Analyzing How the Adoption of Artificial Intelligence Affects Industry Concentration in U.S. AI-using Industries

by

Xiaotong Tan

An honors thesis submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Arts

Business and Economics Honors Program

NYU Shanghai

May 2024

Professor Marti G. Subrahmanyam Professor Christina Wang Professor Wendy Jin	Professor Robert Seamans
Faculty Advisers	Thesis Adviser

Contents

1 Introduction **4**

2 Literature **8**

3 Data and Measurement **12**

3.1 Employment Profiles from LinkedIn 12

3.2 Measure, Summary Statistics, and Validation 14

4 AI Adoption and Industry Concentration **20**

4.1 Industry-level Results 20

4.2 Firm-level Long-difference Results 23

4.3 Firm-level Difference-in-Difference Results 27

4.3.1 Fixed-effect Results 27

4.3.2 Main Difference-in-Difference Results 28

4.3.3 Event Study Results 32

5 Mechanisms **35**

5.1 Product Creation 35

5.2 Operating Cost Savings 38

5.3 Labor Productivity Enhancement 39

6 Conclusion **40**

Abstract

In the context of rapid advancements in artificial intelligence (AI) and growing concerns regarding industry concentration in the United States, this study investigates the relationship between AI adoption and industry concentration. The primary objective is to offer empirical evidence at the firm level, focusing on a sample of U.S. publicly traded companies within AI-related industries. We measure AI adoption using AI-skilled human capital metrics sourced from LinkedIn resume data. Concurrently, we collect firm performance metrics from Compustat and categorize businesses into industries based on 2-digit North American Industry Classification System codes. Our findings reveal a significant impact of AI adoption on an increase in the sales share of the top 20 firms exclusively, indicating a discernible rise in industry concentration attributed to AI adoption. These results withstand rigorous testing through various econometric methodologies, including long difference models, fixed effect models, and difference-in-difference models utilizing the launch of TensorFlow as a quasi-natural experiment. Our empirical findings align with the theory positing that AI, as a general-purpose technology, drives growth across all industries, particularly benefiting larger firms. The study contributes to the literature primarily by providing novel and direct evidence of the impact of AI adoption on the increase in industry concentration across a wide range of industries.

Key Words: *artificial intelligence, technology adoption, industry concentration, superstar firms, economic growth*

1 Introduction

Artificial intelligence (AI) has experienced explosive growth in recent years. As per "The AI Index 2022 Annual Report" from Stanford University, AI publications doubled from 162,444 in 2010 to 334,497 in 2021. This AI research and development surge has been paralleled by increased investment and adoption. The report also notes that global private investments in AI reached $93.5 billion in 2021, more than double the 2020 total. In addition, McKinsey & Company reports that global AI adoption in 2021 was 2.5 times greater than in 2017, with 50% of survey respondents indicating the adoption of AI in at least one business area. According to a 2018 survey conducted by Deloitte, private sector investments in commercial applications of AI have primarily focused on three key areas: machine learning, natural language processing, and computer vision. AI adoption has extended to various industries, including automotive, healthcare, retail, finance, and manufacturing. Examples include McDonald's acquiring an AI startup in Tel Aviv in March 2019 and Sanofi entering an AI drug discovery deal with Atomwise in August 2022.

This remarkable growth of AI leads to the interest in closer examination of the nature of AI and its application in business and economy. AI is broadly defined as "the capacity of computers or other machines to exhibit or simulate intelligent behavior," according to the Oxford English Dictionary. The U.S. Census Bureau's 2019 Annual Business Survey offers a similar, encompassing definition, characterizing AI as a branch of computer science and engineering dedicated to making machines intelligent. A narrower focus of AI centers on Machine Learning, encompassing well-known subfields such as Deep Learning and Reinforcement Learning, along with popular applications like Recommendation Systems, Computer Vision, and Natural Language Processing. From an economic perspective, AI is often defined as a general purpose technology (GPT), given its ability to be applied in various fields and raise research and development productivity (Agrawal et al., 2018; Crafts, 2021). Such an economic definition of AI raises the question of whether AI exhibits similar economic impacts to other general-purpose technologies, particularly automation and information technology.

On the other hand, another significant contemporary economic trend, parallel to advances in information technology, is the rise in industry concentration among U.S. sectors, dating back to the 1980s. Many studies posit that adopting advanced technologies, including AI, is one potential driver of this ongoing rise in market concentration. Autor et al. (2020) attribute the surge in concentration to the emergence of "superstar firms" and suggest that scale-biased technological change may be a contributing factor. Crouzet and Eberly (2019) and Ridder (2024) also highlight "intangible capital," which encompasses software, information technology, intellectual property, brands, and innovative business processes, as a force driving increased concentration.

Under this context, the primary interest of this thesis is to empirically investigate whether the adoption of AI leads to an increase in industry concentration within U.S. AI-using industries, and if so, what mechanisms are involved. While this question has not been extensively explored in existing literature, recent studies offer some but limited support. Babina et al. (2024) present empirical evidence indicating that AI-powered growth, stemming from product and process innovation, tends to favor larger firms, ultimately resulting in higher industry concentration. Beyond AI, Firooz et al. (2023) identify a similar mechanism in the realm of robot adoption. They suggest that a greater likelihood for automation within more productive and larger firms facilitates their expansion by bolstering labor productivity. This, in turn, contributes to market concentration. Following the literature, we hypothesize that AI adoption increases industry concentration as AI-generated growths favor large firms.

Our identification strategy hinges on the observation that there has been a swift and exogenous increase in the availability and applicability of AI technologies (algorithms and tools) within AI-using industries, as demonstrated earlier in this section. Conversely, the avenues through which non-tech businesses expand their sales and capture more market shares are expected to evolve at a much slower pace. As firms change their AI adoption in response to technological advancements, variations across different firms and over time in the relationship between AI adoption and sales share emerge, allowing for the identification of potential causal relationships.

In a specific methodology design, with direct estimation of the relationship between AI adoption and concentration as a preface, I dive into firm-level investigation using panel data. As the concentration is normally measured by the sales share of top firms, the equivalent question to whether AI adoption increases industry concentration is whether AI adoption exclusively increases the sales share of top firms. In designing the specific methodologies, I begin by directly estimating the relationship between AI adoption and industry-level concentration, serving as a preface for delving into firm-level investigation. Given that concentration is typically measured by the sales share of top firms, the central question to answer empirically becomes whether AI adoption leads exclusively to an increase in the sales share of these top firms. To estimate the overall impact of AI adoption on the sales share of top firms, we utilize the long difference model to capture the cumulative effects over time under technological progress. To address potential omitted variable bias and reversed causality problems, we extended our analysis by developing a difference-in-difference model, using the launch of TensorFlow as a quasi-natural experiment (Rock, 2019).

We need to measure both sales share and AI adoption. We measure firm's AI adoption level as the fraction of employees with AI skills, based on data extracted from LinkedIn member profiles. We also collect data on various human capital factors such as total headcount, average employee tenure, and distribution of employees by education level. We then combine this human capital data with sales and financial data from Compustat, aligning with methodologies utilized in previous studies on AI adoption and investment (Babina et al., 2024; Rock, 2019; etc.). This integration constructs panel data covering from 2009 to 2019 for U.S. publicly traded firms and thus allows us to analyze causal relationships effectively.

Overall, we discover that the adoption of AI significantly increases the sales share of the top 20 firms exclusively, as defined by their sales share in their industries. This suggests a rise in industry concentration attributable to AI adoption. Our results remain robust across various econometric approaches and time cut-offs. As secondary findings, we demonstrate that following AI adoption, leading and non-leading firms exhibit differing performance in product creation, yet they show no significant difference in operating costs and labor productivity. These results exclude operating costs and labor productivity as drivers underlying the uneven change in sales share, highlighting the product creation channel for further examination.

This thesis introduces several innovations that expand upon existing literature. First, we provide novel evidence of the impact of AI adoption on the increase in industry concentration across a wide range of industries, linking the recent surge in AI advances with longstanding concerns regarding increasing concentration. Second, We creatively employ firm-level data to examine industry-level aggregated trends, particularly concentration. This approach enables us to more effectively control for the endogeneity in AI adoption concerning firm performance, and consequently, industry characteristics. We thus complement recent research that provides evidence of the impact of AI investment on aggregated industry-level measures (Babina et al., 2024). Thirdly, we innovatively explore the differences in the relationship between AI adoption and product creation, operating costs, and labor productivity. This extends beyond previous studies, which primarily focused on the general relationship between AI adoption and these growth channels.

The remainder of the paper is organized as follows. Section 2 summarizes related literature for the following empirical examination. The data and measures are discussed in Section 3. In Section 4, detailed discussions on econometric models are provided, along with the reported results. Section 5 provides preliminary empirical evidence on possible channels that may influence concentration under AI adoption. Section 6 concludes the thesis and discusses future directions.

2 Literature

This paper is closely related to Babina et al. (2024), who used firm-level data to study the effect of AI investment on a firm's growth. My study differs from Babina et al. (2024) by directly focusing on industry concentration and offering firm-level evidence. Additionally, we employ a distinct source of human capital data.

AI-Driven Business Expansion

We summarize the literature discussing how AI drives business growth in two main points. First of all, as a predictive technology, AI enables firms to efficiently learn from large datasets, enhancing business decision-making under uncertainty—a crucial aspect of a firm's growth (Brynjolfsson and McElheran, 2016; Agrawal et al., 2019). What sets apart popular AI techniques such as machine learning, natural language processing, and computer vision from traditional data analysis methods is their capability to learn from extensive volumes of high-dimensional data—encompassing text, speech, and image data—with enhanced accuracy in predictions.

Furthermore, AI automates various business processes such as data entry, processing, and customer service. Aghion et al. (2017) model AI as the latest form of automation. Statistics from the 2019 Annual Business Survey support the view of AI as a tool for automation, aligning with how it's commonly used in business practices — 55% of AI users report adopting AI for automation (Acemoglu et al., 2022).

To look deeper into how AI enhances businesses and supports the aforementioned points, we explore potential channels through which AI can benefit businesses, summarizing three such channels identified in the literature. Firstly, AI could be a driver for product innovation by either facilitating the creation of new products or improving existing ones. On the one hand, by enhancing decision-making through prediction capabilities, AI can mitigate uncertainty in experimentation and enhance the efficiency of identifying promising projects, thereby expediting the process of creating new products (Babina et al., 2024). For example, an AI-proficient company can prioritize features more effectively during product development. On the other hand, AI creates more opportunities for product enhancement through the integration of AI models into existing products, thereby harnessing automation capabilities like chatbot functions. In addition, AI's prediction power can also enhance understanding of customer preferences and personalize product design and engagement marketing in existing products (Kumar et al., 2019). Empirically, Wu et al. (2020) show that data analytics skills are particularly linked to innovation necessitating extensive information processing, while novel innovation or narrow recombination seem not to benefit from data analytics.

Secondly, we consider labor productivity enhancement as another possible channel in which AI might more effectively replace or synergize with human labor. Agrawal et al. (2019) propose that AI directly replaces capital with labor for prediction tasks; however, they also suggest that the indirect impact of AI on decision tasks, which are closely intertwined with prediction tasks, remains ambiguous. Thirdly, cost-saving is another possible channel. While its mechanism mirrors the second one, cost-saving emphasizes absolute reduction, whereas productivity enhancement is a more relative measure. Likewise, since AI has the potential to automate certain tasks, it could lead to a reduction in labor costs. What's more, with its enhanced prediction accuracy, AI has the potential to mitigate error costs, as both overly optimistic and pessimistic forecast errors can diminish productivity (Tanaka et al., 2019). Notably, while the literature shows the potential of these three channels as drivers of AI-generated growth, it lacks a definitive direction, necessitating further examination within the proposed context, particularly regarding the second and third channels.

Growth Advantage for Large Firms

Why do we anticipate an increase in industry concentration despite AI's potential to drive substantial business growth? Based on the existing literature, we outline two primary growth advantages for larger firms. Firstly, large firms have big data ownership, which serves as the fuel for AI applications. According to Farboodi et al.'s theoretical model (2019), data accumulation influences business dynamics by skewing the distribution of firm sizes towards larger companies that generate more data and invest more in active experimentation. Begenau et al. (2018) also propose that big data predominantly advantages large firms in the realm of finance, due to their substantial data generation capacity. This, coupled with advancements in processor speed, enables more intricate data analysis that enhances investor predictions, diminishes equity uncertainty, and ultimately decreases the cost of capital for these sizable enterprises. From the market competition perspective, Furman and Seamans (2018) argue that the requirement for large data sets is a barrier to entry.

Secondly, large firms benefit from economies of scale. When we view AI investments—such as algorithms, computing technology, and data infrastructure—as fixed costs, large companies with diverse lines of business possess more application scenarios and potential network effects. This economies of scale is especially evident in key technologies like algorithms, where the winner often dominates all. Brynjolfsson and McElheran (2016) present evidence indicating that a firm's size strongly correlates with its adoption of data-driven decision-making practices, aligning with the concept of economies of scale.

Human Capital Measure: A Proxy for AI Adoption

We measure technically skilled human capital from employee profile data and employ those with AI skills as a proxy measure for AI adoption, in accordance with existing literature. Fedyk and Hodson (2022) utilize detailed resume data from Cognism to quantify the technical human capital of U.S. firms in IT, Software Engineering, Data Analysis, and other fields and then link this data to measures of performance among publicly traded firms. Babina et al. (2024), building on the work of Fedyk and Hodson, utilize the same dataset, with a particular emphasis on AI-skilled human capital, as a proxy for measuring AI investment. Additional studies utilize LinkedIn profile data. Tambe et al. (2020) create measures of IT labor by examining the employment histories of individuals on LinkedIn with job titles indicating they work in

IT. Meanwhile, Rock (2019) utilizes detailed AI skills data from LinkedIn to assess AI talent and merge it with Compustat measures of financial performance.

TensorFlow: A Shock for AI Adoption

We construct the launch of TensorFlow as a quasi-natural experiment following Rock (2019). TensorFlow is a machine learning framework developed by Google that provides tools and libraries to build and deploy, empowering developers to train and optimize AI solutions quickly and efficiently. The open-source version was launched by Google on November 9, 2015. TensorFlow demonstrates remarkable performance in commercial applications by utilizing dataflow graphs to efficiently handle computation and state across distributed environments (Abadi et al., 2016). TensorFlow emerged as the leading AI software library, experiencing immediate and explosive popularity upon its launch. This surge was evident in the rapid increase in GitHub stars, starting from its launch in 2015 (Zhang et al., 2021). GitHub serves as a platform where AI researchers and developers can host their code, with users having the ability to "star" projects for saving, thereby enabling the measurement of open-source library popularity as researchers upload packages referencing these libraries. Then Keras and PyTorch enter as competitors for TensorFlow.

Following Rock's argument (2019), we assert that the launch of TensorFlow can be viewed as a quasi-natural experiment for two main reasons. Firstly, Google's decision to open-source TensorFlow was unexpected, leading to little preparation and thus self-selection concern. Secondly, the launch of TensorFlow has had a significant impact on the proportion of AI-skilled employees by reducing the technical barriers to implementing deep learning, which has enabled individuals with limited AI expertise to utilize advanced algorithms effectively.

3 Data and Measurement

3.1 Employment Profiles from LinkedIn

Member profile information from LinkedIn is my primary data source for gauging the fraction of AI-skilled workers within each company. This resume-like approach's effectiveness in addressing my research question can be attributed to its dual-layered rationale. Firstly, human capital measures serve as effective proxies for assessing AI adoption. Specifically, AI-skilled labor is crucial for effective AI implementation since other factors, such as data availability and computing infrastructure, complement AI-skilled human capital. Hence, utilizing this human capital metric enables us to evaluate the varying degrees of adoption of AI technologies among different firms. Secondly, LinkedIn provides comprehensive human capital data owing to its detailed information and extensive coverage. LinkedIn connects organizations with their employees' skills, educational backgrounds, and job histories, providing firm-level data that includes detailed information about individual employees. In addition, it has become a standard tool for job seekers in various labor markets. As of Daniel's report (2019), LinkedIn boasts over 575 million members across 200+ countries and territories, including more than 150 million active members. The platform features representations from over 26 million companies, 60 thousand educational institutions, and 35 thousand skills.

To fully leverage LinkedIn member profile data, we conducted several processing steps to compile panel human capital records at the firm level. In specific, in each individual's profile page, we capture key details from each employment record, including start and end dates, job title, company name, and job description. To ensure consistency, we standardize the reported organizations, enabling us to link each employment record to a specific firm. Within each record, we assess the presence of AI-related skills by examining the job description for relevant keywords. We then aggregate the total number of employees possessing such skills for each firm. This yields a series of aggregated counts of AI workers, aggregated at the firm-month level. Additionally, we extract various employment variables from LinkedIn resume profiles, including total headcount, average employee tenure (in years), and counts of employees holding different education degrees (high school, associate, bachelor, master, MBA, doctorate).

Measurement error of LinkedIn-derived data

The LinkedIn panel's representativeness is limited by factors such as demographic coverage, time span, and the presence of inaccurate information.

Firstly, LinkedIn data exhibits coverage gaps, particularly among less educated and blue-collar workers. According to the Pew Research Center's Social Media Use study, 50 percent of adults in the US with a bachelor's or advanced degree are LinkedIn users, compared to only 10 percent of individuals whose education level does not exceed high school. Additionally, a study conducted by LinkedIn indicates that in 2021, 44 percent of LinkedIn users reported annual incomes exceeding $75,000, surpassing the US national median income for the same year. This gap may be attributed to the fact that certain professions, such as blue-collar jobs, do not extensively utilize LinkedIn for job searching and networking, resulting in fewer incentives for these workers to share their information on the platform. This can result in potential discrepancies between the observed LinkedIn population, their job types, and the broader employee demographic.

Moreover, variations in human capital data across different years could be influenced by the gradual adoption of LinkedIn throughout the workforce. As LinkedIn gained popularity and attracted more users, the completeness of its users' work histories became dependent on members providing detailed information on their profiles. If users were less motivated to fill in their work history, the increase in certain human capital metrics over the years might be due to new users joining the platform rather than reflecting actual changes in the labor market.

Furthermore, concerns about data accuracy and potential inflation or underestimation of self-reported skills arise from the incentives individuals may have to exaggerate their skills. In particular, individuals who are actively seeking employment usually keep their profiles more up-to-date, including information on their employment history and skills. However, despite the potential consequences, dishonest reporting on LinkedIn does occur among job seekers. This could lead to an overestimation of true skill levels within firms, thereby establishing lower bounds in subsequent regression analyses. Additionally, underreporting on the LinkedIn platform is a more concerning issue. It's challenging to derive an accurate estimate of skill stocks as workers frequently omit their qualifications on their profiles. Consequently, regression estimates may represent upper bounds.

Several strategies are employed to address potential biases. Firstly, firm, industry-time, and time fixed effects are incorporated into all regression specifications. Additionally, we construct a "high education share" variable (define as the fraction of employees with master, mba, or doctor degree) for each firm in each time period, serving as a control for the overall level of human capital and thus mitigating potential confounding factors related to variation in firm-level general human capital stocks.

3.2 Measure, Summary Statistics, and Validation

We measure the adoption of AI within firms by calculating the percentage of employees classified as AI-skilled. This metric, termed AI share, is calculated by dividing the number of AI-skilled employees by the total number of employees within each firm during each time period. An employee is considered to be AI-skilled if the description under an employee's position in their profile has AI keywords.

We also integrate employment profile data with firm performance and industry data. This involves aggregating employment records to firm-year and firm-quarter levels, and then merging them with Compustat Fundamentals Annual files and Fundamentals Quarterly files, respectively. The resulting records include various metrics such as NAICS classification, sales figures, asset values, equity, cash reserves, cost of goods sold, common shares outstanding, debt obligations, net income, and long-term debt. From these Compustat measures, we derive primary outcome variables such as sales shares, calculated as a firm's sales value divided by the total sales value of its industry (industry defined at the 2-digit NAICS level). Additional variables encompass cash/assets ratio, markup, Tobin Q, market leverage, and return on assets (ROA). To address skewed distributions, we logarithmically transform the counts of employees, sales, and Tobin Q (adding 1 to zero entries).

Out of the 4,739 US-based firms identified in LinkedIn resume data spanning from 2009 to 2019, 2,811 are matched with publicly listed US firms on NYSE or Nasdaq, and 2,600 are matched with records from Compustat. Among the subset of 2,226 firms not classified as tech or public administration (by excluding 2-digit NAICS industries 51, 54, and 92), 2,061 firms exhibit positive sales and employment figures. We further refine the sample by retaining only firms with non-missing control variables throughout to ensure stability in sample composition. Table 1 presents summary statistics for the unbalanced panel data at both the firm-year and firm-quarter levels covering the period from 2009 to 2019.

We analyzed our constructed measures of AI adoption and verified that they exhibit intuitive properties. Firstly, we observed a natural increase in these measures over time, with a more than five-fold increase from 2009 to 2019. As illustrated in Figures 1, there is a notable surge in the proportion of employees classified as AI-skilled: beginning at 0.07% in 2009 and escalating to 0.29% by 2018.

Variables	Obs.	Mean	Std. Dev.	Min.	25th Perc.	Median	75th Perc.	Max
Firm-Year Level								
AI Share(%)	15,498	.1493	.4594	0	0	0	.1201	15.75
ln(Employment)	15,498	6.79	2.029	-.087	5.368	6.912	8.238	12.49
Duration	15,498	3.571	1.42	.013	2.523	3.525	4.488	19.56
High Edu Share(%)	15,498	15.38	11.9	0	7.44	11.8	19.5	87.5
Sales Share(%)	15,498	.6360	2.128	.0000	.0079	.0612	.3645	50.71
ln(Sales)	15,498	6.194	2.612	-6.215	4.708	6.538	7.973	12.98
Cash/Assets	15,498	.1503	.1912	-.0008	.0234	.0798	.1917	.9936
Markup	15,498	.4498	1.107	-8.739	.2153	.4232	.8051	6.48
ln(Tobin Q)	15,498	.5660	.6301	-1.163	.1064	.4223	.8707	8.519
Market leverage	15,498	.1696	.1716	0	.0304	.1234	.256	.963
ROA	15,498	-.1029	2.89	-397.4	-.0273	.0240	.0666	9.738
Firm-Quarter Level								
AI Share(%)	62,339	.1441	.4441	0	0	0	.1149	16.67
ln(Employment)	62,339	6.845	2.072	-.4055	5.394	7.002	8.331	13.05
Duration	62,339	3.493	1.432	.0509	2.431	3.431	4.42	22.17
High Edu Share(%)	62,339	14.78	11.59	0	6.915	11.54	18.91	100
Sales Share(%)	62,339	.6735	2.193	.0000	.0099	.0742	.4129	52.01
ln(Sales)	62,339	4.932	2.597	-6.908	3.462	5.296	6.696	11.83
Cash/Assets	62,339	.1456	.1836	.0008	.0251	.0786	.1867	1
Markup	62,339	.3759	1.015	-10.58	.2073	.3988	.697	7.486
ln(Tobin Q)	62,339	.5785	.6298	-1.794	.1388	.4364	.8814	8.519
Market leverage	57,412	.1693	.1736	0	.0261	.1243	.256	.9812
ROA	62,339	-.0181	.8162	-52.11	-.0061	.0078	.0196	203.5

Table 1: Summary Statistics

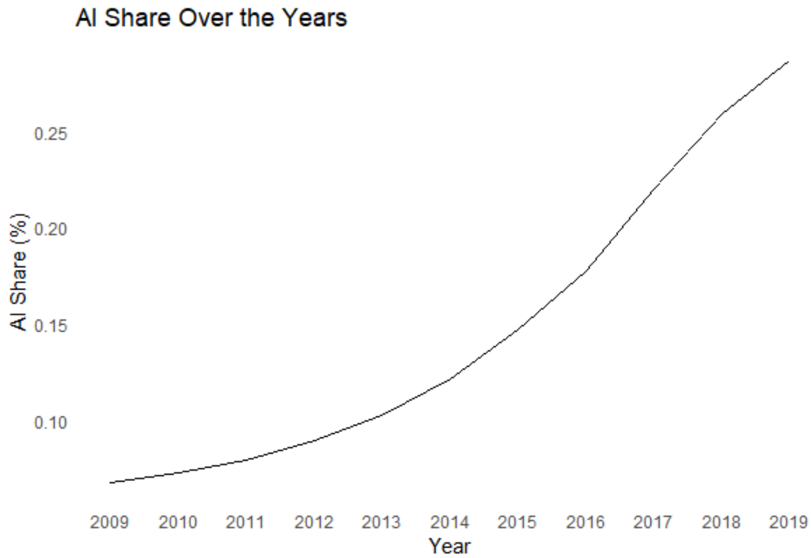

Figure 1: AI Share Over Years

Note: Line chart of fraction of AI-skilled employees as a function of year from 2009 to 2019.

There is large variability in the growth of AI-skilled labor among individual firms, which provides the necessary heterogeneity to explore the relationship between AI adoption and firm outcomes. Across our entire sample, the median firm experiences a modest increase of 0.0076% in the AI workforce share. However, this increase is more substantial for firms at higher percentiles: 0.4352% at the 90th percentile, 0.8052% at the 95th, and 2.448% at the 99th percentile.

Secondly, we observe a discernible pattern in the distribution of AI-skilled workers across industries. Figures 2 illustrate the proportion of AI-skilled workers in public firms within each 2-digit NAICS sector, separately for the periods 2009–2014 and 2015–2019. The data underscores that the Manufacturing sector, which encompasses manufacturers of machinery, computer and electronic products, and electrical equipment and appliances, hosts the highest share of AI-skilled workers. Over time, the proportion of AI workers in this sector rises from 0.19% in the earlier period (2009–2014) to 0.43% in the later period (2015–2019). Additionally, nearly all sectors experience a notable increase in AI-skilled labor.

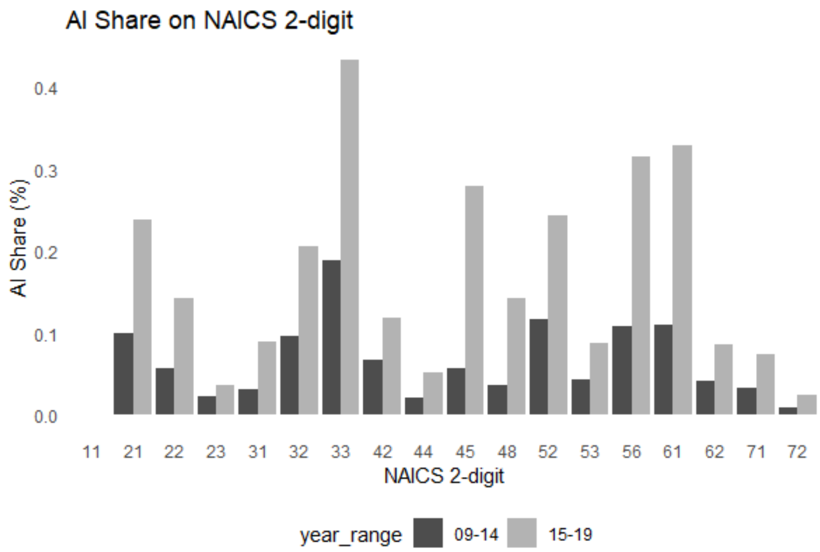

Figure 2: AI Share on 2-digit NAICS

Note: Bar chart of fraction of AI-skilled employees across industries. X-axis is the industry classified by 2-digit NAICS codes. Dark gray bars represent fraction of AI-skilled employees from 2009 to 2014 while light gray bars represent annual fraction of AI-skilled employees from 2015 to 2019.

Thirdly, we present a clear geographical distribution of identified AI workers. In Figure 3, we illustrate the proportion of AI-skilled employees in public firms across different states, while Figure 4 displays the average annual count of AI-skilled employees in each state. Idaho stands out with the highest AI share, attributed to its fewer firms and a significant fraction of AI workers employed in prominent positions at Micron – a leading semiconductor manufacturing company. Despite excluding the tech sector from our data, California and Washington still exhibit a substantial share of AI workers, reaching 0.33% and 0.36%, respectively, between 2009 and 2019.

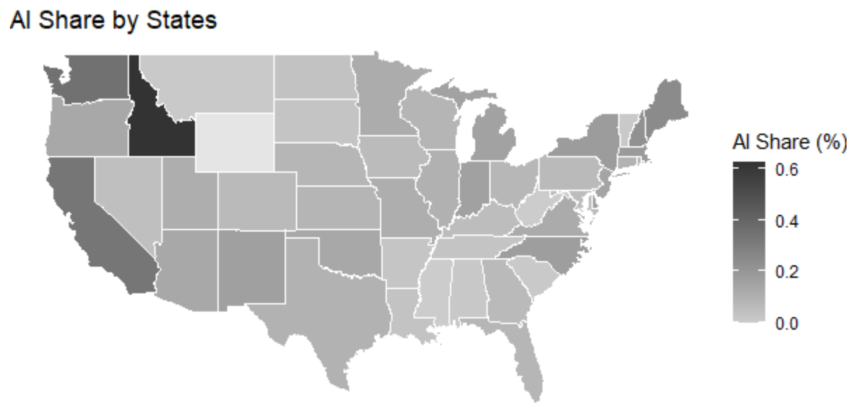

Figure 3: AI Share by State

Note: Heat map of annual fraction of AI-skilled employees across states from 2009 to 2019.

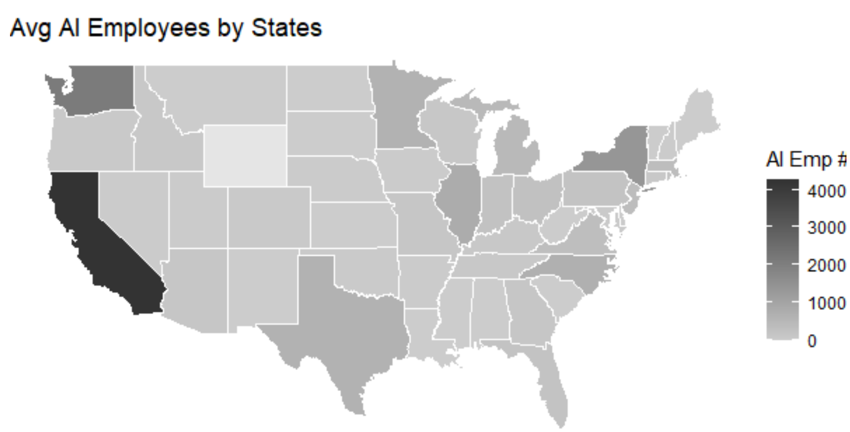

Figure 4: AI Employees by State

Note: Heat map of average annual number AI-skilled employees across states from 2009 to 2019.

4 AI Adoption and Industry Concentration

This section delves into the empirical relationship between AI adoption and industry concentration within the U.S. non-tech sector. Initially, we provide evidence indicating a positive correlation between AI adoption, as measured by AI share, and industry concentration. Subsequently, we present causal evidence illustrating the substantial impact of AI adoption on industry concentration. We consider alternative explanations for this relationship, including the possibility of reverse causality, where industries with higher concentration—i.e., larger firms with a greater sales share—are more inclined to adopt AI. Additionally, we address concerns regarding omitted variables, such as concurrent adoption of other technologies or demand shocks, which may drive both increases in industry concentration and AI adoption.

4.1 Industry-level Results

Our dataset comprises an unbalanced panel encompassing 18 industries (classified by 2-digit NAICS categories) spanning the 11-year period from 2009 to 2019. Table 2 presents a summary of key variables. Notably, the AI share exhibits considerable variations within our sample. For instance, IQR of AI share is approximately 0.1042%, close to the sample mean. Moreover, the standard deviation of AI share nearly matches the mean. These trends reflect both shifts in AI adoption within industries and the heterogeneity of AI adoption across different sectors. Our sample also demonstrates significant variation in industry concentration. On average, the sales share of the top 8 firms constitutes around 53%, with an IQR of roughly 25% and a standard deviation of 18%. Similarly, the sales share of the top 20 firms averages approximately 73%, with an IQR of about 22% and a standard deviation of 17%. We calculate the correlations of AI adoption with industry concentration, controlling for the total number of employees, the total sales value, industry and year fixed effects. To account for sampling weight, we utilize the total number of employees. Given

Variables	Obs.	Mean	Std. Dev.	Min.	25th Perc.	Median	75th Perc.	Max
AI Share(%)	187	.1081	.1023	0	.0354	.076	.1396	.504
CR8(%)	187	53.08	18.49	15.41	40.52	52.05	65.29	98.85
CR20(%)	187	73.39	17.34	33.64	63.29	75.69	85.3	99.99

Table 2: Summary Statistics for Industry-level Measure

our inability to obtain information from non-public firms and those with unreported employee data on LinkedIn, we strive to mitigate this sampling bias by controlling for the total number of employees within each industry. The total sales value is used to measure the time-varying industry scale. Specifically, we estimate the following specification

$$Concentration_{jt} = \beta AIShare_{jt} + \gamma_1 ln(emp_{jt}) + \gamma_2 ln(sale_{jt}) + \alpha_j + \delta_t + \epsilon_{i,t}$$

where the dependent variable $Concentration_{jt}$ is a measure of industry concentration in industry j and year t. $ln(emp_{jt})$ is logged value of a total number of employees, and $ln(sale_{jt})$ is logged total sales value (1 is added to entries with zero). α_j and δ_t are industry and year fixed effects, respectively. The key independent variable is the AI share $AIShare_{jt}$. The term $\epsilon_{i,t}$ denotes the regression residual. The coefficient of interest, β, measures the semi-elasticity of industry concentration with respect to AI share, controlling for aggregate conditions and other industry characteristics.

Table 3 reports the estimation results of the regressions. It shows that AI share is positively correlated with sales concentration (i.e., measured as the sales share of both the top 8 and top 20 firms). The estimated correlation is not statistically significant due to the limited observation and large variation, which inspires us to further firm-level investigation. The point estimate in Column (1) implies that a 1% increase in AI share is associated with an increase in CR8 measure of industry concentration by almost 1%. The point estimate in Column (3) implies that a 1% increase in AI share is associated with an increase in the CR20 measure of industry concentration

VARIABLES	cr8		cr20	
	(1) From 2009	(2) From 2010	(3) From 2009	(4) From 2010
AI Share(%)	1.007	0.670	0.441	0.374
	(0.17)	(0.12)	(0.09)	(0.08)
ln(Emp)	3.335	4.108	0.168	0.888
	(1.02)	(1.25)	(0.06)	(0.34)
ln(Sale)	12.315***	12.359***	9.260***	9.268***
	(5.85)	(5.84)	(5.46)	(5.44)
Observations	187	169	187	169
R-squared	0.271	0.303	0.221	0.252
Number of naics_l2	18	18	18	18
Year FE	Y	Y	Y	Y

Table 3: Industry Level Regression Results

Note: The dependent variable in columns (1) and (2) is CR8 (calculated by summing up the sales share of top 8 firms), and the dependent variable in columns (3) and (4) is CR20 (calculated by summing up the sales share of top 20 firms). Columns (1) and (3) regress on the full sample from 2009 to 2019, and columns (2) and (4) regress on the sub-sample from 2010 to 2019. All columns use the fixed effect model. Robust t-statistics are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

by 0.44%.

4.2 Firm-level Long-difference Results

The statistically insignificant result regarding the estimated correlation suggests the need for further investigation at a more detailed level. Moreover, the correlations observed between AI share and industry concentration do not necessarily imply causal effects, given that both AI adoption and industry concentration are endogenous variables. Omitted variable bias may occur when additional factors, such as those predisposing large firms in certain industries to increase their market share and adopt AI, lead to a simultaneous rise in AI adoption and market concentration within the industry. Therefore, to delve deeper into understanding how the adoption of AI may have impacted industry concentration, we utilize firm-level data, enabling us to control for confounding factors such as other human capital shocks and financial performance at the initial stage.

We begin the analysis by examining whether firms holding a leading position in their respective industries experienced a more substantial gain in sales share through AI adoption from 2009 to 2019. The observation that top firms experience a larger increase in sales share aligns with the rising CRX measure of market concentration. Given the characteristic slow-paced nature of processes like technological progress (e.g., Acemoglu and Restrepo, 2020), our primary analytical approach involves a long-differences regression. This regression assesses changes in a firm's sales share from 2009 to 2022 against changes in AI adoption, approximated by the share of AI workers. Subsequently, we conduct a heterogeneous analysis comparing top firms to non-top firms. As noted by Babina et al. (2024), AI investments unfold gradually over time, with effects that may not manifest immediately. Therefore, this strategy suits our context well for studying the overall impact of AI adoption. By employing first differences in both independent and dependent variables, the long-differences specification ensures that time-invariant firm characteristics do not drive the outcomes.

Specifically, we estimate the following regression:

$$\Delta SaleShare_{i,[t,2019]} = \beta \Delta AIShare_{i,[t,2019]} + \gamma Control_{i,t} + IndustryFE + \epsilon_i$$

where the dependent variable $\Delta SaleShare_{i,[t,2019]}$ measures the annual change in sales share in firm i from year t to 2019. The main independent variable, $\Delta SaleShare_{i,[t,2019]}$, captures the annual change in the share of AI-skilled workers based on the resume data in firm i from year t to 2019. t takes the value from 2009 to 2014. We regress on varying time ranges to study potential time effects and check the robustness of the results. As in the previous section, this analysis focuses on firms in non-tech sectors. Industry FE is 2-digit NAICS category industry fixed effects. We include a rich set of controls that are all measured at the start of the sample period t: (i) the initial firm size (log sales, log employment); (ii) the initial firm-level characteristics that could predict changes in firm performance in the future (cash/assets, log markup, log tobin Q, market leverage, roa); (iii) characteristics of the general human capital (the share of high education workers, log average duration). In this analysis, top firms are defined as the top 20 firms in sales share. Table 4 reports the regression results for top firms, Table 5 reports the regression results for non-top firms, and Table 6 reports the hypothesis test results for the difference in the effect of AI adoption in sales share.

Our results show that regression of annual change of sales shares returns a coefficient of 3.428% per 1% increase in the annual change of AI share for the top 20 firms in the specification (1) from 2009 to 2019 in Table 4. In alternative specifications for the regression of annual changes in sales shares for the top 20 firms, other point estimates range from approximately 1.995% to 3.344%. These estimates exhibit a diminishing trend in both value and precision as the time difference range decreases. Regression of annual change of sales shares returns a coefficient of 0.009% per 1% increase in the annual change of AI share for the non-top firms in the specification (1) from 2009 to 2019 in 5. Results in all specifications for regression of non-top firms

VARIABLES	Δ sales share(pp) for Top 20 firms					
	(1) t=2009	(2) t=2010	(3) t=2011	(4) t=2012	(5) t=2013	(6) t=2014
Δ AI Share(pp)	3.428***	3.344***	3.161***	2.782**	2.579**	1.995*
	(2.85)	(2.86)	(2.75)	(2.53)	(2.42)	(1.94)
ln(Emp)	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(3.09)	(3.17)	(3.46)	(3.00)	(3.16)	(2.97)
Duration	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.83)	(-0.54)	(-0.63)	(-0.59)	(-0.32)	(-0.64)
Highedu	-0.010**	-0.008**	-0.009**	-0.007*	-0.005	-0.003
	(-2.57)	(-2.46)	(-2.14)	(-1.65)	(-1.35)	(-0.82)
ln(Sale)	-0.001**	-0.001**	-0.001**	-0.001	-0.000	-0.000
	(-2.30)	(-2.27)	(-2.06)	(-1.58)	(-1.05)	(-0.47)
Cash/Assets	0.004**	-0.001	0.001	0.002	0.001	0.003
	(2.49)	(-0.46)	(0.52)	(0.56)	(0.53)	(1.00)
ln(Markup)	-0.000	0.000	-0.000	-0.000	-0.000	-0.000
	(-0.41)	(0.01)	(-0.11)	(-0.84)	(-1.33)	(-1.42)
ln(Tobin Q)	0.001	0.001	0.001*	0.001	0.001	0.001*
	(1.60)	(1.57)	(1.69)	(1.51)	(1.57)	(1.91)
Market leverage	0.002	0.001	0.002	0.002	0.001	0.001
	(1.50)	(0.83)	(1.28)	(1.10)	(0.43)	(0.84)
ROA	-0.004	-0.007	-0.011*	-0.006	-0.010**	-0.006
	(-1.40)	(-1.63)	(-1.67)	(-0.98)	(-1.97)	(-1.07)
Observations	285	285	286	282	282	285
R-squared	0.202	0.175	0.153	0.136	0.138	0.145
Year FE	Y	Y	Y	Y	Y	Y

Table 4: Long Difference Regression Results for Top 20 Firms

Note: The dependent variable is the difference in sales share for top 20 firms. All columns use fixed effect model. Robust t-statistics is reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

VARIABLES	Δ sales share(pp) for Non-top firms					
	(1) t=2009	(2) t=2010	(3) t=2011	(4) t=2012	(5) t=2013	(6) t=2014
Δ AI Share(pp)	0.009*	0.009*	0.007	0.007	0.006	0.004
	(1.81)	(1.73)	(1.48)	(1.62)	(1.47)	(1.08)
ln(Emp)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
	(5.01)	(4.94)	(4.64)	(5.20)	(4.99)	(4.45)
Duration	-0.000***	-0.000***	-0.000**	-0.000**	-0.000**	-0.000**
	(-2.88)	(-3.05)	(-2.33)	(-2.49)	(-2.01)	(-2.17)
Highedu	0.000	0.000	0.000	0.000*	0.000	0.000
	(0.25)	(1.01)	(0.97)	(1.91)	(1.47)	(1.16)
ln(Sale)	-0.000**	-0.000**	-0.000	-0.000**	-0.000***	-0.000***
	(-2.29)	(-2.13)	(-1.54)	(-2.35)	(-2.94)	(-2.66)
Cash/Assets	0.000	0.000	0.000	0.000	-0.000	-0.000
	(1.19)	(0.03)	(1.15)	(0.13)	(-1.02)	(-1.02)
ln(Markup)	0.000	0.000	0.000	0.000	0.000***	0.000***
	(1.60)	(1.43)	(0.93)	(1.22)	(3.10)	(3.43)
ln(Tobin Q)	0.000	0.000	0.000	0.000	0.000**	0.000**
	(1.29)	(0.87)	(1.32)	(1.23)	(2.05)	(2.33)
Market leverage	0.000	0.000	0.000	0.000	0.000	0.000
	(0.61)	(0.93)	(0.93)	(1.17)	(1.11)	(1.10)
ROA	-0.000	0.000	0.000	0.000*	0.000***	0.000***
	(-0.19)	(0.06)	(0.54)	(1.89)	(2.74)	(2.58)
Observations	285	285	286	282	282	285
R-squared	0.202	0.175	0.153	0.136	0.138	0.145
Year FE	Y	Y	Y	Y	Y	Y

Table 5: Long Difference Regression Results for Non-top Firms

Note: The dependent variable is the difference in sales share for not top 20 firms. All columns use fixed effect model. Robust t-statistics is reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1) t=2009	(2) t=2010	(3) t=2011	(4) t=2012	(5) t=2013	(6) t=2014
$\beta^{top} - \beta^{nontop}$	3.419***	3.335***	3.154**	2.775**	1.573**	1.991**
	(8.87)	(8.97)	(8.28)	(6.99)	(6.44)	(4.11)
Control	Y	Y	Y	Y	Y	Y

Table 6: Heterogeneity Test for Long Difference Regression Results

Note: The chi2 is is reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

are generally small and less statistically significant, showing that non-leading firms gain little in adopting AI. Our main interest is the heterogeneity between the effect of AI adoption among top and non-top firms, which is quantified as $\beta^{top} - \beta^{nontop}$ obtained by heterogeneous analysis. Specification (1) in 6 shows that 1% increase in the annual change of AI share leads to a 3.914% of difference in annual change of sales shares for a firm with and without top 20 position, controlling for other confounding variables.

4.3 Firm-level Difference-in-Difference Results

The above long-difference regression shows a significant effect of AI adoption in the further expansion of large firms. This relationship remains robust after accounting for initial firm characteristics that could potentially influence both sales share and AI adoption over subsequent years. However, these estimates may not solely represent the causal effects of AI adoption, as there are concerns regarding reverse causality and the omission of relevant variables. One such factor could be the introduction of other information technologies, which may simultaneously increase both sales share and the presence of AI-skilled workers within large firms. To mitigate these biases, we employ a difference-in-difference (DID) approach in this section. Following the methodology outlined in Daniel's (2019) study, we utilize the launch of TensorFlow in November 2015 as a natural experiment to construct our DID model. To capture the dynamic effects of this event more effectively, we analyze the data at the quarterly level within this sector.

4.3.1 Fixed-effect Results

Before employing the DID approach, we first employ a fixed effect model to test the correlation between AI share and the top firm's sales share in quarter-level measures. Specifically, we estimate the following specification for the top 20 firms and non-top firms:

$$SaleShare_{it} = \beta AIShare_{it} + \gamma Control_{it} + \mu_i + \nu_t + \epsilon_{it}$$

where the dependent variable $SaleShare_{it}$ is the sales share of firm i in year-quarter t. The main independent variable, $AIShare_{it}$, is the fraction of AI-skilled workers working in firm i in year-quarter t. Regression results using an unbalanced panel of publicly traded firms for all quarters from 2009-2019 are shown in Table 7. The specifications in columns (1) and (2) only control for firm and time fixed effects. In column (1), a regression of sales share returns a coefficient of 0.9% per 1% increase in the AI share for the top 20 firms, with a 90% significance level. The specifications in columns (2) and (3) include a variety of controls for firm performance and human capital as listed in the previous sector.

4.3.2 Main Difference-in-Difference Results

As established in the preceding section, there exists a correlation between AI adoption and industry concentration, even after adjusting for various controls. This correlation demonstrates a statistically and economically significant relationship. To assess the causal impact more thoroughly, we propose employing a series of difference-in-difference methodologies. Considering the launch of TensorFlow as a natural experiment, we set up the following difference-in-difference framework. The treatment group comprises firms with positive AI-skilled workers, while the control group consists of firms without AI-skilled workers, serving as a proxy for AI adoption. In our empirical estimation, we incorporate various confounding factors to isolate the effect of AI investment. In particular,

$$SaleShare_{it} = \beta_1 AIShare_{it} + \beta_2 [PostTensorFlow_t * AI_{it}] + \gamma Control_{it} + \mu_i + \nu_t + \epsilon_{it}$$

The coefficient β_2 is of main interest, indicating the causal impact of the TensorFlow shock (if all necessary assumptions for identification are upheld). This coefficient

VARIABLES	Sales Share(%)			
	$\beta^{top} - \beta^{nontop} = 0.9$***		$\beta^{top} - \beta^{nontop} = 0.539$***	
	(1)	(2)	(3)	(4)
	Top 20	Not Top 20	Top 20	Not Top 20
AI Share(pp)	0.905*	0.005	0.543	0.004
	(1.66)	(1.18)	(1.59)	(0.98)
ln(Emp)			-1.086*	-0.004
			(-1.77)	(-0.41)
Duration			0.237	-0.001
			(0.94)	(-0.47)
Highedu			11.745**	-0.024
			(2.27)	(-0.99)
ln(Sales)			2.444***	0.044***
			(7.34)	(12.06)
Cash/Assets			0.696	-0.001
			(1.57)	(-0.18)
ln(Markup)			-0.322	-0.019***
			(-1.60)	(-8.98)
ln(Tobin Q)			0.208	0.001
			(1.15)	(0.63)
ROA			-0.931	-0.000**
			(-0.89)	(-2.20)
Observations	14,838	47,501	14,838	47,496
R-squared	0.013	0.033	0.377	0.171
Number of Firms	471	1,422	471	1,422
Firm FE	Y	Y	Y	Y
Quarter FE	Y	Y	Y	Y
Control	N	N	Y	Y

Table 7: Quarter Level Fixed Effect Regression Results

Note: The dependent variable is the sales share. Columns (1) and (3) regress on the top 20 firms, and columns (2) and (4) regress on the non top 20 firms. All columns use fixed effect model. Robust t-statistics are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

shows the effect of AI adoption during the post-period. Firm and time fixed effects (μ_i and ν_t) are integrated into the analysis to account for idiosyncratic characteristics of firms and time-specific variations, respectively, thereby adjusting for unit-specific and time-specific disparities in means across firms. The estimation derived from this equation delineates the incremental increase in sales share for firms employing AI technology during the post-period. Upon accounting for time-varying levels of AI adoption, trend variables, and an extensive array of controls for firm performance and human capital, any observed rise in the sales share of AI-utilizing firms compared to non-AI employing counterparts in the post-period signifies evidence of a causal effect after the acquisition of AI skills has become more accessible. Critical assumptions necessary for drawing such conclusions include: 1) the increasing ease of AI adoption during the post-period, 2) parallel trends in sales share between AI-utilizing and non-AI employing firms prior to the accessibility of AI skills, suggesting that AI-using firms would not have exhibited higher sales shares in a counterfactual world without TensorFlow, 3) the absence of a relationship between pre-existing sales share and exposure to TensorFlow, thus negating the possibility that firms with higher sales shares had preferential access, and 4) adherence to the stable unit treatment value assumption (SUTVA). Notably, Google is excluded from all analyses to mitigate potential endogeneity concerns, as the tech sector has been omitted.

Furthermore, to analyze industry concentration, we employ a DID model separately for top firms and non-top firms, comparing the causal effect of TensorFlow shocks. We first examine the impact of TensorFlow launches on sales share growth for AI-using versus non-AI using firms within both top and non-top firm groups. Subsequently, we assess the difference in the net gain from TensorFlow launches between top and non-top firms. The study encompasses three years preceding and following the launch of TensorFlow, thereby encompassing all quarters from 2012 to 2017. Furthermore, to ensure robustness, two definitions of sales firms are employed: top 8 and top 20 in sales share. The results are below in Table 8.

VARIABLES	Sales Share(%)			
	$\beta_2^{top} - \beta_2^{nontop} = 0.19^{***}$		$\beta_2^{top} - \beta_2^{nontop} = 0.21^{**}$	
	(1)	(2)	(3)	(4)
	Top 20	Not Top 20	Top 8	Non Top 8
AI Share(%)	0.473***	0.000	0.431**	-0.004
	(2.61)	(0.04)	(2.16)	(-0.72)
PostTensorFlow $*$ AI	0.199***	0.010***	0.238*	0.030***
	(2.75)	(3.94)	(1.69)	(5.16)
ln(Emp)	-1.030	0.002	-1.286	0.014
	(-1.39)	(0.18)	(-0.77)	(0.86)
Duration	0.193	-0.001	0.574*	-0.004
	(1.00)	(-0.47)	(1.71)	(-0.92)
Highedu	7.250	-0.006	20.799**	0.041
	(1.31)	(-0.36)	(2.22)	(1.14)
ln(Sale)	2.083***	0.019***	3.185***	0.040***
	(13.14)	(8.61)	(11.60)	(8.06)
Cash/Assets	0.297	-0.004	0.425	-0.001
	(0.88)	(-0.76)	(0.56)	(-0.07)
ln(Tobin Q)	0.256	0.001	0.538	-0.002
	(1.46)	(0.58)	(1.20)	(-0.66)
ROA	-0.737	-0.000	-2.825**	-0.000
	(-1.33)	(-1.07)	(-2.46)	(-1.15)
Observations	7,607	25,090	3,255	29,442
R-squared	0.398	0.121	0.498	0.100
Number of Firms	408	1,276	197	1,453
Firm FE	Y	Y	Y	Y
Quarter FE	Y	Y	Y	Y

Table 8: DID Results

Note: The dependent variable is the sales share. Columns (1) and (3) regress on the top 20 firms, and columns (2) and (4) regress on the non top 20 firms. All columns use fixed effect model. Robust t-statistics is reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Results report economically and statistically significant impacts effects of AI adoption in the post period. The coefficients on the AI x TensorFlow post-period variable vary for top firms between 0.199% in column (1) to 0.238% in column (3), with 99% and 90% confidence levels, respectively. However, it's important to note that including a time dummy for the launch of TensorFlow introduces a caveat. This 'DID' variable encapsulates all changes that occurred during that specific time window, potentially resulting in an overestimation of TensorFlow's impact at launch.

4.3.3 Event Study Results

The results in the previous sector establish that there is a post-period effect, though it does not assign a mechanism. Indeed, the effect might not even coincide with the TensorFlow launch. We estimate therefore a new regression, interacting with the AI share with each time period to create an event study version of the equation:

$$SaleShare_{it} = \beta_1 AIShare_{it} + \sum_{t=1}^{13}[\beta_{2t} AI_{it} * Z_t] + \gamma Control_{it} + \mu_i + \nu_t + \epsilon_{it}$$

In this equation, each β_{2t} is unique for the time period t. For event study specifications, I follow the specifications of columns (1) and (2) in Table 8. The results of estimated β_{2t} are summarized in Figure 5 and 6 below.

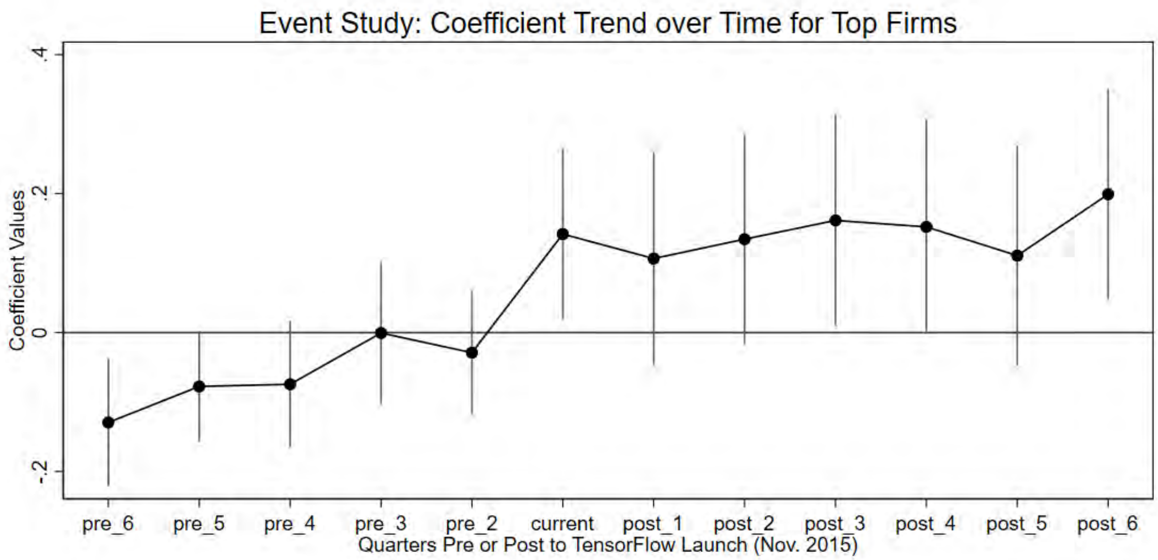

Figure 5: Event Study Results for Top 20 Firms

Note: Line chart of estimated coefficients on TensorFlow launch for top 20 firms as a function of quarters pre or post to TensorFlow launch. X-axis shows the time difference in the unit of quarter to TensorFlow launch. Y-axis is the estimated impact of TensorFlow launch in that period.

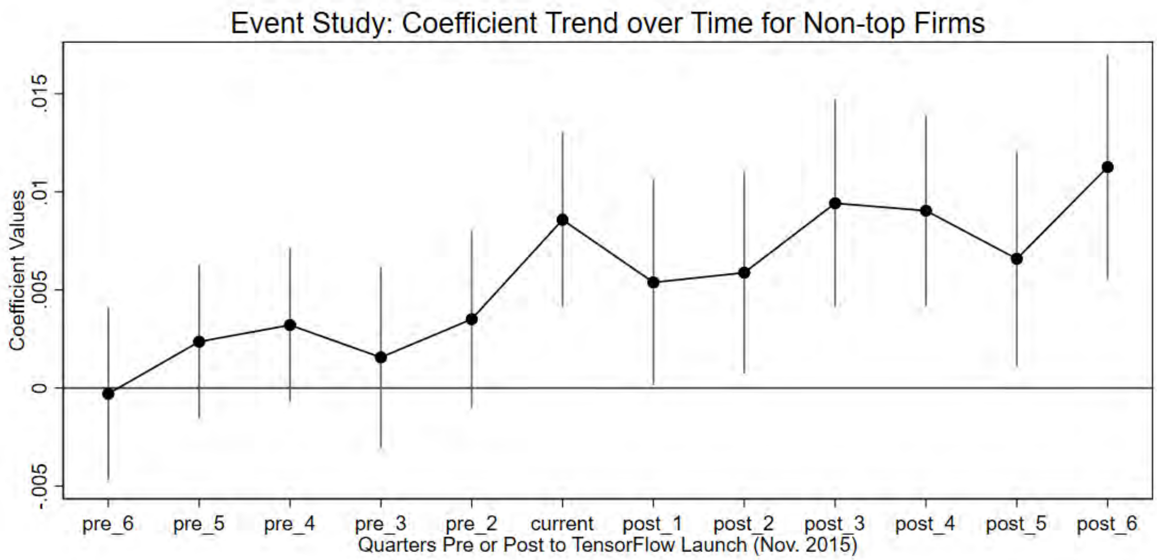

Figure 6: Event Study Results for Not Top 20 Firms

Note: Line chart of estimated coefficients on TensorFlow launch for not top 20 firms as a function of quarters pre or post to TensorFlow launch. The X-axis shows the time difference in the unit of quarter to the TensorFlow launch. The Y-axis is the estimated impact of the TensorFlow launch in that period.

In our result, following the introduction of TensorFlow, there is an immediate uptick in sales share among firms utilizing AI in comparison to the baseline of firms abstaining from AI implementation (where the interaction term of AI share and time dummies for non-AI firms consistently yields zero). The robustness of the estimated effect is proved as the majority of the post-period observations exhibit statistically significant results at the 95% confidence level. Additionally, firms employing AI at the forefront experience substantial increases in sales share, whereas those employing AI without holding a leading position observe more modest increments.

Additionally, the estimated coefficients for the period before the TensorFlow launch can be used as a pre-trend test: if firms adopting AI are on similar growth trends as other firms prior to the TensorFlow lauch (AI becoming cheaper to adopt), β_{2t} with $t < 0$ should be statistically indistinguishable from zero. The graphs suggest no evidence of pre-trends in sales shares: conditional on the controls we include, firms adopting AI in any given year show comparable paths of sales share in prior years and start diverging only afterward. This provides additional evidence that our results are not capturing the reverse effect of industry concentration (top firm position) on AI adoption or the effect of omitted variables placing AI-investing firms on differential growth trajectories, helping to bolster a causal interpretation of our DID results.

5 Mechanisms

We explore how the adoption of AI contributes to biased growth, leading to industry concentration among leading firms compared to non-leading ones. This examination focuses on three potential channels, as outlined in Section 2: product creation, operating cost savings, and labor productivity enhancement. We explore this by analyzing the varied performance across these three areas upon adoption.

5.1 Product Creation

As outlined in Section 2, AI can foster firm growth through product innovation, both by expediting the creation of new products and by enhancing existing ones. To empirically explore this, we require firm-level data on products and services. However, due to challenges in data availability, I could only obtain proxies to measure new product creation using the filing cases of trademarks from USPTO. Trademarks must be filed and registered whenever new products or services are prepared for commercialization, making them a suitable proxy for tracking the creation of new products and services (Hsu et al., 2021).

We acquired raw trademark data, including case details, owner information, and filing dates. Subsequently, we processed and aggregated this data to the firm-year level. We then matched the names of trademark owners with company names in

Compustat to obtain consolidated panel data for publicly traded U.S. firms from 2009 to 2019. To account for the notion that trademark registration (representing new product creation) is believed to respond more promptly to technological advancements compared to industry concentration, we opted for fixed-effects regression instead of employing long-difference regression to analyze the impact of AI adoption. In addition, in the event of a significant shift in the underlying mechanism, such as the ascendancy of deep learning (and reinforcement learning) over traditional machine learning methods, we conduct regressions on both the full sample from 2009 to 2019 (results are shown in Table 9) and the sub-sample from 2015 to 2019 (results are shown in Table10). Control variables include logged total headcount, average employee tenure, fraction of employees with high education, logged sales value, cash/assets ratio, markup, Tobin Q, market leverage, and return on assets.

VARIABLES	log(trademark)		log(cogs)		log(productivity)	
	$\beta^{top} - \beta^{nontop}$ = -0.51***		$\beta^{top} - \beta^{nontop} = 0.01$		$\beta^{top} - \beta^{nontop} = 0.04$	
	(1) Top 20	(2) Not Top 20	(3) Top 20	(4) Not Top 20	(5) Top 20	(6) Not Top 20
AI Share	-0.501*	0.055	-0.008	-0.011	0.038	-0.025
	(-1.68)	(1.01)	(-0.29)	(-0.39)	(0.37)	(-0.94)
Observations	1552	4440	3329	12169	3329	12169
R-squared	0.183	0.084	0.917	0.780	0.040	0.180
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Control	Y	Y	Y	Y	Y	Y

Table 9: Fixed Effect Regression for Trademark, Cost, and Productivity for 2009-2019

Note: The dependent variable in columns (1) and (2) is logged number of trademark filling cases, the dependent variable in columns (3) and (4) is logged costs of goods sold, the dependent variable in columns (5) and (6) is logged labor productivity. Columns (1), (3), and (5) regress on top 20 firms, columns (2), (4), (6) regress on non top 20 firms. All columns regress on a full sample from 2009 to 2019 and use the fixed effect model. Robust t-statistics are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Column 1 in Table 9 indicates a significant decrease in the trademark portfolio

VARIABLES	log(trademark)		log(cogs)		log(productivity)	
	$\beta^{top} - \beta^{nontop}$ = -0.30		$\beta^{top} - \beta^{nontop}$ = -0.10		$\beta^{top} - \beta^{nontop}$ = 0.11	
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 20	Not Top 20	Top 20	Not Top 20	Top 20	Not Top 20
AI Share	-0.066	-0.010	-0.031	-0.026	0.151	-0.033
	(-0.10)	(-0.07)	(-0.63)	(-0.81)	(1.62)	(-1.31)
Observations	676	2274	1506	6309	1506	6309
R-squared	0.002	0.072	0.881	0.699	0.014	0.158
Firm FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Control	Y	Y	Y	Y	Y	Y

Table 10: Fixed Effect Regression for Trademark, Cost, and Productivity for 2015-2019

Note: The dependent variable in columns (1) and (2) is logged number of trademark filling cases, the dependent variable in columns (3) and (4) is logged costs of goods sold, the dependent variable in columns (5) and (6) is logged labor productivity. Columns (1), (3), and (5) regress on top 20 firms, and columns (2), (4), and (6) regress on non top 20 firms. All columns regress on the sub-sample from 2015 to 2019 and use a fixed effect model. Robust t-statistics are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

from 2009 to 2019 due to AI adoption, with a 1% increase in AI share correlating to a 38.5% decrease in trademark filings, significant at the 90% confidence level. Other results do not show significance. In addition, the impact of AI adoption demonstrates significant heterogeneity between leading and non-leading firms. Surprisingly, leading firms do not gain a larger market share through new product creation; instead, this channel reveals some correlation with counterforces. One plausible explanation is that industry leaders primarily adopt AI to enhance existing products rather than create entirely new ones. This approach may be more cost-effective, as it allows them to consolidate market dominance with their top products and leverage network effects. This hypothesis is supported by observations in the literature indicating that AI excels at combining existing knowledge rather than generating new knowledge (Wu and Lynn, 2020). My empirical results also provide some support to this explanation. From 2015 to 2019, adopting AI does not significantly decrease the trademark portfolio for leading firms, and the heterogeneity was no longer significant. Such changes in estimations using the 2015-2019 sample are consistent with intuition as advanced AI technologies are better at fostering innovation. Hence, it is advisable for future studies to incorporate patent data that differentiate between new product creation and existing product improvement.

5.2 Operating Cost Savings

We will then examine whether leading firms boost their market shares by leveraging AI adoption to reduce operating costs. We gauge operating costs using Compustat's costs of goods sold (COGS). Similarly, we employed a fixed-effects model on both the full sample spanning from 2009 to 2019 and a sub-sample covering the years 2015 to 2019. Control variables include logged total headcount, average employee tenure, fraction of employees with high education, logged sales value, cash/assets ratio, Tobin Q, market leverage, and return on assets. All findings yielded insignificance, indicating no discernible relationship between AI adoption and operating costs. Furthermore, our analysis reveals no significant difference in the impact of AI adoption on operating costs between leading and non-leading firms. This suggests that cost savings may not be the factor explaining the exclusive increase in market share for leading firms following AI adoption.

5.3 Labor Productivity Enhancement

We finally examine whether prominent firms augment their market shares by leveraging the channel through which AI adoption enhances labor productivity. Labor productivity, quantified as sales per worker, is assessed using data sourced from Compustat. Similarly, we employed a fixed-effects model on both the full sample spanning from 2009 to 2019 and a sub-sample covering the years 2015 to 2019. Control variables include logged total headcount, average employee tenure, fraction of employees with high education, cash/assets ratio, markup, Tobin Q, market leverage, and return on assets. While coefficients for leading firms show positive signs and coefficients for non-leading firms show negative signs, estimations are all insignificant. There is also no significant difference in the impact of AI adoption on labor productivity between leading and non-leading firms, excluding labor productivity as the channel that drives the industry concentration.

Overall, our findings suggest that the rise in industry concentration resulting from AI adoption may be associated with performance differences in product development. Consistent with previous literature (Babina et al., 2024), we also show that AI adoption has no significant relationship with operating costs and labor productivity for both leading and non-leading firms. However, the analysis in this sector is mostly preliminary, and the reasoning is largely in the hypothesis stage due to challenges in data availability. This area will be a focus for future studies.

6 Conclusion

By employing employees' AI skills as a proxy for AI adoption, we discover a notable impact of AI adoption on the sales share of leading firms, suggesting a rise in industry concentration. In addition, we find a significant, immediate, and enduring impact of TensorFlow's launch on the sales share of leading firms, supporting the overall causal effect of AI adoption on concentration using a specific AI shock. Furthermore, in exploring the underlying drivers for the increase in concentration, we discover that the impact of AI adoption on product development varies between leading and non-leading firms, while there is no significant difference in its impact on operating costs and labor productivity.

These relationships are not necessarily static, as further advancements in AI technologies and their commercialization could revolutionize the market landscape. One potential trend is the adoption of AI, which is driving the growth of small firms, particularly start-ups. Furman and Seamans (2018) suggest that AI could lower the costs of conducting small businesses. Additionally, according to Farboodi et al. (2019), small data-savvy firms can outpace established competitors by efficiently harvesting and utilizing data, if they can sustain the investment phase.

References

[1] Zhang, Daniel, Nestor Maslej, Erik Brynjolfsson, John Etchemendy, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Michael Sellitto, Ellie Sakhaee, Yoav Shoham, Jack Clark, and Raymond Perrault. 2022. "The AI Index 2022 Annual Report." *arXiv* 2205.03468.

[2] Agrawal, Ajay, John McHale, and Alexander Oettl. 2018. "Finding needles in haystacks: Artificial intelligence and recombinant growth." *The economics of artificial intelligence: An agenda* 149–174.

[3] Crafts, Nicholas. 2021. "Artificial intelligence as a general-purpose technology: an historical perspective." *Oxford Review of Economic Policy* 37(3): 521–536.

[4] Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. 2020. "The fall of the labor share and the rise of superstar firms." *The Quarterly Journal of Economics* 135(2): 645–709.

[5] Crouzet, Nicolas and Janice C. Eberly. 2019. "Understanding Weak Capital Investment: The Role of Market Concentration and Intangibles." National Bureau of Economic Research Working Paper 25869.

[6] Ridder, Maarten De. 2024. "Market power and innovation in the intangible economy." *American Economic Review* 114(1): 199-251.

[7] Firooz, Hamid, Zheng Liu, and Yajie Wang. 2023. "Automation and the Rise of Superstar Firms." Federal Reserve Bank of San Francisco Working Paper.

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

[9] Rock, Daniel. 2019. "Engineering Value: The Returns to Technological Talent and Investments in Artificial Intelligence." Working paper.

[10] Brynjolfsson, Erik and Kristina McElheran. 2016. "The Rapid Adoption of Data-Driven Decision-Making." *American Economic Review* 106(5): 133-139.

[11] Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb. 2019. “Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction." *Journal of Economic Perspectives* 33(2): 31-50.

[12] Acemoglu, Daron, Gary W. Anderson, David N. Beede, Catherine Buffington, Eric E. Childress, Emin Dinlersoz, Lucia S. Foster, Nathan Goldschlag, et al. 2022. “Automation and the Workforce: A Firm-Level View from the 2019 Annual Business Survey." National Bureau of Economic Research Working Paper 30659.

[13] Wu, Lynn, Lorin Hitt, and Bowen Lou. 2020. “Data Analytics, Innovation, and Firm Productivity." *Management Science* 66(5): 39-57.

[14] Kumar, V., Bharath Rajan, Rajkumar Venkatesan, and Jim Lecinski. 2019. “Understanding the Role of Artificial Intelligence in Personalized Engagement Marketing." *California Management Review* 61(4): 135-155.

[15] Tanaka, Mari, Nicholas Bloom, Joel M. David, and Maiko Koga. 2020. “Firm performance and macro forecast accuracy." *California Management Review* 114: 26-41.

[16] Farboodi, Maryam, Roxana Mihet, Thomas Philippon, and Laura Veldkamp. 2019. “Big Data and Firm Dynamics." *AEA Papers and Proceedings* 109: 38-42.

[17] Furman, Jason and Robert Seamans. 2018. “AI and the Economy." National Bureau of Economic Research Working Paper 24689.

[18] Begenau, Juliane, Maryam Farboodi, Laura Veldkamp. 2018. “Big data in finance and the growth of large firms." *Journal of Monetary Economics* 97: 71-87.

[19] Fedyk, Anastassia and James Hodson. 2022. “Trading on Talent: Human Capital and Firm Performance." Working Paper 30659.

[20] Tambe, Prasanna, Lorin Hitt, Daniel Rock, and Erik Brynjolfsson. 2020. “Digital Capital and Superstar Firms." National Bureau of Economic Research Working Paper 28285.

[21] Abadi, Martín, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur,

Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. "TensorFlow: A system for large-scale machine learning." *arXiv*, 1605.08695, cs.DC.

[22] Zhang, Daniel, Saurabh Mishra, Erik Brynjolfsson, John Etchemendy, Deep Ganguli, Barbara Grosz, Terah Lyons, James Manyika, Juan Carlos Niebles, Michael Sellitto, Yoav Shoham, Jack Clark, and Raymond Perrault. 2021. "The AI Index 2021 Annual Report." *arXiv*, 2103.06312, cs.AI.