenAI, 2023). This suggests that the release of this model should be considered a major event when analyzing the development of generative AI.

1.0.2 Definition

There are many definitions of artificial intelligence, reflecting the diverse and continuously evolving nature of the field. For the purpose of this thesis, which explores recent advancements in generative AI and their implications for firm valuations, we have adopted a similar definition of AI and generative AI, as the one used by Brynjolfsson et al. (2023):

In our study, "artificial intelligence" is defined as a computer system capable of sensing, reasoning or acting like a human. "Machine learning" is defined as a subset of AI, which uses algorithms that to learn from data, identify patterns, and make informed decisions or make predictions. This process is done without the need for explicit programming. LLMs and tools built around LLMs, for example, ChatGPT, are an application of ML. These models are capable of generating new content and are a form of "generative AI".

This definition encapsulates the essential capabilities of generative AI models and is therefore a seemingly appropriate definition for this thesis. All references to AI and generative AI in this thesis therefore adopt this definition. Accordingly, all references to AI within this thesis, if not mentioned

otherwise, are to be understood in the context of this definition. References to exposure score are to be understood as generative AI exposure score if not stated otherwise.

1.1 Purpose and Contribution

While public access to generative AI has considerably expanded over the last year, research on its effects on market valuation remains scarce. A notable study by Eisfeldt et al. (2023) investigates these effects in the U.S. public market, revealing significant impacts on firm value. However, this research focus is predominantly U.S.-centric, leading to a gap in the literature regarding other geographical areas. Our research aims to fill this gap by investigating the impact on firms within the EU, a region distinctly different from the U.S. in market dynamics and regulatory frameworks. This is particularly relevant given the European Parliament's significant step on June 14, 2023, towards enacting the world's first comprehensive AI law (European Parlament, 2023), and the EU's stringent data privacy regulations. We will employ a similar method to that of Eisfeldt et al. (2023) to explore how exposure to generative AI influences EU firms' market value. This thesis aims not only to contribute to a better understanding of the cross-border implications of AI but also to offer insights into how different regulatory environments shape AI's economic impact. By comparing these findings with existing U.S. data, our research could provide a comprehensive view of AI's role in global market valuation.

Hence, by looking at the European large-cap public market, we try to answer the following research question: *What are the effects of recent advances in generative AI on the value of firms within the European Union?*

By answering this question, we aim to explore the broader implications of generative AI in a diverse regulatory landscape. This study provides insights into both the immediate and longer-term financial impacts following the release of ChatGPT.

1.2 Scope of Investigation

The scope of this investigation is limited to 500 large-cap companies whose headquarters are located within the EU. This limitation ensures that all examined companies are subject to the overarching EU regulatory framework. However, it is acknowledged that these companies may also be influenced by the specific national laws of the EU member state in which they are headquartered. Consequently, European companies from non-EU countries such as the United Kingdom, Norway, and Switzerland have been excluded, as they are not subject to the same EU regulations. As such, any reference to Europe in this study specifically pertains to the European Union (EU).

The study analyzes daily company return data ranging from 2022-10-01 to 2023-09-30 and focuses on comparing the period before and after the release of ChatGPT-3 (30 November 2022) and ChatGPT-4 (14 March 2023). The chosen time frame allows for an assessment of both immediate and longer-term market reactions to these generative AI models' public releases.

2 Literature Review and Theory

The following chapter aims to review existing literature and theories concerning the impact of generative AI on firm performance. Worth noting, is that while the topic of AI contains a substantial amount of research, the recent advancements within generative AI have opened up a new field of investigation. This entails that literature on generative AI has predominantly emerged in recent years, which might introduce limitations due to its novelty. Such limitations include scarcity of academic publications and challenges in peer-reviewing. Nevertheless, this also ensures that the impact is evaluated based on the most current findings in a rapidly evolving field. The review is done by initially investigating the occupational-level implications of generative AI, focusing specifically on its effects on workforce productivity and effectiveness. This investigation is crucial for understanding how rapid advancements in AI are reshaping workforce dynamics and, consequently, why exposure to these technologies matters. Subsequently, the chapter transitions to an exploration of how firm valuation has been influenced by the workforce exposure of generative AI. Through this, we aim to provide both a comprehensive overview and a benchmark for understanding AI exposure's impact on firm valuation. Thereafter, the chapter examines literature that investigates potential disparities in the adoption of AI technologies between the U.S. and the EU. This segment aims to identify how the distinct approaches and regulatory environments in these regions influence a firm's ability to adopt generative AI. In doing so, we aim to provide insights into the differing landscapes and how that may influence firm valuation. Lastly, the chapter investigates relevant theories, utilizing frameworks and insights from the literature review, to inform our understanding of how exposure to generative AI may influence firm performance. By integrating and applying theories and models related to technological innovation, AI, and market reactions, we aim to construct a robust theoretical framework to explain the empirical findings of our study.

2.1 Literature Review

2.1.1 Generative AI in Workforce Productivity and Contribution

A study conducted by Noy and Zhang (2023) investigated generative AI's impact on productivity in writing tasks. They assigned occupation-specific writing tasks to 444 college-educated workers and exposed 50% of them to ChatGPT. The results found that workers exposed to ChatGPT improved their average productivity, decreasing the time to complete a task by 0.8 standard deviations while improving the quality of output by 0.4 standard deviations. This demonstrates that using generative AI improves workers' productivity and contribution. Similarly, a working paper by Brynjolfsson et al. (2023) explored the effects of generative AI on customer support agents. The study included data from 5,179 agents and revealed that the integration of generative AI tools led to a 14% average increase in the rate of issues resolved per hour. However, the study also observed a variance in impact across workers with differing skill levels improving the productivity of low-skilled workers by 35%, while leaving a small to no impact on high-skilled workers. This suggests that lower-skilled workers may experience a greater impact on efficiency. The study by Dell' et al. (2023) arrives at a similar conclusion, where below average performance workers tend to exhibit the most significant improvements in both quality and quantity of output when exposed to generative AI tools. However, Dell' et al. (2023) found that an improvement was seen across all skill levels. Furthermore, the article extends its analysis by categorizing tasks according to their compatibility with GPT-4's capabilities and examining the implications of this compatibility. The study revealed that workers utilizing GPT-4 for tasks exceeding its current scope, saw a 19 percentage point decrease in result accuracy. This was seen even if the task itself maintained a consistent level of complexity for humans.

From this, we can conclude that the usage of generative AI has consistently shown an increase in workforce efficiency. However, careful consideration should be taken when selecting its use cases as the quality of the output may vary depending on the task (Dell' et al., 2023). To further understand the underlying characteristics of such tasks, the study by (Eloundou et al., 2023) highlights characteristics associated with varying degrees of exposure to Large Language Models (LLMs). The

research identifies a negative correlation between exposure to LLMs and skills related to critical thinking. Conversely, there is a notable positive correlation between programming and writing skills. These findings imply that occupations requiring programming and writing skills are more influenced by LLMs compared to those demanding higher levels of critical thinking. Additionally, the study suggests that occupations involving physical tasks are not impacted by generative AI. Hence, occupations involving a higher proportion of physical tasks are expected to have lower exposure to generative AI.

It should be stated that one ought to be weary of the practical considerations in the used literature related to performance enhancements, as the existence of hurdles that accompany AI integration into the workplace has been observed. As noted by Ångström et al. (2023), 70% of companies reports minimal impact from AI investments and only 13% of data science projects reach production stages.

2.1.2 Generative AI and Firm Valuation

Building upon the premise that generative AI has the potential to significantly improve workforce effectiveness, the NBER paper *Generative AI and Firm Value* analyzes how this technology impacts firm valuation in the U.S. The study measures how workforce exposures to generative AI correlate with performance, analyzing value-weighted portfolios to capture economic effects. It finds that firms with higher exposure to AI earned an excess daily return of 0.4% compared to companies with lower exposure (Eisfeldt et al., 2023; Brynjolfsson et al., 2023; Noy and Zhang, 2023). Upon detailed examination, the validation of the exposure variable is conducted over a relatively short time period of 10 observable days. This limited time frame may impact the measure's reliability, as it could potentially overlook crucial trends that significantly influence the measure.

2.1.3 EU and U.S. Distinctions

Historically, AI adoption in the EU has progressed at a slower pace compared to the U.S., a trend that may possibly extend to the adoption of generative AI. This is mainly due to differing

regulatory landscapes and market dynamics (Hoffmann and Nurski, 2021). In the EU, the strict regulatory environment highlighted by the General Data Protection Regulation (GDPR), emphasizes data privacy. This regulation imposes limitations on data availability and usage, an essential component for AI development (Hoffmann and Nurski, 2021). Moreover, the anticipated AI Act in the EU underscores the region's focus on ethical and responsible AI deployment, suggesting a cautious approach to AI integration (European Parlament, 2023). While these regulations protect consumer privacy, they can potentially slow the pace of AI adoption and restrict the extent of AI experimentation as companies must adhere to such guidelines (Ångström et al., 2023). In contrast, the U.S. presents a more forgiving regulatory framework regarding data privacy and AI deployment. This environment encourages more rapid AI innovation and implementation, allowing for wider experimentation and quicker integration of AI technologies across various sectors (Hoffmann and Nurski, 2021).

2.2 Theory

As proposed by the literary review, multiple papers suggest that generative AI can, in successful cases, increase the efficiency of a firms labor force. However, one should remain cautious of when to use it as usage of such tools may have implications on the quality of output (Dell' et al., 2023). These findings serve as the basis for our theoretical framework. As such, we hypothesise that these improvements in efficiencies will increase firm-level free cash flows, as suggested by Eisfeldt et al. (2023). The article mentions that there are two possible scenarios leading to this. Firstly, the firm may utilize generative AI-based systems to fully automate existing tasks. In turn, this can be expected to reduce labor related input costs as these systems are generally cheaper than human labor. The second scenario suggests that some tasks are able to be augmented by the usage of generative AI, leading to increasing quantity and quality of outputs for the same labor input. Consequently, this behavior leads to lowered input costs as the completion time for each task is lowered. Furthermore, increasing the quality of outputs reduces costs related to quality corrections. Building on this understanding, we expect a direct impact on firm valuation, as underlined by the discounted cash

flow (DCF) model, which serves as a common approach for valuation (David T. Larrabee and Jason A. Voss, 2012). Firms that quickly incorporate these AI-driven efficiencies should see an earlier and more pronounced increase in their valuation due to the time value of money principle inherent in the DCF approach. In applying these findings to our research, we expect that U.S. firms, given their more rapid adoption of AI (see), to realize cash flow enhancements earlier than companies residing within the EU. These varying rates of adoption may result in distinct valuation trajectories, with firms headquartered in the EU experiencing a lesser impact in valuation from ChatGPT releases.

Furthermore, we anticipate that occupations involving physical tasks or requiring a significant degree of critical thinking will be less impacted by generative AI. In contrast, occupations centered around programming and writing skills are likely to experience a higher degree of exposure. (Eloundou et al., 2023)

Although prior literature has found, on average, that valuation is impacted positively by exposure to generative AI, there are reasons to believe that these impacts may vary in size between firms and industries (Eisfeldt et al., 2023). As suggested by Eisfeldt et al. (2023), firms may possess assets that diminish in value or become obsolete due to the introduction of ChatGPT-3 and 4. These assets, supposedly in the form of proprietary internal systems or offerings, which once provided a competitive edge to the firm, may have diminished in significance as their capabilities are increasingly matched or surpassed by those of ChatGPT-3 and 4. Furthermore, some industries may also see an increased threat of entry as competitive ideas become cheaper to execute. This may lead to a negative impact for incumbents finding themselves in an industry with a high degree of exposure, suggesting a market opportunity.

Another important aspect to consider in our analysis is the market reaction to a significant release, such as the releases of ChatGPT-3 and ChatGPT-4. Close to this concept lies the *Efficient Market Hypothesis* (EMH). This theoretical model assumes that markets are efficient and that the valuation

of a publicly traded firm reflects all publicly available information. This implies that when new information is available to the market, prices adjust instantly as investors modify their portfolios to incorporate the new information and its impact on the security. With the assumption that markets are efficient, we anticipate that firm valuations will adjust subsequent to the releases of ChatGPT-3 and ChatGPT-4, reflecting the influence of these tools on labor efficiency and contribution adjusting for risks.

It is also important to highlight that while exposure to generative AI doesn't guarantee its adoption, the competitive dynamics of the market and the clear advantages of a successful implementation outlined in contemporary studies suggest that exposed firms are inclined to implement these technologies (Michael E. Porter, 1996). However, it should be noted that numerous companies encounter challenges in implementing artificial intelligence (AI), suggesting the existence of a gap between the theoretical value-add of implementing AI and its practical application Ångström et al. (2023). While this might still be the case for firms today, it is important to recognize that this study is based on data which is from 2019, a period before ChatGPT. As such, one could argue that our exposure measure, which is based on more recent AI capabilities, is comparatively better fitting in illustrating the practical applications of AI implementation. This is most likely due to the easy-to-use and accessible nature of ChatGPT, in contrast to previous AI tools. Thus, we posit that the implementation challenges identified in the 2019 study may not be as pertinent in the context of ChatGPT.

3 Method

The following chapter outlines the methodology used in the study to answer the question: *What are the effects of recent advances in generative AI on the value of firms within the European Union?*. Our approach builds upon the work by Eisfeldt et al. (2023), with slight deviations to handle any inconsistencies between our datasets. Initially, we assess the exposure of each company to generative AI. To achieve this, we first determine an exposure level for seven different occupational groups. By utilizing data on the occupational group distribution within each company, a company-level exposure score can be computed using a weighted average approach. Furthermore, these scores are validated through linear regression analysis on various firm ratios to ensure that our measure effectively explains what it is intended to. The ratios investigated are Return on Assets (ROA), Tangibility, and Labor Intensity. Finally, we explore the correlation between firms' exposure to generative AI and their market valuation before and after significant generative AI events. Echoing the findings of Eisfeldt et al. (2023), we identify the launch of ChatGPT-3 as a pivotal event in this context. Additionally, recognizing the enhanced capabilities, this thesis extends the investigation to include the impact of ChatGPT-4. The analysis involves constructing industry-neutral quintile portfolios based on generative AI exposure and assessing their relative performance over seven distinct time intervals. This approach aims to capture the relationship between generative AI exposure and firm performance.

It should be noted that for all regressions in our study, significance levels are computed using two-tailed tests assuming a t-distribution for the test statistics. This is consistent with the methodology used by Eisfeldt et al. (2023). Significance levels are evaluated in accordance with Ivanova et al. (2023) where ****, ***, ** and * represent significance at the 0.1%, 1%, 5% and 10% levels respectively.

3.1 Calculating Firm Exposure to Generative AI

The subsequent section is divided into two parts, outlining the steps taken to achieve the exposure of each individual firm to generative AI. Represented by a score from 0 to 100, the measure is designed to rank firms based on their relative exposure to generative AI where a higher score indicate a greater exposure. Given that the concept of AI exposure score is not widely recognized or standardized, this paper calculates the exposure scores using a similar methodology as the study by Eisfeldt et al. (2023).

3.1.1 Occupational Exposure Score

The first step in calculating the firm level exposure score is deriving it from an occupational level. Employing data from the International Standard Classifications of Occupations (ISCO-08) database, sourced from the International Labour Organization (ILO), this study conducts a detailed examination of the key tasks associated with each occupation. The ISCO-08 framework was chosen over the O*NET system used by Eisfeldt et al. (2023), following the recommendation by the European Commission (2009). The choice of framework facilitates a cross-country analysis, aligning more appropriately with the European labor context, rather than relying on data specific to the U.S. labor market. To ensure the highest accuracy in the analysis, tasks were analyzed at a level-4 precision, which represents the most detailed occupational classification level in the ISCO-08 framework. In total, 3305 tasks were classified and distributed across 427 occupations.

The study adopts a similar classification system to Eisfeldt et al. (2023) to measure the degree to which each task is exposed to automation or enhancement through generative AI. It is important to note that both quality and efficiency dimensions are incorporated into our classifications to account for compromises in quality for some tasks, as highlighted in our literature review (Dell' et al., 2023). More specifically, the classifications are labeled as *C0*, *C1*, *C2* and *C3*, where *C* denotes *classification*. The details for each classification are as follows:

C0 - No Exposure: This refers to tasks that remain unaffected by AI capabilities, such as those needing physical interaction, detailed visual review, or mandatory human legal intervention. These tasks neither benefit from AI assistance nor experience reduced completion times due to AI tools.

C1 - Direct LLM Efficiency: Tasks that are likely to experience an efficiency improvement of over 50% through the application of existing Large Language Models (LLMs), like ChatGPT, without compromising quality. These tasks primarily involve text and code manipulation, translation, summarizing, feedback provision, question generation and other text-based activities. Direct access to AI models, like ChatGPT, is sufficient to complete these tasks with sufficient quality.

C2 - LLM Powered Applications: Tasks where 50% product gain is possible by using additional software built on top of existing LLM models while preserving existing quality standards. These tasks include summarizing extensive documents, retrieving up-to-date information, making data-driven recommendations, and analyzing written information, among others.

C3 - Image Capabilities: Tasks where supplementary image processing is needed to realize a 50% efficiency gain and maintain the same quality as before. Examples of these tasks include reading and extracting text from images, generating and processing images, or editing images as per specific instructions.

To categorize each task into these classifications, we developed a script using Python and integrated the *gpt-4-0613* model with an API key, provided by OpenAI. This gives the program access to the same underlying model as ChatGPT-4 as of June 13, 2023. We were then able to run the model using parameters consistent with those used in the study by Eisfeldt et al. (2023). More specifically, the model was given a temperature of 0, which in the context of LLM-models regulates the level of randomness or creativity of the output from the model. A temperature of 0 will ensure that its responses are highly deterministic, producing the most likely output with minimal

randomness. Additionally, two prompts will be used to guide the model, a system prompt and a user prompt. The system prompt provides specific instructions, establishing a framework for the model to operate within. Within this prompt, explicit directives will be incorporated outlining how the model should categorize each task. The user prompt will be the task description of each occupation that the model needs to respond to. To ensure the model's accuracy, we run the program three times, allowing for a comparison of outputs across these runs. The results from the first run are utilized in our analysis. For further clarification, details regarding the model's specific setup and the prompts utilized are provided in Appendix A along with a sample of the output in Appendix C, Table 9. Additionally, information on the model's accuracy level can be found in Appendix C, Table 8a.

After classifying each task into the different categories, the study applied the following formula to calculate the occupational level exposure score:

$$E_O = \frac{N_1 + 0.5 \cdot N_2}{N_0 + N_1 + N_2 + N_3} * 100$$

The formula can be broken down into the following components:

- E_O: The exposure score
- N_0: Total number of tasks within an occupation classified as *C0*
- N_1: Total number of tasks within an occupation classified as *C1*
- N_2: Total number of tasks within an occupation classified as *C2*
- N_3: Total number of tasks within an occupation classified as *C3*

Dividing the numerator by the denominator gives the proportion of tasks within an occupation that are impacted by LLM, taking into account the full impact of tasks classified as 1 and half the impact of tasks classified as 2. Note that we also multiply the proportion by 100 for easier analytical interpretation. The result is a measure that represents the exposure score for each occupation from 0 to 100, where a higher score indicates a greater exposure.

3.1.2 Firm Level Exposure Score

Once the exposure score is calculated on an occupational level, we map the score to the firm level using data provided by RevelioLabs. Occupational distribution data is derived from a snapshot taken on 15 October 2022, representing the distribution before the release of ChatGPT-3. RevelioLabs provides a breakdown of the occupational structure within a firm, categorizing it into seven categories: *Sales*, *Engineering*, *Administration*, *Operations*, *Science*, *Marketing*, and *Finance*. The platform sources its data from platforms such as LinkedIn and other resume profiles, and employs sophisticated proprietary methodologies to adjust and refine the data, ensuring its accuracy and reliability. Due absence of information for certain European firms, this study focuses on the 602 largest companies headquartered within the European Union. This approach is employed to derive exposure scores for a representative sample of 500 companies.

Given that ISCO-08 and RevelioLabs utilize different occupational categorization methods, we developed a Python script to categorize each occupation from the ISCO-08 framework into one of the seven categories provided by RevelioLabs. To ensure consistency, the program utilized the same *gpt-4-0613* model used for task categorization. Similarly, the model operated with a temperature setting of 0 and responded to both a system prompt and a user prompt. The system prompt provided detailed instructions for categorizing occupations, while the user prompt included the occupational title and its associated tasks from the ISCO-08 list. The specific setup and prompts used for this model can be found in Appendix A, along with a snippet of the output as an illustrative example in Appendix C, Table 2. Similar to the script used for the categorization of tasks, we ran the program three times to ensure its accuracy. A comparison of the runs is presented in Appendix C, Table 8b.

From here, we calculate the category level exposure score by taking the average score for each of the seven occupational groups. The exposure score for each of the occupational groups can be found in Appendix B, Table 4. Subsequently, the firm-level exposure score, E_f, was calculated using the following formula:

$$E_f = \sum_{\text{Occupational Groups in } f} (OGShare_{f,O} \cdot E_{OG})$$

where E_{OG} represents the exposure score of each occupational group, and $OGShare_{f,O}$ the share for each occupational group within firm f. This method offers a comprehensive analysis of the firm's overall AI exposure by considering the specific AI exposure level of each occupational group relevant to the firm and its relative proportions within the firm.

3.2 Validation of Exposure Variable

Although our methodology for generating the exposure score variable is similar to that of Eisfeldt et al. (2023), differences in data quality and other discrepancies between the studies could influence the construction of this variable. To examine the behavior of our measure in relation to the findings of Eisfeldt et al. (2023) we conduct an analysis of various firm ratios. More specifically, the ratios that will be examined in this paper are:

Return on Assets (ROA): Financial metric that measures a company's profitability in relation to its total assets. This ratio reflects how effectively a company generates earnings from its asset base. Previous studies have established a significant negative correlation between ROA and a company's exposure to generative AI technology(Eisfeldt et al., 2023). This suggests that firms with more exposure to generative AI tend to have a lower return on assets. ROA is calculated by using the following formula:

$$\text{ROA} = \frac{\text{Net Income}}{\text{Total Assets}}$$

Tangibility: Tangibility quantifies the proportion of a company's property, plant, and equipment (PPE) in relation to its total assets. It is an indicator of a company's investment in physical assets relative to its overall asset base. To be consistent with the definition used in prior literature, tangibility is calculated using the logarithm of the ratio of PPE to Total Assets (Eisfeldt et al., 2023). Previous literature supports a negative relationship between the ratio and a firm's exposure to generative AI.

Hence, the formula for calculating tangibility is as follows:

$$\text{Tangibility} = \log\left(\frac{\text{PPE}}{\text{Total Assets}}\right)$$

Labor Intensity: Measures the extent to which human labor, as opposed to capital, is employed in the production of goods or services. It is quantified by comparing the number of employees to the capital invested in the production process. A labor-intensive process is marked by a higher proportion of labor costs relative to capital investment. Previous research has identified a positive and significant correlation between labor intensity and a firm's exposure to generative AI technology (Eisfeldt et al., 2023). This suggests that companies with greater workforce exposure to generative AI tend to exhibit higher labor intensity. To calculate labor intensity, the following formula is used:

$$\text{Labor Intensity} = \frac{\text{Number of Employees}}{\text{PPE}}$$

The variables used for computing these ratios, *Total Assets, PPE, Number of Employees, Net Income, and Firm Industry*, are derived for each firm from Capital IQ, using the latest annual reporting data as of October 19, 2023.

Moreover, to conduct the analysis we use two different regression models for each of the ratio variables. Firstly, we conduct a univariate regression analysis to evaluate the direct relationship between the firm level exposure score as the dependent variable and the ratio as the explanatory variable. This simple regression provides insights into the basic correlation between these two factors. Subsequently, we implement a fixed effects regression model, which is designed to adjust for variations across the different industries. This approach allows us to account for industry-specific characteristics that might influence the relationship between the exposure score and the examined ratio.

To ensure the accuracy and credibility of our regressions, we modify the data by excluding any

companies that have missing values. It is important to note that this removal was specific to each regression analysis; thus, missing values in variables not included in a particular regression will not lead to the exclusion of companies from that analysis. Moreover, our fixed effects model includes a further adjustment by only including industries that are represented by more than six companies. This criterion is set to ensure sufficient data representation within each industry, thereby enhancing the reliability of our findings. By implementing this threshold, we aim to avoid skewed results that could arise from analyzing industries with too few data points, which might not accurately reflect broader industry trends.

Furthermore, we use Cook's Distance to identify influential observations. By comparing results before and after adjustments we want to ensure the robustness of our tests. More specifically, we employ a Cook's Distance threshold value of $4/n$, where n represents the number of observations (Tom Van der Meer and Pelzer, 2006). Values above this threshold are considered influential points. This cut-off point is carefully selected due to its dynamic nature, enabling an adaptable threshold that varies proportionally with the size of the dataset. Such an approach ensures that our identification of influential points remains sensitive and tailored to the specific characteristics of our data, thereby enhancing the reliability of our models. By adjusting for these influential points, we aim to gain insights into the model's sensitivity to specific data points.

Lastly, we address potential multicollinearity issues by conducting a Variance Inflation Factor (VIF) tests for the different fixed effects models. As suggested by Peter Kennedy (2008), harmful collinearity is identified for VIF-values greater than 10. This ensures that our model estimations are not overly influenced by high inter-correlations among the predictors, further ensuring the validity of the models.

3.3 Mapping Firms Exposure Score to Firm Valuation

The following section outlines the methodology used to explore how a firm's exposure to generative AI impacts its market valuation before and after the releases of ChatGPT-3 and ChatGPT-4. It is structured into three key parts. The first part explains the construction of industry-neutral quintile portfolios, providing the basis for our analysis. The second part introduces the two linear regression models used to analyze the relationship between generative AI exposure and portfolio performance. In the last part, we describe the time periods used to assess both immediate and long-term market effects.

3.3.1 Portfolio Construction

Industry neutral portfolios are constructed by first categorizing each firm into 20 different industries based on the *North American Industry Classification System* (NAICS) using 2-digit precision sourced from Capital IQ. The classification system is selected for comparability with the findings of Eisfeldt et al. (2023) and the 2-digit precision is chosen to ensure sufficient sample sizes within each industry. For each industry, firms are ranked by their exposure scores into five industry quintiles. Industry Quintile 1 (IQ1) includes the firms with the top 20% highest exposure scores, while Industry Quintile 5 (IQ5) comprises the firms with the 20% lowest exposure scores. These industry quintiles are then consolidated across all segments to form new quintile portfolios, Q1 to Q5, which represents a cross-section of firms from each industry.

In order to retrieve the returns for each portfolio, the performance of each firm within a quintile was weighted according to its market capitalization, meaning larger firms exert a greater influence on the portfolio's overall performance.

For analytical purposes, this thesis chooses to focus on the portfolio extremes, specifically Q1 and Q5. In addition, the study leveraged the concept of an 'Artificial Minus Human' (AMH) portfolio, as outlined by Eisfeldt et al. (2023). This AMH portfolio is constructed by adopting an equal-weighted

long position in the Q1 portfolio, characterized by higher generative AI exposure, and a short position in the Q5 portfolio, characterized by lower generative AI exposure. The purpose of this AMH portfolio was to give insight into disparities driven by differential AI exposure. This methodology allows for a nuanced analysis of how varying degrees of generative AI exposure correlate with firm valuation and performance.

3.3.2 Regression Models

To investigate the relationship between AI exposure and firm valuation, this thesis will employ two linear regression models. The first model, referred to as model A will take the following form:

$$r_{i,t}^{p} - r_{t}^{f} = \alpha_t + \varepsilon_{i,t}$$

where $r_{i,t}^{p}$ is the daily return of portfolio i at time t, $\varepsilon_{i,t}$ is the error term for portfolio i at time t, representing random effects that are not explained by model. Furthermore, r_{t}^{f} is the daily risk-free rate at time t. The regression estimates the average daily excess return of portfolio i net of the daily risk-free rate is captured by α, which is the intercept term of the model.

To estimate the risk-free rate, our paper utilizes data from the European Central Bank on the Euro yield curve spot rate with a 1-year maturity. This rate is based on nominal government bonds issued by Euro-area countries with a AAA credit rating. The focus on these high-credit-quality bonds is intended to align with our objective of capturing the dynamic and diverse economic environment of the Eurozone. The 1-year maturity of the risk-free rate has been selected to align with the time frame of the portfolio returns, to ensure the most accurate result (Lally, 2007). It is important to note, that not all EU countries are part of the Euro-area and thus do not contribute to this specific yield curve data.

The second model, referred to as model B, incorporates market risk adjustments by employing the Capital Asset Pricing Model (CAPM) (Sharpe, 1964), a fundamental framework for understanding

the relationship between systematic risk and expected return. This model extends the simpler model A by including a market factor-adjusted term, capturing the portfolio-specific abnormal returns above or below what would be predicted by CAPM. The model takes the following form:

$$r_{i,t}^{p} - r_{t}^{f} = \alpha_t + \beta_i(r_{t}^{M} - r_{t}^{f}) + \varepsilon_{i,t}$$

where the r_t^M is the daily market return at time t and β_i quantifies the sensitivity of portfolio i to market fluctuations. To derive a reasonable measure for the European market return, our analysis calculated an average of daily returns from a curated selection of European stock market indices. These indices were selected based on two key criteria. Firstly, to avoid the generality of a broad European index, our selection targets a market return that is representative of the European Union. This approach ensures geographical alignment with the scope of our investigation. Secondly, by choosing indices focused on large capitalization, we align our market benchmark with the large capitalization characteristic of our dataset. This ensures that the market return is representative to the size characteristic of our dataset. The data was gathered from Yahoo Finance for the period October 1, 2022, to September 30, 2023. These indices collectively reflect the performance of stocks traded across various European exchanges centred within the European Union. A comprehensive list of these indices, along with their detailed descriptions, can be found in Appendix C, Table 1.

Given the intrinsic properties of financial time-series data, as discussed in "Heteroskedastic Time Series with a Unit Root" by Cavaliere and Robert Taylor (2009), and following the methodology of Eisfeldt et al. (2023), we computed confidence intervals using Newey-West standard errors with five lags. This approach directly tackles the issues of autocorrelation and heteroskedasticity in our dataset, thereby addressing two of the underlying assumptions of linear regression models (Poole and O'farrell, 1971). Furthermore, our study addressed other underlying assumptions for linear modeling by assessing the normality of error terms and presuming a linear relationship between the independent and dependent variable. Given that our analysis involved univariate regression models, multicollinearity was not seen as a concern.

3.3.3 Periodic ChatGPT Impact Analysis

Period Name	Time Period
Pre-ChatGPT period	October 1, 2022 - November 29, 2022
ChatGPT-3 release period	November 30, 2022 - December 14, 2022
ChatGPT-3 extended release period	November 30, 2022 - December 30, 2022
Interim period	December 31, 2022 - March 13, 2023
ChatGPT-4 release period	March 14, 2023 - March 28, 2023
ChatGPT-4 extended release period	March 14, 2023 - April 14, 2023
Post-ChatGPT-4 period	April 15, 2023 - September 30, 2023

The regression models were employed across seven distinct time intervals with specifications for each period provided in the table above. The initial interval, referred to as the *Pre-ChatGPT period*, covers the time before ChatGPT's public release. This period serves as a baseline providing a valid comparison for subsequent periods. In recognition of limitations identified in prior research, our analysis adopts an extended validation period (Eisfeldt et al., 2023). This approach is designed to encompass a broader range of variability and trends, thereby enhancing the robustness and accuracy of the exposure variable. The following period, referred to as the *ChatGPT-3 release period* aligns with the release period, as suggested by Eisfeldt et al. (2023). This provides comparability between the markets. To encompass possible informational frictions between the regions, our release period is extended. This also increases the number of observations, potentially increasing the validity of our results. The third observational window, referenced as the *Interim period*, provides insight into the evolving impact of ChatGPT-3 over time. A similar period structure is repeated for the release of ChatGPT-4. To our knowledge, the examination of the ChatGPT-4 in the valuation context represents a novel contribution to academic literature.

4 Results

The following chapter presents the empirical results of our study, focusing on the impact of generative AI exposure on firm value in the European large-cap public market. We first present the findings from constructing the exposure score, including comparisons between industries and occupational groups. Thereafter, we provide validation results for the exposure measure, examining correlations with key financial metrics and linkage to prior findings Eisfeldt et al. (2023). This section also includes results from robustness testing of the validation results. The chapter culminates with a section presenting the results from our portfolio analysis, particularly highlighting the performance dynamics surrounding the release of ChatGPT-3 and ChatGPT-4. The section additionally includes tests for robustness to ensure result validity.

4.1 Generative AI Score Findings

The findings presented in Table 3 reveal that among the NAICS industries, the *Finance and Insurance* and *Real Estate and Rental and Leasing* industries obtained the highest generative AI exposure scores, at 32.14 and 28.29 respectively. The lowest scores were given to the *Transportation and Warehousing* and *Wholesale Trade* industry with a score of 22.51 and 22.42 respectively. Comparing our results with those of Eisfeldt et al. (2023), we find that *Finance and Insurance* is indeed the industry with the highest exposure, while *Transportation and Warehousing* is among the less exposed industries, signifying a general alignment in results. However, it is noteworthy full alignment is not seen. Moreover, variation of industry exposure differs between the two studies, with the U.S. paper showing larger discrepancies.

Moving to occupational groups it can be noted that the variation of exposure differs greatly between the different groups. *Finance* and *Marketing* obtained the highest generative AI exposure score, 48.84 and 39.88 respectively (see Table 4). The lowest exposure groups were *Operations* and *Engineer*, with the scores 10.66 and 12.43 respectively.

4.2 Generative AI Score Validation

The comparative analysis of results, both with and without adjustment for influential points, was conducted employing Cook's Distance, as elaborated in the methodology chapter (3.2). This comparative assessment revealed no significant variances across the datasets, thus underscoring the robustness of our findings. For simplification purposes, the results presented below are those adjusted for influential points, as provided in Table 6. Furthermore, concerns regarding multicollinearity within the fixed effects models reveal no concern, as evidenced by the Variance Inflation Factor (VIF) values not exceeding the threshold of 10 (Table 7).

ROA: The univariate regression (see Table 6a) supports a statistically significant negative relationship between Return on Assets (ROA) and the generative AI exposure score at the 0.1% significance level. The coefficient for the generative AI exposure score is estimated to -0.4433, indicating that, on average, a one-unit increase in the AI exposure score is associated with a 0.4433 percentage point decrease in ROA. However, adjusting for industry variation using fixed effects, the exposure score is no longer significant at a p-value of 0.11 (see Table 6b).

Tangibility: The finding from the univariate regression (Table 6c) supports a statistically significant negative relationship between tangibility and generative AI exposure at the 0.1% significance level. The estimated coefficient of approximately -0.0124 suggests that for every one-unit increase in the generative AI exposure score, tangibility decreases by 0.0124 units. In contrast, the second regression (Table 6d) does not provide statistical support for this relationship, as evidenced by a non-significant p-value of 0.53.

Labor Intensity: Firstly, looking at the simple linear regression (Table 6e), the model supports that labor intensity is negatively correlated with the exposure score variable at the 99.9% confidence interval. The coefficient of the exposure score is approximately -0.0598 suggesting that an increase in generative AI exposure score by one unit decreases the labor intensity of a firm by 0.0598 units.

Similar to the first regression for labor intensity, the industry adjusted model (Table 6f) supports a negative relationship between AI exposure and labor intensity.

Furthermore, a consistent feature across all univariate linear regression models presented in this section is the low R-squared values. These low values indicate that the generative AI exposure score accounts for only a minor fraction of the variability in the dependent variables.

4.3 Generative AI and Firm Value

Results, both with and without adjustments for outliers, are presented in Tables 10 and 11 respectively. Without adjusting for outliers, Model A, representing an intercept-only model indicates notable excess returns for both the Q1 and Q5 portfolios prior to the ChatGPT release, at 0.37% and 0.34% respectively (Table 10). The excess return for Q1 is significant at the 1% level and for Q5 at the 5% level. However, the AMH portfolio's marginal change of 0.03% suggests parallel trends for these portfolios, reflecting no significant disparity related to AI exposure. During the ChatGPT-3 release period, a positive daily excess return of 0.09% is observed for the AMH portfolio, significant at the 10% level. A similar trend is noted for the AMH portfolio for the extended release period with a positive excess return of 0.09% at the 5% significance level. However, in the subsequent interim period, this effect is no longer significant. Moreover, we find no support for an impact on valuation post the release of ChatGPT-4. Upon examining the results from our market adjusted model (Model B), we observe no significant movements for the pre-ChatGPT period. In contrast, the AMH portfolio exhibited a significant daily excess return of 0.12% during the ChatGPT-3 release, highlighting a disparity in performance between the highly exposed Q1 and less exposed Q5 portfolios at the 5% significance level. Similarly to Model A, we also note a significant excess return for the extended release period.

After adjusting for outliers, our analysis yields consistent results across both models, as provided in Table 11. We observe a significant daily excess return of 0.19% for the AMH portfolio during

the ChatGPT-3 release period at the 0.1% significance level. The extended release period also finds support for this, with Model A indicating a 0.11% increase and Model B a 0.12% increase in daily excess return for the AMH portfolio, both significant at the 1% level. However, the interim period shows no significant changes, similar to the findings before outlier adjustments. Contrasting with the non-adjusted return data, the outlier adjusted analysis reveals a significant positive daily excess return for the AMH portfolio in both the shorter and extended ChatGPT-4 release periods. Specifically, both models report a 0.21% excess daily return during the shorter release period, significant at the 10% level. During the extended release period, however, the models diverge slightly: Model A shows a 0.14% daily excess return, while Model B shows 0.18%, with both significant at the 5% level. Similar to the interim period, no significant effects are observed following the extended release period.

To conclude the findings above, we find support that the ChatGPT-3 release had a significant positive impact on firms with greater exposure to generative AI in relation to those with lower exposure. The effect, though temporary for the ChatGPT-3 release periods, is evident during both the two week release period and the extended one-month release period. Moreover, these findings are consistently supported by both models (A and B), both before and after adjusting for outliers. This indicates robustness of the results. In contrast, the impact of generative AI exposure on firm value during and after the release of ChatGPT-4 presents mixed evidence. After adjusting for outliers, our analysis indicates a significantly positive effect on the valuation of highly exposed firms compared to less exposed ones. However, the robustness of these findings is not as pronounced due to contrasting results before adjusting for outliers. Therefore, while our results after outlier adjustments indicate a significant impact on firm value after the release of ChatGPT-4, they should be interpreted with caution.

Moreover, upon evaluating the normality of error terms in each regression we find no apparent issues. This finding, coupled with the previously addressed assumptions detailed in the methodology

chapter, confirms consistency with the underlying assumptions for linear regression, as outlined in the work of Poole and O'farrell (1971). This means that there are no apparent deviations that might compromise the validity of our analysis.

5 Analysis & Discussion

The following chapter will focus on interpreting our findings by analyzing and discussing the observed results. We will examine the implications of these results and look at how they relate to our research question and theoretical framework. Additionally, we will consider the limitations of our methodology and future research opportunities.

It is important to clarify that these interpretations are not asserted as the sole or definitive explanations for the observed outcomes. Rather, our discussion is centered around the rational application of our theoretical framework, aiming to provide a thoughtful, albeit not exhaustive, analysis of the findings.

5.1 Analysis

5.1.1 Generative AI Score Findings

The data presented in Table 3 and 4, provides a snapshot of the current landscape of generative AI exposure across various occupations and industry sectors. When interpreting these observations it is important to recognize that these are early observations based on current GPT capabilities (ChatGPT-4 as of June 13th 2023) and not final truths.

As highlighted in the results chapter, the *Finance and Insurance* sector emerges as the most exposed industry. According to our theoretical framework, this could imply that tasks in this sector align more closely with programming and writing skills, as opposed to other industries (Eloundou et al., 2023). As discussed in the theoretical framework, chapter, tasks involving these skills have been identified as more receptive to AI augmentation and automation. Conversely, sectors such as *Agriculture, Forestry, Fishing and Hunting*, *Transportation and Warehousing*, and *Wholesale Trade* show the lowest exposure. This result suggests that these sectors predominantly involve a larger number of tasks where performance can't be leveraged using generative AI. From our theoretical framework, this may be due to a higher proportion of manual labor, which is currently less impacted

by AI advancements according to Eloundou et al. (2023). Our findings generally align with those of Eisfeldt et al. (2023), reinforcing the validity of these exposure assessments. However, it is important to acknowledge some slight deviations. These could potentially be attributed to differences in the datasets, as the study by Eisfeldt et al. (2023) focuses on U.S. data, while our analysis is based on European data.

Within the occupational groups, *Finance* and *Marketing* are classified to have the highest exposure to generative AI. In line with the earlier industry-level discussion, it is anticipated that these occupational categories are constructed of occupations encompassing a larger share of tasks requiring writing and programming skills. Consequently, it can be expected that these occupational groups will benefit the most from generative AI capabilities as a greater proportion of the tasks are considered within the technological frontier (Eloundou et al., 2023; Dell' et al., 2023). This suggests that productivity can be increased while retaining a similar quality level. In contrast, the occupational groups *Operations* and *Engineering* received the lowest scores. This trend underscores the inherent limitations in applying generative AI within these fields, primarily ascribed to the task characteristics. Occupations in these categories can be expected to demand a greater degree of critical thinking, which is an area where current generative AI systems may not offer substantial leverage. As observed in the study by Dell' et al. (2023), there is a risk of decreased quality in outputs if tasks exceed the capabilities of generative AI. Additionally, these findings highlight a notable variation in generative AI exposure across different occupational groups, implying that the opportunities for AI-driven augmentation and automation vary significantly among them.

5.1.2 Generative AI Score Validation

The analysis for validating our exposure measure reveals that two out of three ratios align with the previous literature by Eisfeldt et al. (2023). We observe similar relationships in the ratio analysis on ROA and Tangibility (figure 4a, 4b). However, our ratio analysis on Labor Intensity indicates a deviation from prior findings (figure 4c). There are many reasons why this might be the case.

However, notable is the consistently low R^2 value for all models. Such a pattern hints at the potential presence of an *omitted variable bias*. This means that other relevant factors, not included in our models, could significantly influence the outcomes. Consequently, excluding important variables may lead to biased and unreliable estimates of the impact of the generative AI exposure score. This may potentially be an explanation for why our exposure characteristics differ slightly from those of Eisfeldt et al. (2023). However, these findings point towards a general alignment in exposure score characteristics which we deem sufficient for further analysis. However, it is crucial to carefully consider and seek additional validation signs to ensure the robustness of our variable. The alignment in industry exposure, as discussed in the previous section, could serve as such an indicator.

5.1.3 Firm Value

The following section is directed towards the explicit goal of addressing the research question: *What are the effects of recent advances in generative AI on the value of firms within the European Union?*

Firstly, it is important to note that the exposure score shows no effect on valuation during the pre-ChatGPT period, yet it demonstrates significant impact during periods where an effect is anticipated Eisfeldt et al. (2023). This finding, combined with prior validations (see 5.1.2), reinforces that our approach successfully isolates the influence of generative AI from other variables affecting market dynamics.

Our empirical findings indicate a significant impact on firm value for highly exposed firms after the release of ChatGPT-3 (Table 11). Firms which obtained a higher degree of exposure towards generative AI experienced a daily excess return of 0.19% in relation to lower exposed firms during the release period of ChatGPT-3. This finding supports a 2.11% valuation premium for the 11 trading days as opposed to the 4.49% found by Eisfeldt et al. (2023) on the U.S. market. These results were obtained by compounding the excess return across the observed period. For the extended period, consisting of 22 trading days, the total valuation adjustment amounts to 2.56%. This indicates a

value premium for firms that are highly exposed to generative AI compared to their less exposed counterparts. Note that this calculation uses the combined average daily excess return of the AMH portfolio for the extended period, which is 0.115%. This average is obtained from the results of the two models presented in Table 11. Additionally, we observe that this effect diminishes in the subsequent period, suggesting that market prices have adjusted in response to the new information. This trend could indicate a transition towards a steady state where the initial impacts of the ChatGPT-3 release have been fully absorbed and reflected in the market prices. Adjusting in response to new information is consistent with the *Efficient Market Hypothesis* (EMH), as discussed in the theory chapter. However, our results also indicate a continuation of this effect for an extended time period. Such a continuation, beyond the two week release period, implies that while the market responds to new information, it may not do so instantaneously, suggesting frictions in the market and thus deviating from the concept of efficient markets. However, understanding why this is the case requires further investigation.

Moreover, empirical evidence supports the theory that firms highly exposed to generative AI, see positive valuation adjustments due to anticipated future cash flow increases. This expectation stems from the belief that competition will drive these firms to sooner or later adopt generative AI for tasks where labor efficiency and effectiveness could be seen. Subsequently, this suggestively leads to lower labor costs and hence a positive effect on free cash flows. Notably, firms exhibiting greater exposure to generative AI demonstrate a broader spectrum for potential application, thus experiencing more pronounced effects on valuation. Intriguingly, this premium appears to be driven by mere exposure to generative AI, not the actual implementation.

In our ChatGPT-4 analysis, we find support for a statistically significant positive excess return of 0.21% for the AMH portfolio during the first two weeks following the release. With 11 trading days incorporated, this translates to a 2.33% valuation premium for firms highly exposed to generative AI compared to those with lower exposure during the two-week release period. Similar to the

ChatGPT-3 release, we also observe a statistically significant excess return for the AMH portfolio during the extended period. By compounding the daily excess returns of the AMH portfolio across the extended period, which encompasses a total of 23 trading days, the portfolio's total valuation premium for this period is calculated to 3.75%. Note that 0.16% was used as the average of the two models in Table 11. However, when examining the results from the unadjusted dataset, we do not see these significant results. Hence, caution is advised regarding the robustness of these findings. On one hand, adjusting for outliers is logical as excessive returns not attributable to generative AI can skew the results. On the other hand, not adjusting for outliers might incorrectly modify large values that are, in fact, related to the releases.

In comparing our results with those presented by Eisfeldt et al. (2023), notable discrepancies emerge regarding the valuation premium subsequent to the release of ChatGPT-3. Specifically, our findings indicate that the impact on the European market is approximately 42% lower than that observed on the U.S. market. These findings are in line with the theoretical framework suggesting that US firms may see a more significant increase in valuation compared to firms headquartered within the EU. As discussed in the theory, this may come as a result of the faster adoption rate leading to an anticipated earlier positive impact on cash flows.

5.2 Limitations

While we remain confident that every identifiable effort has been made to ensure accuracy and validity, it is essential to recognize the limitations of this study in order to maintain transparency and integrity. This section aims to outline the challenges and limitations encountered during the study with the goal of providing readers with a comprehensive understanding that can be taken into account for future research.

5.2.1 Exposure Scoring Across Occupational Categories

A limitation of this paper is the access to the occupational distribution of each occupation within a firm. Our research utilizes RevelioLabs trial data, which offers insight into each company's workforce distribution across seven distinct categories: *Sales*, *Engineer*, *Operations*, *Admin*, *Finance*, *Marketing* and *Scientist*. The data does not provide a detailed description of specific occupations within each company beyond these seven categories. Consequently, the exposure score calculated using occupational data from ISCO-08 for each of these categories may overlook variations within each occupational group as well as firm specific nuances. Additionally, given the limited details in the study conducted by Eisfeldt et al. (2023), it is important to acknowledge that differences in the level of occupational detail may influence our results. Moreover, the connection between ISCO-08 occupations and the occupational groups identified by RevelioLabs is confidential. To address this, our analysis employs a simplified approach using GPT classification, as outlined in the methodology chapter.

5.2.2 Large-Cap Companies

The study specifically concentrates on large-capitalization companies. This is to ensure data availability. However, by concentrating on this specific segment of the market, our research does not capture the impact of generative AI exposure across all company sizes and geographies. Smaller sized companies, which may exhibit different market dynamics and a varying degree of responsiveness to AI technologies, are not captured.

5.2.3 Previous Literature

As our study focuses on the advancement of generative AI during the last year, it is important to note that there is still limited research done within the field. This implies that there is currently a scarcity of academic publications and challenges in peer-reviewing. This limitation may result in a reliance on working papers that may not have undergone the same careful examination as peer reviewed papers. However, this approach ensures that the study is based on the most current information.

5.3 Future Research

As outlined by previous papers, the field of generative AI and its impact on organizations is still relatively uncharted. Consequently, there exists a multitude of potential research areas. Our study specifically addresses the effects of generative AI on the value of large-cap firms. However, expanding this across various sizes is a potential area of expansion. As Ångström et al. (2023) suggests, companies of different sizes face unique challenges when adopting AI. Therefore, investigating these effects on small and medium-sized enterprises, alongside large corporations, would be an interesting field to study further. Such research would not only offer a more holistic view of AI implementation challenges and effectiveness but also provide critical insights into the scalability and adaptability of generative AI in diverse business contexts.

While our study focused on the cross-industry effects of generative AI on firm value, there is potential for further research in examining the impact within specific industries. Such an investigation would add depth to the field, revealing how different industries might experience more pronounced effects than others promising a more nuanced understanding of generative AI's diverse impacts.

Lastly, given that generative AI is a relatively recent addition to the open market, there lies a significant opportunity to extend the research timeline and examine its long-term effects. Doing this would give a more thorough understanding of the temporal dynamics associated with public access to generative AI over an extended period.

5.4 Final conclusions

In the study, we examine how exposure to generative AI impacts European firm valuation following the introductions of ChatGPT-3 and 4. This is done by utilizing an existing methodological framework developed by Eisfeldt et al. (2023) and answering the following research question: *What are the effects of recent advances in generative AI on the value of firms within the European Union?*

Our findings indicate that firms with higher exposure to generative AI experienced a statistically significant positive excess daily return of 0.19% following the release of ChatGPT-3, relative to less exposed firms. This effect was observed during the initial two weeks, culminating in a total valuation adjustment of 2.11%. When compared with similar studies in the U.S., our analysis suggests a 42% lower impact on firm valuation for the same period. Additionally, our research identifies an extended valuation effect. Over the extended period, this totaled a valuation premium of 2.56% for higher exposed firms. Additionally, our data support similar effects on firm valuation following the release of ChatGPT-4, with firms highly exposed to generative AI experiencing an average daily excess return of 0.21%, translating to a 2.33% valuation premium over the first two weeks post-release. Looking at the valuation effect for the whole month following the release of ChatGPT-4 it can be concluded that the premium amounts to 3.75% for highly exposed firms in relation to lower exposed firms. However, due to certain limitations in data robustness, results tied to the release of ChatGPT-4 should be interpreted with caution. These findings support that the introduction of ChatGPT-4 had a more profound impact on European firm valuation in comparison with ChatGPT-3 when comparing the release periods.

References

Andrew Ng. Andrew Ng: Opportunities in AI - 2023. *Stanford Online*, (Accessed 25/11/2023), 9 2023. URL https://www.youtube.com/watch?v=5p248yoa3oE&t=238s.

Rebecka C. Ångström, Michael Björn, Linus Dahlander, Magnus Mähring, and Martin W. Wallin. Getting AI Implementation Right: Insights from a Global Survey. *California Management Review*, 11 2023.

Erik Brynjolfsson, Danielle Li, and Lindsey R Raymond. Generative AI at Work. *NBER Working paper 31161*, 4 2023.

Giuseppe Cavaliere and A M Robert Taylor. Heteroskedastic Time Series with a Unit Root. *Econometric Theory*, 25(5):1228–1276, 10 2009.

David T. Larrabee and Jason A. Voss. Valuation Techniques : Discounted Cash Flow, Earnings Quality, Measures of Value Added, and Real Options. 11 2012.

Fabrizio Dell', Acqua Saran, Rajendran Edward Mcfowland, Lisa Krayer, Ethan Mollick, François Candelon, Hila Lifshitz-Assaf, Karim R Lakhani, and Katherine C Kellogg. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *Harvard Business School, Working Paper 24-013*, 9 2023.

Dr. Gwendolyn Stripling. Introduction to Generative AI. *Google*, (Accessed: 27/11/2023), 5 2023. URL https://www.cloudskillsboost.google/course_templates/536.

Andrea L Eisfeldt, Gregor Schubert, and Miao Ben Zhang. GENERATIVE AI AND FIRM VALUES. *NBER Working Paper 31222*, 5 2023.

Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. *OpenAI, OpenResearch, University of Pennsylvania*, 8 2023.

European Commission. COMMISSION RECOMMENDATION of 29 October 2009 on the use of the International Standard Classification of Occupations (ISCO-08). *European Commission*, 10 2009.

European Parlament. EU AI Act: first regulation on artificial intelligence. *European Parlament*, (Accessed 04/12-2023), 6 2023. URL https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

Mia Hoffmann and Laura Nurski. What is holding back artificial intelligence adoption in Europe? *Bruegel*, pages 11–17, 11 2021.

Mariya N. Ivanova, Henrik Nilsson, and Milda Tylaite. Yesterday is history, tomorrow is a mystery: Directors' and CEOs' prior bankruptcy experiences and the financial risk of their current firms. *Journal of Business Finance and Accounting*, 4 2023.

Martin Lally. Regulation and the Term of the Risk Free Rate: Implications of Corporate Debt. *Accounting Research Journal*, 20 NO 2:73–80, 2007.

Michael E. Porter. What Is Strategy? *Harvard Business Review*, 12 1996.

Mirella Lapata. What is generative AI and how does it work? – The Turing Lectures with Mirella Lapata. *The Royal Institution*, (Accessed 25/11/2023), 10 2023. URL https://www.youtube.com/watch?v=_6R7Ym6Vy_I&t=261s.

Shakked Noy and Mit Whitney Zhang. Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *MIT*, 3 2023.

OpenAI. Research GPT-4. *OpanAI*, (Accessed 01/12/2023), 3 2023. URL https://openai.com/research/gpt-4.

Peter Kennedy. *A Guide to Econometrics, 6th-edition*. 2008.

Michael A Poole and Patrick N O'farrell. The Assumptions of the Linear Regression Model. *Transactions of the Institute of British Geographers*, (52):145–158, 1971.

William F. Sharpe. CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK. *The Journal of Finance*, 19(3):425–442, 1964.

Manfred Te Grotenhuis Tom Van der Meer and Ben Pelzer. Influential Cases in Multilevel Modeling: A Methodological Comment. *Amerikan Sociological Review*, 4 2006.

A Appendix: Methodology Notes

A.1 Set Up for Integrating GPT-4 to our DataSets

In order to integrate ChatGPT-4 into our dataset, two models were developed in Python using an API-key provided by OpenAI. Both models are constructed in the same way, but uses different system prompts as the models serve different purposes. The system prompt provides specific instructions, establishing a framework for the model to operate within. Furthermore the models function at a temperature setting of 0, which in the context of LLM-models regulates the level of randomness or creativity of the output from the model. A temperature of 0 will ensure that its responses are highly deterministic, producing the most likely output with minimal randomness. Finally, to ensure the accuracy level of both these models, we have done three different runs and compared the results. The percentage agreement for task classifications can be found in Appendix C, Table 8a, and the percentage agreement for occupational categorization between ISCO-08 and Revleiolabs can be found in Appendix C, Table 8b.

A.1.1 System Prompt for Task Classification

Below is the system prompt used for classifying tasks each individual task into one of the following categories, *C0, C1, C2* and *C3*, The prompt used was constructed under an established framework created by Eloundou et al. (2023), and is heavily based on the prompt used in the study by Eisfeldt et al. (2023) with some minor modifications to fit our data. The user prompt includes the task provided by ISCO-08.

System Prompt: "Consider the most powerful OpenAI large language model (LLM). The LLM is capable of processing text-based inputs and outputs but has limitations, including no physical interaction and limited access to recent information (those from <1 year ago). You will be given a task. Evaluate the task and state a label (*C0/C1/C2/C3*) for the task based on the following criteria and provide a short explaination:

C0 - No exposure: Label tasks E0 if none of the above clearly decrease the time it takes for an experienced worker to complete the task with high quality by at least half. Some examples: - If a task requires a high degree of human interaction (for example, in-person demonstrations) then it should be classified as E0. - If a task requires precise measurements then it should be classified as E0. - If a task requires reviewing visuals in detail then it should be classified as E0. - If a task requires any use of a hand or walking then it should be classified as E0. - Tools built on top of the LLM cannot make any decisions that might impact human livelihood (e.g.hiring, grading, etc.). If any part of the task involves collecting inputs to make a final decision (as opposed to analyzing data to inform a decision or make a recommendation) then it should be classified as E0. The LLM can make recommendations. - Even if tools built on top of the LLM can do a task, if using those tools would not save an experienced worker significant time completing the task, then it should be classified as C0. - The LLM and systems built on top of it cannot do anything that legally requires a human to perform the task. - If there is existing technology not powered by an LLM that is commonly used and can complete the task then you should mark the task E0 if using an LLM or LLM-powered tool will not further reduce the time to complete the task. When in doubt, you should default to C0."

C1 - Direct exposure: Label tasks C1 if direct access to the LLM through an interface like ChatGPT or the OpenAI playground alone can reduce the time it takes to complete the task with equivalent quality by at least half. This includes tasks that can be reduced to: - Writing and transforming text and code according to complex instructions, - Providing edits to existing text or code following specifications, - Writing code that can help perform a task that used to be done by hand, - Translating text between languages, - Summarizing medium-length documents, - Providing feedback on documents, - Answering questions about a document, - Generating questions a user might want to ask about a document, - Writing questions for an interview or assessment, - Writing and responding to emails, including ones that involve refuting information or engaging in a negotiation (but only if the negotiation is via written correspondence), - Maintain records of written data, - Prepare training materials based on general knowledge, or - Inform anyone of any information via any written or

spoken medium

C2 - Exposure by LLM-powered applications: Label tasks C2 if having access to the LLM alone may not reduce the time it takes to complete the task by at least half, but it is easy to imagine additional software that could be developed on top of the LLM that would reduce the time it takes to complete the task by half. This software may include capabilities such as: - Summarizing documents longer than 6000 tokens and answering questions about those documents, - Retrieving up-to-date facts from the Internet and using those facts in combination with the LLM capabilities, - Searching over an organization's existing knowledge, data, or documents and retreiving information, - Retrieving highly specialized domain knowledge, - Make recommendations given data or written input, - Analyze written information to inform decisions, - Prepare training materials based on highly specialized knowledge, - Provide counsel on issues, and - Maintain complex databases.

C3 - Exposure given image capabilities: Suppose you had access to both the LLM and a system that could view, caption, and create images as well as any systems powered by the LLM (those in E2 above). This system cannot take video as an input and it cannot produce video as an output. This system cannot accurately retrieve very detailed information from image inputs, such as measurements of dimensions within an image. Label tasks as C3 if there is a significant reduction in the time it takes to complete the task given access to a LLM and these image capabilities: - Reading text from PDFs, - Scanning images, or - Creating or editing digital images according to instructions. The images can be realistic but they should not be detailed. The model can identify objects in the image but not relationships between those options.

A.1.2 GPT Prompt for Occupational Categorization

Below is the system prompt used for categorizing each occupation from ISCO-08 provided by the ILO into one of the seven categorize provided by RevelioLabs. The system prompt lays out a structured framework, providing specific instructions and examples on how categorizing occupations

and acts as a guideline for the model to interpret and process the input accurately. It is important to note that these examples are manually assembled, using inspiration from a sample of the skill dynamic breakdown of each category provided by RevelioLabs. The user prompt includes the occupation title and the associated tasks from the ISCO-08 list.

System Prompt: "You will receive a series of occupation titles along with associated tasks. Your job is to analyze each occupation and the corresponding tasks carefully. Your task is to analyze the tasks linked to each occupation and categorize the entire occupation into the most fitting of the following categories: Sales, Engineer, Admin, Operations, Scientist, Marketing, or Finance.

For precice categorization, consider these examples for each category:

Sales: Engaging in business development activities such as identifying new market opportunities or negotiating contracts with potential clients. Performing customer needs assessments and presenting products or services as solutions. Generating leads and converting them into customers through various sales techniques. Establishing and maintaining relationships with key clients to ensure repeat business. Creating and delivering sales presentations and proposals to highlight product benefits.

Engineer: Developing new product designs, conducting prototype testing, and iterating based on feedback. Optimizing manufacturing processes for improved efficiency and cost savings. Applying engineering principles to solve complex technical problems and improve product functionality. Conducting research to develop new engineering methods and technologies. Collaborating with cross-functional teams to integrate engineering solutions into final product designs.

Admin: Managing office supplies, coordinating schedules, and maintaining organizational systems. Handling correspondence, preparing reports, and providing general administrative support. Organizing company records, managing databases, and ensuring data privacy and security. Scheduling and

supporting company meetings, events, and travel arrangements. Performing reception duties such as greeting visitors and answering phones.

Operations: Overseeing supply chain logistics, managing inventory levels, and ensuring smooth operation of production lines. Implementing process improvement initiatives to enhance operational efficiency. Developing and managing quality assurance programs to maintain company standards. Planning and controlling the allocation of resources and the maintenance of equipment. Analyzing operational data to forecast operational needs and project future requirements.

Scientist: Conducting experimental research, publishing findings, and developing new scientific methodologies. Analyzing complex datasets to extract insights and inform further research directions. Designing experiments and trials to test hypotheses and explore scientific questions. Collaborating with academic and industrial partners to advance scientific knowledge. Staying updated on scientific advancements and integrating new findings into research activities.

Marketing: Crafting marketing strategies, overseeing advertising campaigns, and evaluating market research. Developing brand guidelines and coordinating promotional activities. Analyzing market trends and customer feedback to inform marketing decisions. Utilizing digital marketing tools and social media to engage with a broader audience. Measuring and reporting on the performance of marketing campaigns and activities.

Finance: Managing budgets, forecasting financial trends, and conducting risk analysis. Overseeing investment portfolios, performing audits, and ensuring compliance with financial regulations. Analyzing financial statements to advise on business decisions and performance improvement. Developing financial models and conducting investment appraisals. Ensuring accurate and timely financial reporting and control, as well as managing taxation issues.

For each task, you should:

- Identify the key components and objectives of the tasks included in the occupation.
- Evaluate the occupation title combined with the tasks to match the occupation to the appropriate category based on the identified components.
- Provide a brief rationale for your categorization."

B Appendix Figures

Figure 1: Cumulative Daily Return By Portfolio from 2022-10-01 to 2023-09-30. The plot shows the cumulative return percentage for two portfolios, Q1 and Q5, and their difference (AMH). The portfolios are industry-neutral, and calculated by ranking each firm by exposure score into quintiles within each industry, from Q1 (highest exposure score) to Q5 (lowest exposure score). These quintiles are consolidated across all segments to create quintile portfolios that represent a cross-section of firms from each industry.

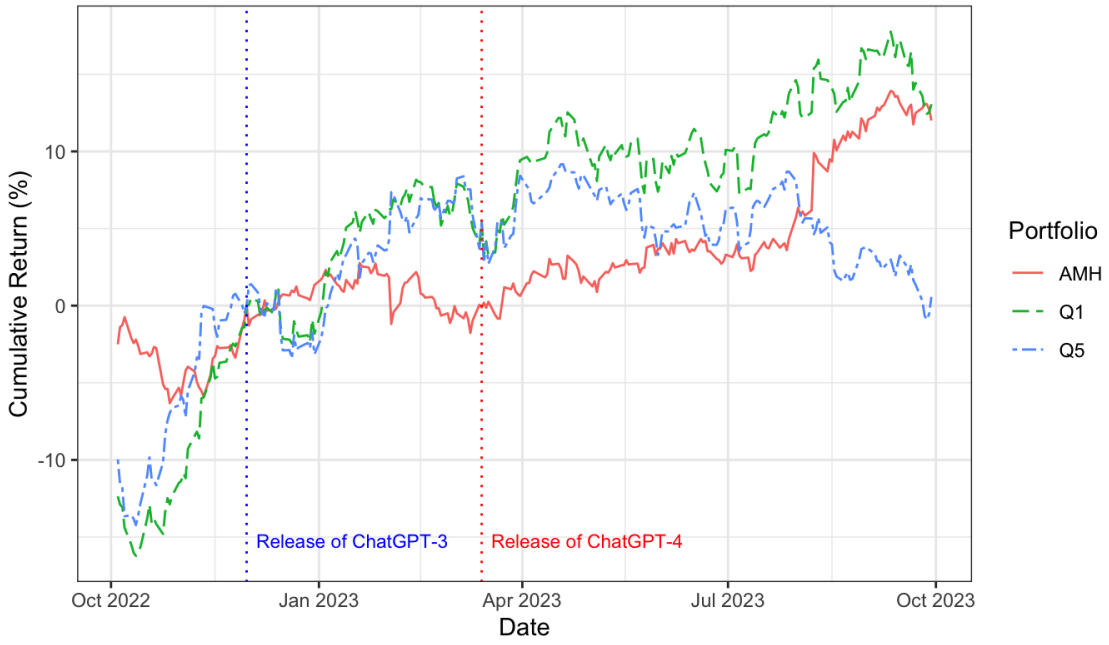


Figure 2: Cumulative Daily Return By Portfolio from 2022-10-01 to 2023-09-30 adjusted for outliers. The plot shows the cumulative return percentage for two portfolios, Q1 and Q5, and their difference (AMH) adjusted for outliers. The portfolios are industry-neutral, and calculated by ranking each firm by exposure score into quintiles within each industry, from Q1 (highest exposure score) to Q5 (lowest exposure score). These quintiles are consolidated across all segments to create quintile portfolios that represent a cross-section of firms from each industry.

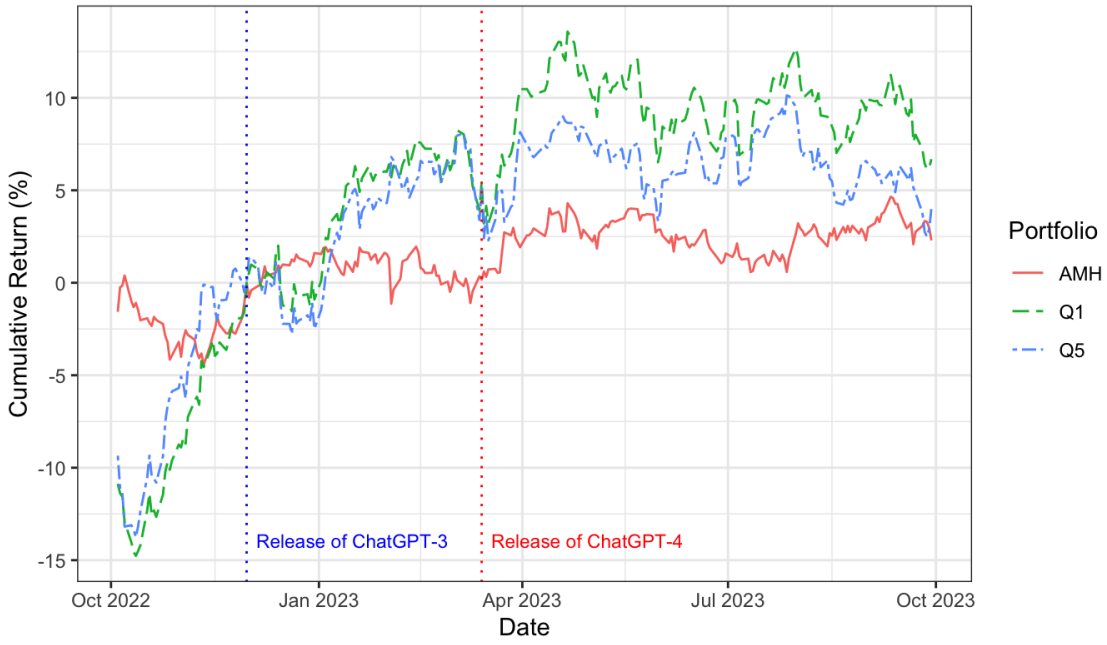


Figure 3: Outlier Adjustment of Daily Returns. This graph illustrates the removal of outliers from our dataset. Outliers are identified independently on a monthly basis using a threshold based on the Z-score method for the 12 months in the sample. Any data point lying more than 3.5 standard deviations from the mean of the particular month's sample is considered an outlier. These outliers are represented as red points on the graph and have been adjusted to a value of 0 to mitigate their impact on the analysis. A total of 1,019 values were adjusted. This means that the percentage of unaffected values amounts to 99.19% after making the adjustments.

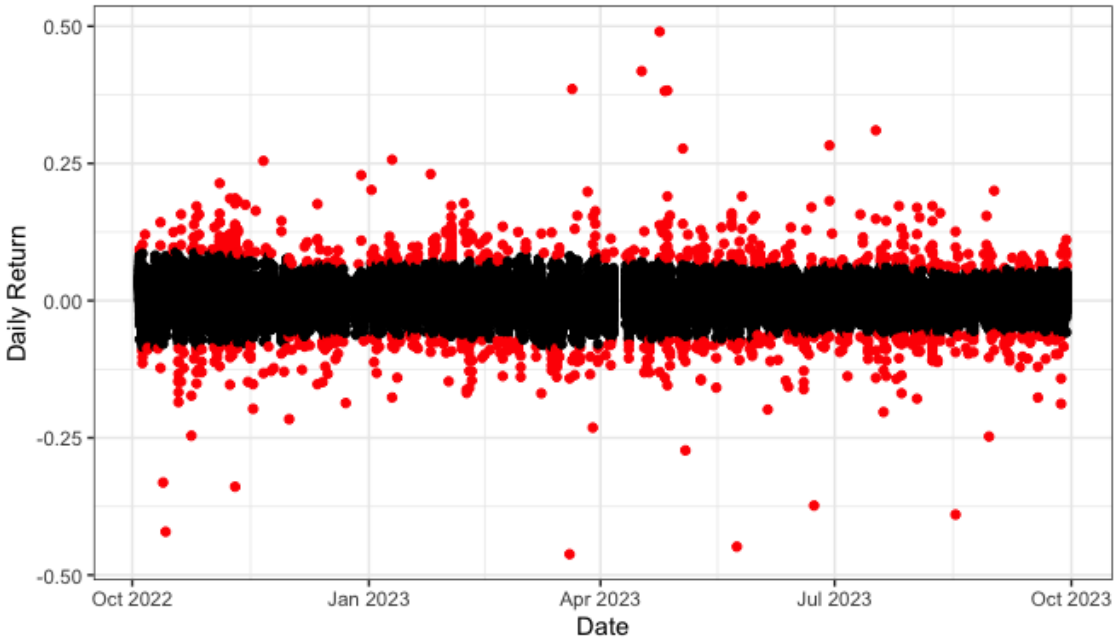
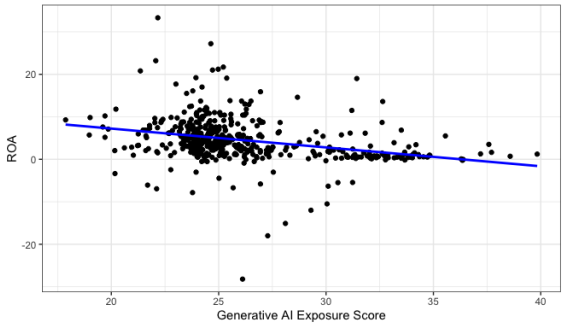
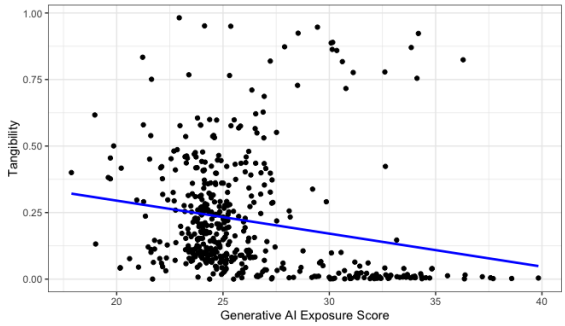


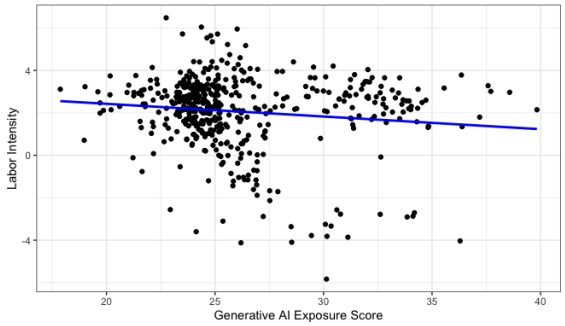
Figure 4: Generative AI Exposure and firm characteristics. The graphs below shows the relationship between generative AI Exposure Scores and various firm characteristics: Return on Assets (ROA), Labor Intensity and Tangibility. Due to missing data for 15 companies, the analysis includes 483 observations, providing a substantial dataset for assessing these relationships.



(a) Generative AI Exposure Score In Relation to ROA



(b) Generative AI Exposure Score in Relation to Tangibility



(c) Generative AI Exposure Score in Relation to Labor Intensity

C Appendix Tables

Table 1: List of European Stock Market Indices Used for Market Rate Calculation. The market rate used in this study was determined by computing the average daily return of the specified indices below. These indices were selected through two key criteria. Firstly, to avoid the generality of a broad European index, our selection targets a market return that is representative of the European Union. This approach ensures geographical alignment with the scope of our investigation. Secondly, by choosing indices focused on large capitalization, we align our market benchmark with the large capitalization characteristic of our dataset. This ensures that the market return is representative to the size characteristic of our dataset. The dataset was retrieved from Yahoo Finance for the period between 2022-10-01 to 2023-09-30.

Exchange	Index
BIT	FTSE MIB
ENXTAM	AEX Index
WBAG	ATX
ENXTPA	CAC 40
ENXTBR	BEL 20
XTRA	DAX
BME	IBEX 35
ISE	ISEQ Overall Index
CPSE	OMXC25
OM	OMX Stockholm 30
HLSE	OMX Helsinki 25
ENXTLS	PSI 20

Table 2: Occupation Categories by ISCO-08 Code. This table presents a randomized sample of occupations classified by the GPT model specified in A. Occupations are assigned to one of seven predefined categories, *Admin, Sales, Operations, Scientist, Engineer, Marketing* and *Finance*. Detailed task descriptions and the rationale behind each categorization are excluded due to space constraints.

ISCO-08	Occupation Title	Category
1112	Senior Government Officials	Admin
1420	Retail and Wholesale Trade Managers	Sales
3119	Physical and Engineering Science Technicians Not Elsewhere Classified	Engineer
3121	Mining Supervisors	Operations
3211	Medical Imaging and Therapeutic Equipment Technicians	Operations
3212	Medical and Pathology Laboratory Technicians	Scientist
3354	Government Licensing Officials	Admin
3522	Telecommunications Engineering Technicians	Engineer
8332	Heavy Truck and Lorry Drivers	Operations
9129	Other Cleaning Workers	Operations

Table 3: Generative AI Exposure Score by Industry. The table below describes the average weighted generative AI exposure score by market cap. for each industry.

NAICS Code	Industry Title	Exposure Score
52	Finance and Insurance	32.14
53	Real Estate and Rental and Leasing	28.29
71	Arts, Entertainment and Recreation	26.77
62	Health Care and Social Assistance	26.55
54	Professional, Scientific, and Technical Services	26.28
51	Information	26.28
55	Management of Companies and Enterprises	26.19
22	Utilities	26.10
21	Mining, Quarrying, and Oil and Gas Extraction	25.85
56	Administrative, Support, Waste Management, and Remediation Services	25.59
23	Construction	25.11
31-33	Manufacturing	24.63
81	Other Services (except Public Administration)	24.23
44-45	Rental Trade	23.61
72	Accommodation and Food Services	23.55
11	Agriculture, Forestry, Fishing and Hunting	23.49
48-49	Transportation and Warehousing	22.51
42	Wholesale Trade	22.42

Table 4: Generative AI Exposure Score by Occupational Group
The table presents each occupational group's unique exposure score that is used for mapping exposure score from occupational level to firm level.

Occupational Group	Generative AI Exposure Score
Finance	48.84
Marketing	39.88
Admin	39.01
Scientist	27.39
Sales	18.59
Engineer	12.43
Operations	10.66

Table 5: Generative AI Exposure Score by Company
The table displays the 10 companies with the highest generative AI exposure scores, as well as the 10 companies with the lowest scores.

Company	Generative AI Exposure Score
FinecoBank Banca Fineco S.p.A.	39.83
Banca Mediolanum S.p.A.	38.58
Hannover Rück SE	37.71
Azimut Holding S.p.A.	37.58
Banca Generali S.p.A.	37.20
SCOR SE	36.38
Aegon Ltd.	36.35
IMMOFINANZ AG	36.29
DWS Group GmbH & Co. KGaA	35.56
...	...
Industria de Diseño Textil, S.A.	20.21
Just Eat Takeaway.com N.V.	20.17
Compañía de Distribución Integral Logista Holdings, S.A.	20.16
Axfood AB	19.86
Jumbo S.A.	19.70
Koninklijke Ahold Delhaize N.V.	19.70
Deutsche Post AG	19.60
DSV A/S	19.01
Ryanair Holdings plc	18.97
B&M European Value Retail S.A.	17.87

Table 6: Consolidated Regression Results of Generative AI Exposure and firm characteristics. The following results describes the relationship between generative AI exposure scores and various firm characteristics: Return on Assets (ROA), Labor Intensity, Tangibility. The tables include both results before and after industry adjustments. Due to missing data for 15 companies, the analysis includes 483 observations, providing a substantial dataset for assessing these relationships.

(a) Regression results of ROA

Coefficient	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	16.0890	1.6327	9.91	0.00 ****
Eo	-0.4433	0.0615	-7.20	0.00 ****
Multiple R-squared: 0.10, Adjusted R-squared: 0.10				

Significance codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1

(b) Regression Results of ROA - Industry Adjusted

Coefficient	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	11.4717	3.5501	3.23	0.00 ***
Eo	-0.1541	0.0975	-1.58	0.11
Multiple R-squared: 0.18, Adjusted R-squared: 0.15				

Significance codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1

(c) Regression Results of Tangibility

Coefficient	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.5428	0.0696	7.80	0.00 ****
Eo	-0.0124	0.0026	-4.70	0.00 ****
Multiple R-squared: 0.04, Adjusted R-squared: 0.04				

Significance codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1

(d) Regression Results of Tangibility - Industry Adjusted

Coefficient	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0039	0.1220	0.03	0.97
Eo	0.0021	0.0034	0.63	0.53
Multiple R-squared: 0.45, Adjusted R-squared: 0.43				

Significance codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1

(e) Regression Results of Labor Intensity

Coefficient	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.6209	0.5681	6.37	0.00 ****
Eo	-0.0598	0.0215	-2.78	0.01 ***
Multiple R-squared: 0.02, Adjusted R-squared: 0.01				
Significance codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1				

(f) Regression results of Labor Intensity - Industry Adjust

Coefficient	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.8412	0.9524	7.18	0.00 ****
Eo	-0.1374	0.0262	-5.26	0.00 ****
Multiple R-squared: 0.48, Adjusted R-squared: 0.46				
Significance codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1				

Table 7: Variance Inflation Factors (VIF) for Each Variable in the Data Validation Models. Note that only segments with at least six companies are considered in the model to ensure validity. Furthermore, segment 5 is acting as baseline in the model and in therefore not included even though observations exceed six for the industry.

Variable	VIF
E_o	2.715535
Mining, Quarrying, and Oil and Gas Extraction	1.041198
Utilities	1.099032
Construction	1.023360
Wholesale Trade	1.048351
Retail Trade	1.081510
Transportation and Warehousing	1.049456
Information	1.177087
Finance and Insurance	2.771274
Real Estate and Rental and Leasing	1.189647
Professional, Scientific, and Technical Services	1.070992
Administrative and Support and Waste Management and Remediation Services	1.031759

Table 8: GPT Model Consistency Across Task Classification and Occupation Mapping
The table presents data on the consistency of the task classification process, as well as the occupational mapping using the ChatGPT-4 API in Python. Each row in the table indicates the percentage of agreement (%) between the different runs.

(a) Comparison of three different runs of the GPT model for classifying tasks into the four distinct exposure categories: 'C0', 'C1', 'C2', and 'C3'

Score comparison	Agreement (%)
GPT #1 vs. GPT #2	94.25
GPT #1 vs. GPT #3	93.71
GPT #2 vs. GPT #3	93.62

(b) Comparison of the three different runs of the GPT model used for mapping occupations from ISCO-08 to one of the seven distinct categories defined by RevelioLabs, which includes *marketing, admin, scientist, sales, finance, operations, and engineer.*

Score comparison	Agreement (%)
GPT #1 vs. GPT #2	96.02
GPT #1 vs. GPT #3	96.49
GPT #2 vs. GPT #3	96.72

Table 9: Sample of the exposure categorization of task into C0/C1/C2/C3 by our ChatGPT model

ISCO 08 Code	Occupation	Task	Exposure Category	Explanation
1111	Legislators	Negotiating with other legislators and representatives of interest groups to reconcile interests and create policies	C0	Task requires high degree of human interaction and negotiation.
1113	Traditional Chiefs and Heads of Villages	Performing ceremonial duties related to cultural events	C0	Ceremonial duties require a human presence and cultural understanding.
1114	Senior Officials of Special-interest Organizations	Determining and formulating policies, rules, and regulations	C1	LLM can assist in policy formulation and document drafting.
1213	Policy and Planning Managers	Representing the organization in negotiations, at conventions, seminars, public hearings, and forums	C0	Role involves personal representation and negotiation skills.
2113	Chemists	Preparing scientific papers and reports	E1	LLM can aid in drafting and information gathering for scientific writing.
2132	Farming, Forestry, and Fisheries Advisers	Managing resources for commercial, recreational, and environmental benefits	C0	Involves complex decision-making and fieldwork.
2149	Engineering Professionals Not Elsewhere Classified	Designing medical devices and imaging systems	C0	Involves specialized knowledge and physical prototyping.
2152	Electronics Engineers	Directing maintenance and repair of electronic systems and equipment	C0	Task requires physical manipulation of electronics.
2166	Graphic and Multimedia Designers	Formulating design concepts	C1	LLM can contribute to ideation and information provision for design.
2512	Software Developers	Modifying software for error correction, hardware adaptation, and performance upgrades	C2	LLM can provide coding assistance and performance optimization suggestions.
3314	Statistical, Mathematical and Related Associate Professionals	Preparing data for presentation in graphical or tabular form	C2	LLM can assist with data preparation and visualization.

Table 10: Realized returns of exposure sorted portfolios without outlier adjustments The table bellow reports excess returns for value weighted industry neutral portfolios. T-statistics are calculated using Newey-West standard errors with five lags. These are provided within the parentheses.

Sample	Q1	Q5	AMH
Model A: Excess returns (%)			
Pre-ChatGPT	0.37***	0.34**	0.03
	(2.59)	(2.13)	(0.25)
ChatGPT-3 release period	0.24	0.14	0.09*
	(1.40)	(0.83)	(1.76)
ChatGPT-3 extended release period	-0.01	-0.12	0.09**
	(-0.12)	(-0.81)	(2.07)
Interim Period	0.11	0.13	-0.03
	(0.97)	(1.05)	(-0.44)
ChatGPT-4 release period	0.21	0.09	0.11
	(1.30)	(0.61)	(0.72)
ChatGPT-4 extended release period	0.32***	0.20*	0.10
	(2.66)	(1.76)	(1.48)
Post-ChatGPT-4	0.01	-0.06	0.06
	(1.22)	(-0.02)	(1.49)
Model B: Market factor-adjusted alpha (%)			
Pre-ChatGPT	0.12	-0.00	0.12
	(1.20)	(-0.02)	(1.03)
ChatGPT-3 release period	0.15***	0.02	0.12**
	(2.89)	(1.02)	(2.07)
ChatGPT-3 extended release period	0.06	-0.02	0.08**
	(1.12)	(-0.51)	(2.06)
Interim Period	0.04	0.03	-0.00
	(0.87)	(0.92)	(-0.00)
ChatGPT-4 release period	0.24**	0.13***	0.10
	(2.19)	(2.59)	(0.68)
ChatGPT-4 extended release period	0.21***	0.05	0.14
	(2.77)	(0.66)	(1.58)
Post-ChatGPT-4	0.06	-0.02	0.07*
	(1.76)	(-0.96)	(1.64)

Significance codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1

Table 11: Realized returns of exposure sorted portfolios with outlier adjusted data: The table bellow reports excess returns for value weighted industry neutral portfolios with outlier adjusted data, where any data point lying more than 3.5 standard deviations from the mean is considered an outlier and has been removed. T-statistics are calculated using Newey-West standard errors with five lags. These are provided within the parentheses.

Sample	Q1	Q5	AMH
Model A: Excess returns (%)			
Pre-ChatGPT	0.33**	0.32**	-0.00
	(2.52)	(2.10)	(-0.02)
ChatGPT-3 release period	0.32*	0.15	0.19****
	(1.75)	(0.85)	(3.68)
ChatGPT-3 extended release period	0.03	-0.09	0.11***
	(0.26)	(0.68)	(2.89)
Interim Period	0.09	0.12	-0.03
	(0.86)	(0.96)	(-0.58)
ChatGPT-4 release period	0.31*	0.08	0.21*
	(1.92)	(0.53)	(1.78)
ChatGPT-4 extended release period	0.36***	0.21*	0.14**
	(3.23)	(1.80)	(2.15)
Post-ChatGPT-4	-0.04	-0.03	-0.02
	(-0.68)	(-0.53)	(-0.75)
Model B: Market factor-adjusted alpha (%)			
Pre-ChatGPT	0.07	0.00	0.06
	(0.96)	(0.02)	(0.71)
ChatGPT-3 release period	0.23****	0.03	0.19****
	(4.49)	(1.29)	(3.68)
ChatGPT-3 extended release period	0.11*	-0.00	0.12***
	(1.80)	(-0.03)	(2.69)
Interim Period	0.03	0.02	-0.00
	(0.73)	(0.85)	(-0.04)
ChatGPT-4 release period	0.34****	0.12***	0.21*
	(3.97)	(2.78)	(1.78)
ChatGPT-4 extended release period	0.24***	0.04	0.18**
	(3.04)	(0.82)	(2.37)
Post-ChatGPT-4	0.00	0.01	-0.02
	(0.13)	(0.56)	(-0.60)

Significance codes: '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1