Investments and Innovation with Non-Rival Inputs: Evidence from Chinese Artificial Intelligence Startups*

Jian Xie[†]and Kang Zhou [‡]

October 2022

Abstract

Large technology firms have substantial advantages in data, a key non-rival input for developing AI technology. We argue that investments by large technology firms stimulate innovation by AI startups through the sharing of data, bringing more than money to the startups. We assemble a unique dataset containing (nearly) the universe of AI-inventing firms in China to examine the innovation effects of these investments. Our difference-in-differences estimation shows that, after receiving investments from large technology firms, AI startups increase the number of AI patent applications by 62% and the number of software products by 56%, relative to their mean values prior to the investments. Using a triple-differences strategy, we further find that the innovation impact of investments by large technology firms is stronger than that of investments by other firms without data advantages. We confirm these findings using an instrumental variables approach based on recent investments by large technology firms in peer startups. Finally, we provide novel evidence that the innovation effect works mainly through sharing non-rival data by leveraging our rich information on non-AI data-related patent applications and data-related online job postings.

Keywords: Innovation, Investments, Artificial Intelligence, Non-Rival Inputs

JEL codes: O30, G20, M13, L20, L63

*We are grateful to Manuel Bagues, Mingli Chen, Mirko Draca, Suqin Ge, Clement Imbert, Ruixue Jia, Yancheng Qiu, Yuan Ren, Yuta Suzuki, Ao Wang, Mike Waterson, David Yang, Noam Yuchtman, Mingang Zhang, Junsen Zhang, Li Zhao and especially James Fenske and Daniel Sgroi, for their helpful comments. We thank participants in the CAGE-AMES Seminar at the University of Warwick, the ACEM Doctoral Students Forum at Shanghai Jiao Tong University, and the Oxford Development Economics Workshop. All errors are our own. Jian Xie: Department of Economics, University of Warwick. Email: J.Xie.7@warwick.ac.uk; Kang Zhou: School of Economics, Zhejiang University. Email: kang_zhou@zju.edu.cn.

[†]University of Warwick

[‡]Zhejiang University

1 Introduction

As a general purpose technology, Artificial Intelligence (AI) can spur economic growth through increased productivity and product innovation across a wide range of sectors (Aghion, Jones and Jones, 2018; Acemoglu and Restrepo, 2018; Athey, 2018; Brynjolfsson, Rock and Syverson, 2018; Furman and Seamans, 2019; Mihet and Philippon, 2019).[1] Developing AI technology relies heavily on data (Furman and Seamans, 2019; Goldfarb, Gans and Agrawal, 2019; Jones and Tonetti, 2020; Acemoglu, Autor, Hazell and Restrepo, 2020; Beraja, Yang and Yuchtman, 2022), which is a non-rival input that can be used by multiple firms simultaneously. Large technology firms have accumulated enormous amounts of data from connected devices, machines, and global systems, all of which are increasing in the spread and depth of data they collect (Aghion, Jones and Jones, 2018; Bessen, Impink, Reichensperger and Seamans, 2018; Furman and Seamans, 2019). At the other end of the spectrum, many of the startups that develop new AI technology lack user-based platforms or other business lines that allow for the collection of training data.

In this paper, we examine whether investments by large technology firms foster or impede innovation by AI startups, and whether these investments are more effective at increasing innovation than investments by other firms without data advantages.[2] Globally, annual investment in AI startups has grown from less than USD 3 billion in 2012 to close to USD 75 billion in 2020.[3] However, little is known about the innovation effects of these investments. With data advantages, large technology firms could be more effective investors for startups since they bring more than money. For traditional industries that intensively use rival inputs, such as capital and labor, incumbents may absorb resources from, instead of transferring resources to, target startups (Hellmann and Puri, 2002). By contrast, for industries that rely on non-rival inputs which are shareable across firms, resource competition might be less of a concern. Investments by incumbents could generate additional gains.

However, investments by incumbents do not always benefit entrepreneurial firms. They might deter innovative startups to preempt future competition, using a tactic called "buy and kill" (Cunningham, Ederer and Ma, 2021). For large technology firms, the critical issue is whether they share their data with the startups in which they invest. Even though data is non-rival, the attraction of reinforcing advantages over new entrants (Cockburn, Henderson and Stern, 2018) and the fear of creative destruction (Jones and Tonetti, 2020) may incentivize large technology firms to keep data private. Finding an empirical answer to the question not only brings insight into the strategic objective of these incumbents, but also provides additional evidence for the political debate on whether large technology firms should be regulated.[4]

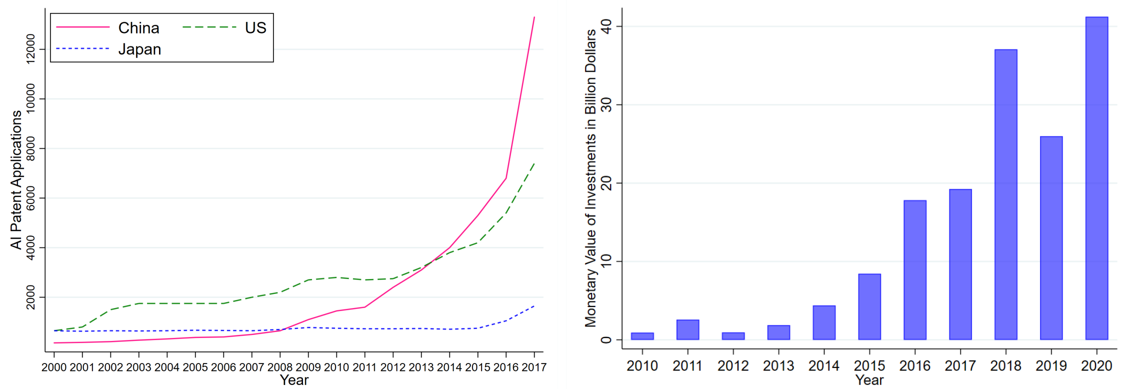
We assemble a unique dataset for our research. Our dataset covers (nearly) the universe of

[1]The Oxford English Dictionary defines Artificial Intelligence as "the theory and development of computer systems able to perform tasks normally requiring human intelligence." Acemoglu, Autor, Hazell and Restrepo (2020) define Artificial Intelligence as a collection of algorithms that act intelligently by recognizing and responding to the environment to achieve specified goals.

[2]Including other CVC (corporate venture capital) and independent VC/PE (venture capital or private equity.

[3]See statistics in (Tricot, 2021).

[4]For example, see a discussion here: https://mitsloan.mit.edu/ideas-made-to-matter/will-regulating-big-tech-stifle-innovation



(a) AI Patent Applications in Three Countries

(b) AI Investments in China

Figure 1: AI Patent Applications in Three Countries and AI Investments in China
Data source: IPO (2019) and www.itjuzi.com

AI startups in China, which refer to firms with at least one AI-related patent application and established between 2010 and 2020.[5] We collect semi-annual information on these AI startups' patent applications, software products, financing history, online job postings and registered capital, among other variables. Information at this level makes it possible to trace the technology outputs of AI startups, which are linked to the records of their external financing. We conduct the research in the context of China for two reasons. First, China had more AI-related patent applications than any other countries in recent years (IPO, 2019) and has become the largest producer of AI research in the world (Beraja, Yang and Yuchtman, 2022). Second, China accounted for nearly one-fifth of global funding of annual private investment in recent years, making it a global leader in funding commercial AI startups (Zhang, Mishra, Brynjolfsson, Etchemendy, Ganguli, Grosz, Lyons, Manyika, Niebles, Sellitto et al., 2021).

To identify the causal effect of investments by large technology firms on AI innovation, we use a staggered difference-in-differences (DID) strategy. We compare changes in technology outputs between AI startups receiving investments from large technology firms and those that do not receive such investments. Using a triple differences (DDD) strategy, we further estimate the innovation effect of type-specific investments from large technology firms and investments from firms without data advantages, including other CVC (corporate venture capital) and independent VC/PE (venture capital or private equity).

A key threat to identification is that startups selected by large technology firms may have higher potential. They could innovate more even in the absence of the investments. We address this issue in three ways. First, we control for firm fixed effects to eliminate impacts from time-invariant characteristics correlated with large technology firms' selection decisions. Second, we include a variety of time-invariant characteristics interacted with period fixed effects, such as establishment year, firm type, industry affiliation, and county, to account for innovation trends associated with these characteristics and for policy shocks. Third, we use an event study approach to show that large technology firms do not select AI startups with trends in AI

[5] It also covers AI firms established before 2010.

technology outputs.

We find that investments by large technology firms spur AI innovation, as measured by AI-related patent applications and registered software products. In magnitude, AI startups file 0.497 more AI-related patent applications and register 4.903 more software products after receiving investments from large technology firms. The number of AI-related patent applications increases by 62% and the number of software products increases by 56%, respectively, relative to their mean values prior to the investments. Despite a sharp increase in technology outputs after the receipt of investments from large technology firms, a similar pre-trend between the treatment and control startups is observed using an event study.

We use a triple-differences (DDD) strategy to estimate the innovation impact of type-specific investments. These estimations suggest that AI startups file 0.426 more AI-related patent applications and register 3.591 more software products after receiving investments from large technology firms, compared to their counterparts that received investments from firms without data advantages, including other CVC and independent VC/PE firms. These estimates represent an increase of 53% and 41%, respectively, relative to their mean values prior to the investments. This evidence further supports the strong effect of investments by large technology firms on AI startups' innovation.

We use an instrumental variables (IV) strategy to address potential concerns about omitted variables or unobserved shocks driving both AI startups' innovation and receipts of investments from large technology firms. Our IV strategy leverages the fact that investment waves vary over time, in the spirit of Acemoglu, Naidu, Restrepo and Robinson (2019) and Acemoglu, He and le Maire (2022), who exploit regional democratization waves, and waves in hiring managers with a business school degree, respectively, as instruments. The core idea is that firms may invest in similar startups within a short period due to global investment waves and sectors becoming 'hot' (Tricot, 2021). We find that the ratios of peer startups receiving investments from large technology firms in the three recent periods strongly predict the probability that a startup will receive investments from large technology firms.[6] Our IV estimates further confirm our baseline results that investments by large technology firms increase technology outputs for AI startups.

To understand why investments by large technology firms are more effective at increasing AI innovation, we investigate three potential mechanisms. First, we examine whether AI startups acquire data from large technology firms, which have substantial data advantages relative to other investors. To a large extent, who owns big data determines which firms benefit from AI (Babina, Fedyk, He and Hodson, 2021). However, it is usually unobservable when large technology firms share data with target startups, making it challenging to investigate empirically. In the spirit of Beraja, Yang and Yuchtman (2022), we measure the quantities of data each firm possesses using the number of non-AI data-related patent applications and the number of data-related online job postings. The underlying assumption is that, if these entrepreneurial firms obtain big data from large technology firms after the investment event, they should have more

[6] Startups within the same county $\times$firm type $\times$industry$\times$ size cell.

technological activities in data processing and recruit more workers to work with the shared data. Empirically, we find AI startups file more non-AI data-related patent applications and post more data-related jobs after receiving investments from large technology firms, relative to investments from other firms. Formal mediation analyses suggest that non-AI data-related patent applications and data-related jobs explain around 60% and 40% of the total effects in the increase of AI-related patent applications. And they explain around 34% and 58%, respectively, of the total effects in the increase of software products.

To ensure the two variables measure the quantity of data, instead of the general innovation capacity or firm size, we conduct two "placebo" tests using other patent (non-AI, non-data-related) applications and online job postings with low education requirements as mediators, which are less related to data processing and AI research. We find that the two variables only account for a tiny fraction of the total effects. Hence, the results of our "placebo" tests suggest that it is not the general innovation capacity or firm size that is driving up innovation by AI startups. This further confirms the data sharing mechanism, which is that investments by large technology firms facilitate the sharing of data. In addition, we examine the alternative mechanisms of whether funding itself (measured by registered capital) or government contracts lead to more AI innovation. We find little evidence to support either of the mechanisms.

We conduct a variety of robustness tests to verify our findings. First, we use the method proposed by Borusyak, Jaravel and Spiess (2021) to mitigate concern about negative weights that could lead to a biased estimation (De Chaisemartin and d'Haultfoeuille, 2020). Second, we use the number of AI-related patents granted and the number of trademarks as alternative proxies of innovation outputs, to test the robustness of our results. Third, we address the concern that our results could be driven by a endogenously selected sample by restricting the sample to startups with AI-related patent applications and without receiving investments by a certain year and examining the effects of investments that occurred after that year.[7] Fourth, we consider several other functional forms to ensure that our results are not driven by a particular function. Fifth, we control for initial AI technology outputs and prior investment rounds by other firms (if any) to further account for startups' innovation potential. Sixth, to address the concern that large technology firms could be joint investors or invest more money than other investors do, we control for joint investments (if any) and the monetary value of investments. Finally, we include county×industry×period fixed effects and county×industry×firm type×period fixed effects in specifications to absorb more specific shocks. In sum, these tests show a robust positive innovation effect of investments by large technology firms and lend additional confidence to our findings.

This paper makes three contributions. First, we reveal a new force driving AI innovation, contributing to a nascent stream of research on the economics of both AI and data. Previous work has analyzed how AI facilitates economic growth (Aghion et al., 2018; Acemoglu and Restrepo, 2018; Athey, 2018; Brynjolfsson et al., 2018; Furman and Seamans, 2019; Mihet and Philippon, 2019) and the impact of AI on labor market outcomes (Ford, 2015; David, 2015; Frey

[7] We conduct a series of tests by setting the cutoff year to be 2013, 2014,..., and 2019.

and Osborne, 2017; West, 2018; Acemoglu et al., 2020; Susskind, 2020; Grennan and Michaely, 2020; Babina et al., 2022).[8] These studies highlight the consequences of adopting AI, whereas our emphasis is on the development of AI itself. The importance of the analysis is highlighted at least in part by the intensified global race to fund, develop, and acquire AI technologies. To the best of our knowledge, we are the first to investigate the impact of external financing on innovation by AI startups. We have unique data with which we can analyze whether investments by large technology firms are more effective at increasing AI innovation than investments by firms without data advantages. Our study complements papers by Beraja, Kao, Yang and Yuchtman (2021) and Beraja, Yang and Yuchtman (2022), who examine the role of government-collected data in shaping innovation by facial recognition firms.

Second, we contribute to the literature that examines the effects of incumbents' strategic choices to cooperate or compete with entrants on entrants' innovation, including their investments in the entrants (Lerner, 2000; Hellmann and Puri, 2002; Fulghieri and Sevilir, 2009; Ivanov and Xie, 2010; Chemmanur, Loutskina and Tian, 2014; Lerner and Nanda, 2020) and more broadly, their mergers and acquisitions (Ornaghi, 2009; Guadalupe, Kuzmina and Thomas, 2012; Phillips and Zhdanov, 2013; Seru, 2014; Haucap, Rasch and Stiebale, 2019; Cunningham, Ederer and Ma, 2021). Incumbents may preempt future competition by deterring innovative startups, and this motive is confirmed by recent evidence from pharmaceutical firms (Cunningham, Ederer and Ma, 2021). Focusing on AI startups, which are frequently the target of investments or acquisitions by incumbents, we find that non-rival data could have mitigated resource competition and increased cooperation between incumbents and new entrants, which could have fuelled commercial AI innovation.

Third, we contribute to the literature that integrates production with non-rival inputs into economic analysis. Previous theoretical papers emphasizing the role of non-rival inputs in economic growth focus more on technologies (e.g., Romer, 1990; Eicher, 1996), ideas (e.g., Romer, 2015) and public goods (e.g., Barro and Sala-i Martin, 1992). Recent theoretical papers emphasize the role of data due to its increased availability and its importance in technology advances (Aghion, Jones and Jones, 2018; Veldkamp and Chung, 2019; Jones and Tonetti, 2020). However, few papers empirically examine the sharing of non-rival inputs, since quantifying non-rival inputs is a big challenge, and it is difficult to observe sharing because it is usually not revealed. We leverage our rich information on non-AI data-related patent applications and data-related online job postings, and provide novel evidence of data sharing. This sheds light on the positive relationship between sharing non-rival inputs and innovation.

The rest of this paper proceeds as follows. In Section 2, we introduce data, key variables and definitions. In Section 3, we discuss empirical strategies and estimation results. In Section 4, we examine the mechanisms underlying our results. In Section 5, we conduct a variety of robustness tests. Section 6 concludes the study.

[8] There is also growing literature that provides descriptive evidence and preliminary analysis on the innovative landscape of AI across time, sectors or space (Keisner, Raffo, Wunsch-Vincent et al., 2015; Cockburn, Henderson and Stern, 2018; Fujii and Managi, 2018; WIPO, 2019; IPO, 2019; Van Roy, Vertesy and Damioli, 2019).

2 Data and Variables

There is no universally accepted definition of AI and thus identifying AI-related patents is a challenge.[9] We combine two literature-based approaches to capture AI-related patents whilst excluding non-AI-related patents. First, we define AI-related patents on the basis of International Patent Classification (IPC) and Cooperative Patent Classification (CPC) codes, which are assigned by the Intellectual Properties Offices to distinguish patent categories. IPO (2019) provides a list of IPC and CPC codes to define AI technology areas, as shown in Table A.1. Second, we complement the sample by adding patents whose title or abstract contains one or more AI-related keywords such as machine learning, computer vision, and natural language processing. We consider 40 AI-related keywords including AI technology components, applications or fields of AI, and their synonyms as displayed in Table A.2, referring to Cockburn, Henderson and Stern (2018) and Van Roy, Vertesy and Damioli (2019).[10]

We assemble a unique dataset in three steps. First, we define AI firms and obtain a universal list of Chinese AI firms. Based on our definition, AI firms are firms with AI-related patent applications, consistent with the definitions used in IPO (2019), Alderucci, Branstetter, Hovy, Runge and Zolas (2020) and Igna, Venturini et al. (2021). We adopt this definition because it shows real research in AI and it is replicable.[11] It also ensures our focus is on AI-inventing firms, instead of AI-using firms. We use patent applications instead of granted patents because it could take several years for a patent to be granted and many factors could impact the processing time (IPO, 2019).[12] We obtain information about Chinese AI patents by using the Global Patent Database (or IncoPat), which contains the universe of Chinese patents with detailed information on titles, abstracts, applicants, and IPC and CPC codes.[13] Using the patent information on applicants, we obtain a universal list of Chinese AI firms.

Second, we collect variables of our interest of these AI firms from Tianyancha, the largest business search platform, covering the universe of Chinese companies.[14] These variables include firms' financing history, number of registered software products, online job postings, registered capital over time, government bids, trademarks, and industry affiliation, among others.[15] We focus on AI startups established between 2010 and 2020 because there were few AI patent applications prior to 2010 in China. The investments by large technology firms are mainly through their corporate venture capital. Our sample only consists of independent startups. We do not consider merged startups or acquired startups, nor other subsidiaries of these

[9]Broadly, AI consists of a diverse set of technologies and components.

[10]Our keywords are not the same as theirs since there are language differences.

[11]There are other approaches to define AI firms on the basis of their business scope (Bessen, Impink, Reichensperger and Seamans, 2018; Beraja, Yang and Yuchtman, 2022) or job postings (Babina, Fedyk, He and Hodson, 2021; Acemoglu, Autor, Hazell and Restrepo, 2020; Babina, Fedyk, He and Hodson, 2022). However, we find more than 800,000 Chinese firms describe themselves as AI firms in their business scope due to exaggeration. Many firms could tag themselves as AI firms even though they only use some AI products. Furthermore, we do not have the universe of online job postings of Chinese firms. For these reasons, we define AI firms according to their patent applications.

[12]Hence, patent application is a more accurate reflection of innovative activities for an inventing firm. Also, our results remain robust when we use the number of granted AI patents as a measure of AI innovation. See Section 5.2 for a detailed discussion.

[13]https://www.incopat.com/

[14]The website of Tianyancha is at https://m.tianyancha.com. The recent influential research by Beraja, Yang and Yuchtman (2022) also uses the platform to collect similar information for Chinese facial recognition firms.

[15]Information on online job postings is only available from 2016.

large technology firms, since independent startups and dependent startups are not comparable.

Third, we validate our key variables using multiple other independent data sources and these validations confirm the high quality of our data. Specifically, we verify the reliability of our patent data from the Global Patent Database by comparing our information with the patent information from the website of the China Intellectual Property Administration, the official website of patent registration and publication in China.[16] We also validate the reliability of firms' financing history using CVSource.[17] CVSource covers the investment events of all the active corporate venture capital, independent venture capital, and private equity firms in China (more than 11,000 in total). Finally, we validate the registered software data using the website of the Copyright Protection Center of China.[18] Chinese software is required to register on this website.

We define large technology firms as the top 50 firms making the highest number of AI-related patent applications. These large technology firms, by this definition, include Baidu, Alibaba, Tencent and Huawei.[19] We choose 50 as the threshold for three reasons. First, data advantages are highly concentrated in some large technology firms (Bessen, Impink, Reichensperger and Seamans, 2018). Second, large technology firms within the top 50 are more likely to invest in AI startups, whereas those outside the top 50 are less likely to invest. We illustrate this with a graph in Figure 2, where we plot the cumulative number of investment events in AI startups by the number of large technology firms. The vast majority of investments are made by the top 50 largest technology firms; the rest of the firms contribute little to funding AI startups. Third, these 50 large technology firms are well-established as technology firms prior to the investments. Their growth is less likely to be driven by investing in other companies. Our results are robust to alternative thresholds, such as the top 20 or top 100.[20]

Our sample consists of 9,775 AI startups in total. 314 of them have received investments from large technology firms. Figure 3 displays the distribution of the AI startups. As shown, these startups geographically disperse across the country, with a higher concentration in the Yangtze Delta and the Pearl River Delta. Table A.3 reports the summary statistics of these AI startups by group.

3 Empirical Strategies and Results

3.1 Two Group Comparison

In this section, we compare the innovation performances of AI startups that do and do not receive investments from large technology firms. This tells us about the effect of these

[16] http://epub.cnipa.gov.cn/

[17] https://www.cvsource.com.cn/

[18] https://www.ccopyright.com.cn/

[19] Foreign large technology firms, such as Google and Facebook, are not included in our sample. This is because their main business is banned in China.

[20] The estimates are displayed in Table A.12, Table A.13, Table A.14 and Table A.15.

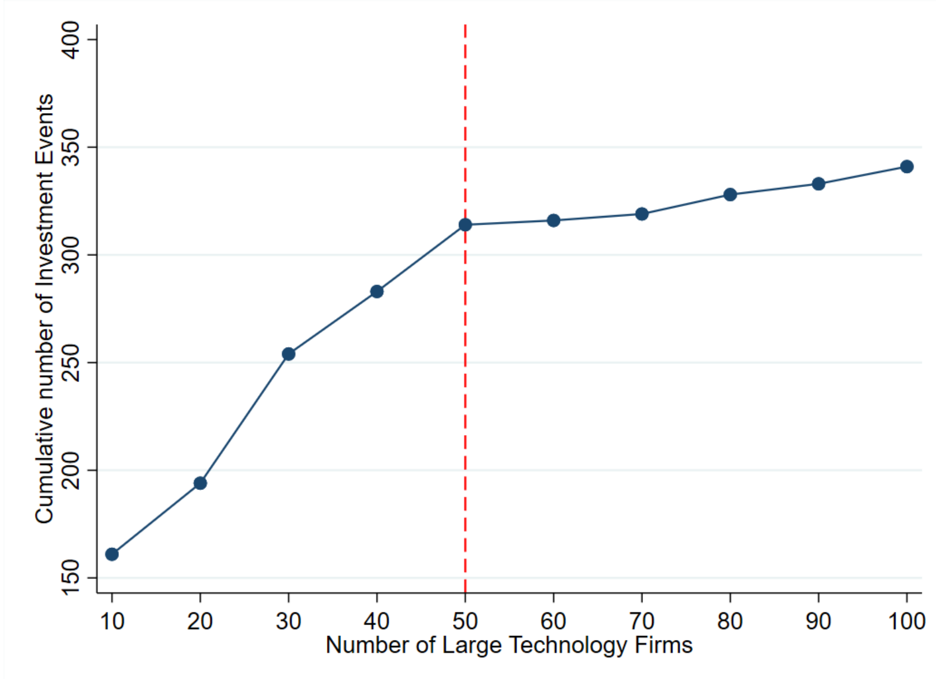


Figure 2: Cumulative Number of Investment Events in AI Startups

investments on AI startups' innovation. Our control group in this section is AI startups that did not receive external investments and that received investments from other firms, including other CVC (corporate venture capital) and independent VC/PE (venture capital or private equity).

3.1.1 Difference-in-Differences

To estimate the consequence of investments by large technology firms on innovation by AI startups, we employ the following staggered difference-in-differences approach:

$$Y_{i,t} = \beta_1 TechInvest_i \times Post_{i,t} + X_i' \times \lambda_t + \mu_i + \delta_{c,t} + \varepsilon_{i,t}, \tag{3.1}$$

where $Y_{i,t}$ refer to the cumulative number of AI-related patent applications and the cumulative number of registered software products, respectively, for firm i up to the 6-month period t, following Beraja, Yang and Yuchtman (2022). Since there were only a few applications for AI-related patents before 2010, we consider AI startups established between 2010 and 2020 (inclusive) and the outcome variables are restricted to the same period.[21] $TechInvest_i$ is a dummy indicating whether firm i ever received an investment from a large technology firm.[22] $Post_{i,t}$

[21]Our results are robust if we consider only the period of 2010-2019 to avoid the influence of Covid-19 pandemic.

[22]There are joint investments by several firms, but the majority of investments are not joint investments. We will tackle this issue

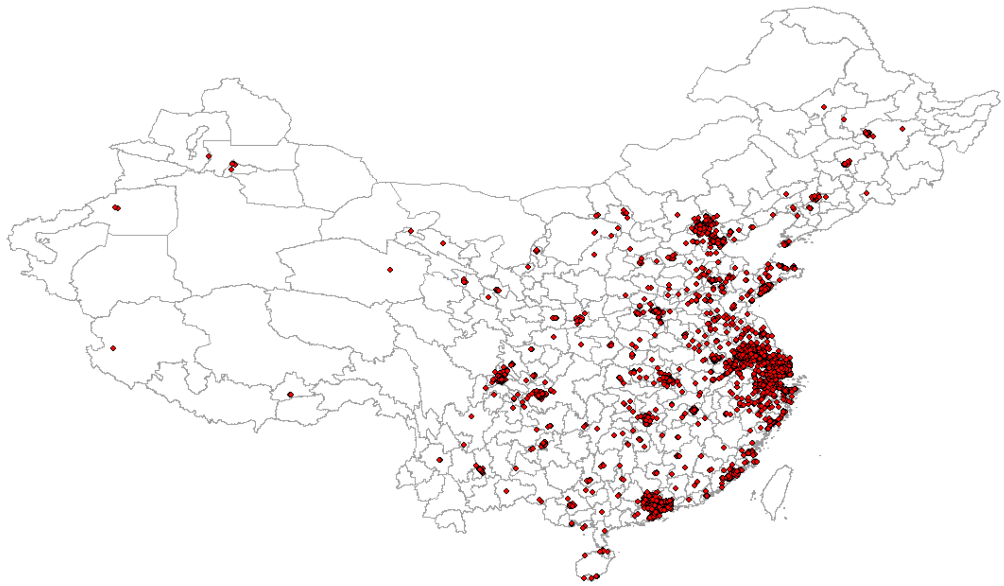


Figure 3: The Distribution of AI Startups across China

is a dummy indicating the periods for a startup after it was invested in by a large technology firm for the first time.

$X_i' \times \lambda_t$ are time-invariant firm characteristics interacted with period fixed effects, including establishment year, firm type, and industry. Firm type includes domestic private firms, state-owned firms, and firms controlled by foreign firms. We include them to account for the innovation trends associated with firm characteristics. For instance, there could be a burst of AI innovation in a certain industry over specific periods. Moreover, we always control for firm fixed effects μ_i, since there could be inherent differences between firms selected by large technology firms and those without these investments. We also include county×period ($\delta_{c,t}$) fixed effects in our preferred specification to account for time-variant county-specific shocks, such as government policies, innovation environments, and the evolving tendency to invest in specific counties. These fixed effects are very restrictive. Outcome variables are winsorized at the 1st and the 99th percentiles to eliminate the extreme values, and standard errors are clustered at the firm level.[23]

One threat to our empirical analysis is that large technology firms may select AI startups with higher innovation potential. To address this, we always account for time-invariant sources of selection by including firm fixed effects in our specification. We also control for a set of characteristics interacted with period fixed effects that could be associated with time-variant sources of selection. In addition, for robustness checks, we further control for several pre-

in the robustness analysis.

[23]Our results are robust to winsorizing outcome variables at the 5th and 95th percentiles.

investment factors interacted with period fixed effects to measure their innovation potential, including the initial number of AI patent applications, the initial number of software products, and the prior number of investment rounds made by other firms (if any). We will use an event study approach to test whether large technology firms select startups with pre-trends in AI technology outputs. To address the possible omitted variable bias, we employ an instrumental variables strategy based on recent investments in peer startups in Section 3.1.3. To mitigate concern about negative weights in two-way fixed effects estimation, we use the method proposed by Borusyak et al. (2021) in Section 5.1.

Table 1 displays the coefficients of interest. In column (1) and column (2), the outcome variable is the cumulative number of AI-related patent applications. In column (3) and column (4), the outcome variable is the cumulative number of registered software products. Both are used to measure AI innovation. Odd columns control only for period fixed effects and firm fixed effects. Even columns additionally control for a full set of time-invariant firm characteristics interacted with the period fixed effects and county×period fixed effects, which are our preferred specifications.

Across specifications, we find a significant increase in AI technology outputs in response to investments by large technology firms. In magnitude, the receipt of an investment from a large technology firm leads to an additional 0.497 AI patent applications and an additional 4.903 software products after the investment. These are large magnitudes, since they represent an increase of 62% (0.497/0.80) and 56% (4.903/8.83), respectively, compared to their mean values prior to the investments.[24] In brief, these results suggest that firms with data advantages exert a strong influence on AI startups' innovation.

Table 1: DID: AI Patent Applications and Software Products

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.699***	0.497***	7.172***	4.903***
	(0.115)	(0.120)	(1.141)	(1.097)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
R^2	0.652	0.693	0.691	0.745
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

[24] They represent an increase of 54% (0.497/0.926) and 65% (4.903/7.574), respectively, compared to the means of the outcome variables.

3.1.2 Event Study

Large technology firms might select AI startups with higher innovation potential to invest in, on the basis of their AI technology outputs prior to the investment. Therefore, the positive effects of the investments could reflect the superior ability of large technology firms to identify startups with higher innovation potential. If this were the case, the target startups would innovate more even in the absence of investments from large technology firms. To test for this possibility, we employ the following event study method:

$$Y_{i,t} = \sum_{T} \beta_{1t} TechInvest_i \times Period_{i,t} + X_i' \times \lambda_t + \mu_i + \delta_{c,t} + \varepsilon_{i,t} \tag{3.2}$$

where $Period_{i,t}$ refers to 6-month period dummies before or since the first period of receiving the investment from a large technology firm (the period prior to the first investment is omitted as a comparison). Other variables are the same as those in Equation 3.1. Outcome variables are winsorized at the 1st and the 99th percentiles. β_{1t} captures the change in differences in AI technology outputs between AI startups with and without investments by large technology firms over time. Figure 4 and Figure 5 plot the coefficients of interest, respectively, with different outcome variables. We find few pre-trends prior to the investments. Yet, the AI technology outputs increased substantially after they received investments from large technology firms and remain at higher levels 8 periods (or four years) later. This pattern is consistent with the possibility that large technology firms contribute to AI startups' technology outputs, instead of selecting AI startups with pre-existing positive trends in AI patent applications.

3.1.3 Instrumental Variables

Another endogeneity concern is that our empirical results may be driven by omitted variables at the firm level, such as startup quality or concurrent organizational changes. There could be an adverse selection problem associated with startup quality: AI startups with sufficient funds and high potential might reject investments from large technology firms. Entrepreneurs will be less willing to sell their equity shares if they know these shares are more valuable than market expectations. On the other hand, lower-quality AI startups have more incentives to send out good signals to large technology firms. If an adverse selection problem is affecting the results, our OLS approach will underestimate the innovation effects of investments by large technology firms. However, some organizational changes could lead to an overestimated OLS result. For example, appointments of capable CEOs could confound our estimates if the new CEOs are capable of both attracting investments from large technology firms and fostering innovation at the same time. To mitigate the endogeneity concern, we employ an IV strategy that leverages the fact that investment waves vary from time to time, which could be

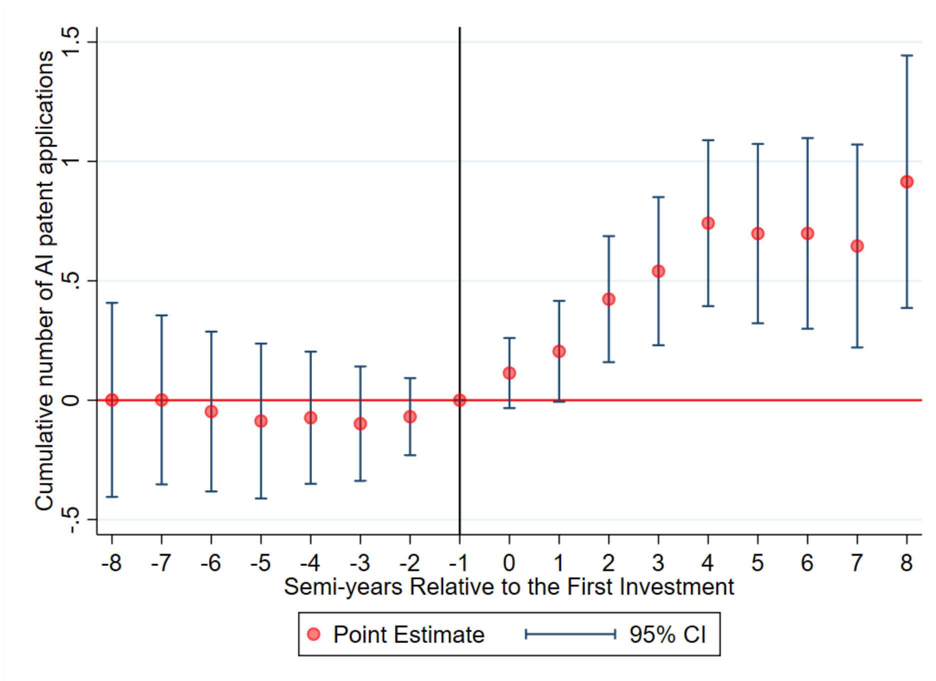


Figure 4: Number of AI Patent Applications

stimulated by exogenous shocks.[25] As a result of the change in investment waves, large technology firms could invest in peer startups within a short period. Similar ideas can be found in Acemoglu, Naidu, Restrepo and Robinson (2019) and Acemoglu, He and le Maire (2022), who exploit regional democratization waves and waves in hiring managers with a business school degree, respectively, as the instrumental variables.

In our context, we define startups within the same county $\times$firm type $\times$industry$\times$ size cell as analogous startups, where the size is measured by the quantiles of their registered capital. The first-stage equation of our IV estimates is:

$$TechInvest_i \times Post_{i,t} = \sum_{k=1}^{3} \beta_k RatioTechInvest_{i,t-k} + X_i' \times \lambda_t + \mu_i + \delta_{c,t} + \varepsilon_{i,t}, \tag{3.3}$$

where $RatioTechInvest_{i,t-k}$ indicates the jackknifed ratio of startups within the same county $\times$firm type $\times$industry$\times$size cell that have received an investment from a large technology firm in period $t-k$.[26] Following Acemoglu, He and le Maire (2022), we consider three prior periods (1.5 years) since there could be lags in the influence of investment waves.[27]

As one can see in Table 2, in the recent 3 periods, if the ratio of peer startups receiving

[25]Such as AlphaGo beating Lee Sedol in 2016. See the following news from Futurism: https://futurism.com/south-korean-government-announces-nearly-1-billion-ai-funding

[26]We exclude firm i when we compute this ratio for firm i.

[27]Our results are robust when we only consider one or two prior periods.

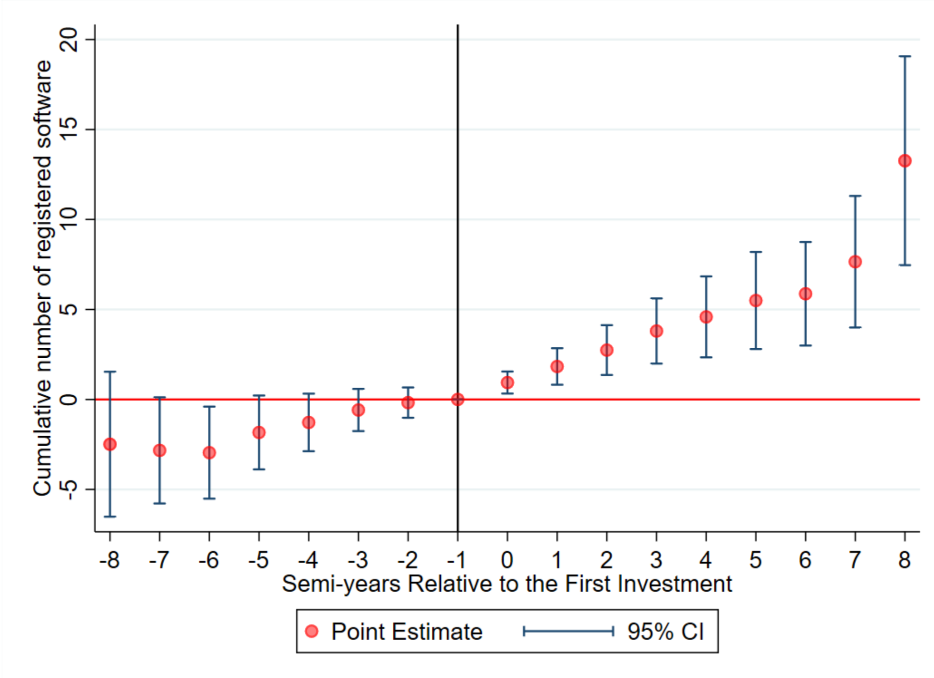


Figure 5: Number of Registered Software Products

investments from large technology firms is higher, firm i is more likely to receive an investment from a large technology firm in this period. The first-stage results are strong across specifications with KP F-statistics greater than 77. For the second-stage results, we still find that AI startups receiving investments from large technology firms delivered a larger increase in AI technology outputs, compared to their peers without these investments. Since we have three instruments and only one endogenous variable, we perform over-identification tests. Hansen J statistics are not statistically different from zero across columns, suggesting that our instruments are appropriately uncorrelated with the error terms. The null hypothesis that the instruments are valid is satisfied.

One threat to the IV strategy is that confounders associated with investment waves in peer startups might impact an AI startup's innovation. For instance, if the government initiates a policy benefiting certain types of AI startups, it could stimulate an investment wave in these startups and facilitate these startups' AI technology outputs simultaneously. If this is the case, our inclusion of county$\times$period fixed effects in specifications can partially relieve the concern, since it can help to absorb policy shocks at the county level. To further address this concern, we include county$\times$industry$\times$period fixed effects and county$\times$firm type$\times$period fixed effects, separately, in our specifications. These fixed effects account for policy shocks and changing economic environments at a more granular level. As reported in Table A.4, the estimates of interest are close to our baseline IV results. This suggests that our IV results are not driven by policy shocks.

In magnitude, the IV estimates are 2.8 to 3.6 times as large as the OLS estimates. The first reason could be adverse selection: conditional on market expectations, some AI startups with higher quality are less willing to receive investments from large technology firms, whereas some of their counterparts with lower quality are eager to sell equity shares. Omitted variables in startup quality could lead to an underestimated OLS result. A second reason might be local average treatment effects: a startup selected by large technology firms in an investment wave might innovate more in response to the investment because effective techniques used to promote peer startups' innovation could also be applied to this startup.

Table 2: IV: AI Patent Applications and Software Products

Second-Stage Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	2.164***	1.807***	17.342***	13.790***
	(0.479)	(0.473)	(3.331)	(3.260)
First-Stage Dependent Var.:	TechInvest×Post			
	(1)	(2)	(3)	(4)
RatioTechInvest t-1	0.358***	0.346***	0.358***	0.346***
	(0.039)	(0.039)	(0.039)	(0.039)
RatioTechInvest t-2	0.133***	0.133***	0.133***	0.133***
	(0.036)	(0.036)	(0.036)	(0.036)
RatioTechInvest t-3	0.282***	0.256***	0.282***	0.256***
	(0.049)	(0.049)	(0.049)	(0.049)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	97162	92184	97162	92184
R^2	0.707	0.725	0.707	0.725
KP F-statistics	88.7	77.7	88.7	77.7
Hansen J-statistics	2.632	3.676	0.182	0.476
(p-values)	(0.268)	(0.159)	(0.913)	(0.788)

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

3.2 Three Group Comparison

Our difference-in-differences estimates suggest that the receipt of investments from large technology firms predicts more AI technology outputs. However, other types of investments could also increase in AI patent applications and software products. In this part, we further compare the innovation impact of investments from large technology firms and those from other firms by examining the innovation performances of three groups of AI startups: those invested in by large technology firms, those invested in by other firms including other CVC and independent VC/PE, and those not receiving external investments. The comparison between investments by large technology firms and by other firms generates additional insight into whether large technology firms bring more than money to the startups.

3.2.1 Triple Differences

To investigate whether investments by large technology firms are more effective at increasing AI innovation compared to investments by other firms, we employ the following triple differences model:

$$Y_{i,t} = \beta_1 Invest_i \times TechInvest_i \times Post_{i,t} + \beta_2 Invest_i \times Post_{i,t} + X_i' \times \lambda_t + \mu_i + \delta_{c,t} + \varepsilon_{i,t}, \quad (3.4)$$

where $Invest_i$ is a dummy indicating whether firm i was ever invested in by a large technology firm or other firms. $TechInvest_i$ is a dummy indicating whether firm i was ever invested in by a large technology firm. We include the same set of fixed effects as in Equation 3.1. $\beta_1 + \beta_2$ captures the effects of the investments by large technology firms, β_2 captures the effects of investments by other firms. Therefore, β_1 non-parametrically estimates the additional advantage caused by investments by large technology firms, relative to investments by other firms.

Table 3 shows the estimates of the triple differences model. The coefficients of the triple interaction terms are positive and significant across columns, which suggests that investments by large technology firms are more effective at fostering innovation by AI startups, relative to investments by other firms (other CVC and independent VC/PE). The magnitude is economically important. Based on our preferred specification, investments by large technology firms result in 0.426 more AI patent applications and 3.591 more software products following the event, compared to investments by other firms. They represent an increase of 53% (0.426/0.80) and 41% (3.591/8.83), respectively, compared to their mean values prior to the investments. The coefficients on $Invest_i \times \text{Post}_{i,t}$ are also positive and significant. These results show that investments by other firms also promote AI innovation, although the effect is not as pronounced as the effect achieved by large technology firms, consistent with our argument that large technology firms bring more than money to the AI startups they invest in.

3.2.2 Instrumental Variables in the Triple Differencess

To address the endogeneity problem caused by omitted variables in the estimation of the triple differences, we use the same instrumental variables as in Section 3.1.3 to predict post-periods of investments by large technology firms. The first-stage equation is:

$$\begin{aligned} Invest_i \times TechInvest_i \times Post_{i,t} = \sum_{k=1}^{3} \beta_k Invest_i \times RatioTechInvest_{i,t-k} + \quad (3.5) \\ \beta_4 Invest \times Post_{i,t} + X_i' \times \lambda_t + \mu_i + \delta_{c,t} + \varepsilon_{i,t} \end{aligned}$$

$RatioTechInvest_{i,t-k}$ is the jackknifed ratio of startups in the same county×firm type×industry×size

Table 3: DDD: AI Patent Applications and Software Products

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.582***	0.426***	5.208***	3.591***
	(0.117)	(0.121)	(1.195)	(1.123)
Invest×Post	0.320***	0.230***	5.391***	4.253***
	(0.040)	(0.041)	(0.436)	(0.438)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
R^2	0.653	0.694	0.695	0.748
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

cell as firm i that have received investments from large technology firms in period $t-k$. We use $RatioTechInvest_{i,t-k}$ to predict $TechInvest_i \times Post_{i,t}$. The intuition is the same as Section 3.1.3: if the ratio of peer AI startups receiving investments from large technology firms is higher in the recent three periods, then firm i would have a higher probability of receiving this kind of investment this period, due to investment waves.

Table 4 displays the IV estimates in the triple differences. The first-stage results show that $Invest \times RatioTechInvest_{i,t-k}$ strongly predicts the triple interaction terms with KP F-statistics higher than 85 across the columns. The second-stage results confirm that large technology firms bring additional advantages to AI startups in their innovation, relative to other investors. Our over-identification tests (Hansen J-statistics reported in Table 4) suggest that we do not violate the null hypothesis. The instruments are independent of the error process, supporting their validity.

Analogous to Section 3.1.3, we tackle the threat of violating the exclusion restrictions by adding county×industry×period fixed effects and county×firm type×period fixed effects. These fixed effects absorb more policy shocks and other economic changes that could lead to more AI innovation. Table A.5 displays the estimates with these fixed effects. It shows that the coefficients of interest are close to the baseline IV estimates. Similar to the analysis in Section 3.1.3, the increased magnitude in IV estimates relative to OLS estimates could be caused by adverse selection or local average treatment effects.

Table 4: IV in DDD: AI Patent Applications and Software Products

Second-Stage Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	1.982***	1.709***	10.885***	9.181***
	(0.468)	(0.449)	(3.142)	(2.904)
Invest×Post	0.257***	0.185***	5.380***	4.107***
	(0.044)	(0.042)	(0.440)	(0.424)
First-Stage Dependent Var.:	Invest×TechInvest×Post			
	(1)	(2)	(3)	(4)
Invest×RatioTechInvest t-1	0.403***	0.400***	0.403***	0.400***
	(0.042)	(0.043)	(0.042)	(0.043)
Invest×RatioTechInvest t-2	0.155***	0.163***	0.155***	0.163***
	(0.042)	(0.044)	(0.042)	(0.044)
Invest×RatioTechInvest t-3	0.309***	0.293***	0.309***	0.293***
	(0.057)	(0.057)	(0.057)	(0.057)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	97162	92184	97162	92184
R^2	0.717	0.733	0.717	0.733
KP F-statistics	95.9	85.4	95.9	85.4
Hansen J-statistics	1.795	2.589	1.015	1.008
(p-values)	(0.408)	(0.274)	(0.602)	(0.604)

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

4 Mechanisms

4.1 Non-Rival Inputs: Data

Large technology firms have substantial advantages in data (Aghion, Jones and Jones, 2018; Bessen, Impink, Reichensperger and Seamans, 2018; Furman and Seamans, 2019). The importance of data access for AI innovation has been widely documented in the literature (Furman and Seamans, 2019; Goldfarb, Gans and Agrawal, 2019; Jones and Tonetti, 2020; Acemoglu, Autor, Hazell and Restrepo, 2020; Beraja, Yang and Yuchtman, 2022). For AI startups, connections to large technology firms could bring more proprietary data for their innovation process. There are anecdotes about large technology firms sharing data with their closely connected startups.[28] As a non-rival input, data could be used simultaneously across many firms to train algorithms for AI products.

It is challenging to measure empirically the amount of data each firm possesses, since this is usually not revealed. AI firms could obtain their data from product users, mobile application (APP) users, websites of firms, governments, and various organizations. Beraja, Yang and Yuchtman (2022) use the number of data-complementary (non-AI) software products that foster data transmission and data management to measure the quantities of data. In a similar approach, we employ two proxies to measure the quantity of data for each firm: non-AI data-related patent applications and data-related online job postings. For the second proxy, we only have observations from 2016 onwards.

We start by identifying non-AI data-related patents and data-related jobs and then test the data sharing mechanism. We define non-AI data-related patents as patents relevant to data processing and data management based on descriptions of IPC and CPC codes. We list these codes in Table A.6 in the Appendix. AI-related patents are excluded. Similarly, data-related jobs are jobs analyzing or managing data, such as data analyst or data engineer. We search for a list of relevant keywords such as big data, data analysis and data management, in the titles of jobs posted online. If its title contains at least a keyword, then the job is defined as data-related. The complete list of these keywords is in Table A.7 in the Appendix. The underlying assumption is, if a startup obtains more data from a large technology firm after receiving its investment, then the startup will have more data-related patent applications and data-related online job postings, compared to a startup invested in by other firms without the same data advantages.

We test this mechanism on the basis of Equation 3.1. In the first two columns of Table 5, we replace the previous outcome variables with the cumulative number of non-AI data-related patent applications. Startups invested in by data-rich large technology firms have a larger increase in non-AI data-related patent applications relative to those invested in by other firms, both when we include time-invariant firm characteristics interacted with period fixed effects

[28] For example, there are two media reports regarding Alibaba's and eBay's data sharing with their connected startups:
hbr.org/2018/09/alibaba-and-the-future-of-business
www.sohu.com/a/208725444_692906

and when we do not. From column (3) to column (6), the regressions are based on those in Table 1, but we additionally control for the cumulative number of non-AI data-related patent applications as a mediator. We can observe a reduction in the magnitude of coefficients across columns. Specifically, a formal mediation analysis proposed by Imai, Keele and Yamamoto (2010) suggests that the mediation effects account for around 60% of the total effects when the outcome variable is the cumulative number of AI-related patent applications, and around 34% of the total effects when the outcome variable is the cumulative number of software products.[29] These results demonstrate that there should be data sharing after the investments.

We conduct a similar mediation analysis using data-related online job postings. In the first two columns of Table 6 and Table 7, the outcome variable is the cumulative number of data-related jobs posted online. It shows that receiving an investment from a large technology firm leads to more data-related job postings, compared to investments by other firms. Since the data of online job postings is only available from 2016, we reestimate Equation 3.1 using the sample from 2016. Column (3) and column (4) of Table 6 display the results using the cumulative number of AI patent applications as the outcome variable and the corresponding columns of Table 7 show the results using the cumulative number of registered software products as the outcome variable. In the last two columns of these two tables, we additionally control for the cumulative number of data-related online job postings as a mediator. The mediation analysis suggests that the mediator explains around 40% of the total effect in Table 6 and around 60% of the total effect in Table 7. It further improves the possibility that large technology firms share data with the AI startups they invested in.

One might worry that these two proxies could measure other firm characteristics instead of the quantity of data a firm possesses, such as the general innovation capacity or the size of startups. An AI startup at expansion will apply for more patents and post more jobs, including data-related patents and data-related jobs, and it is likely to have more AI technology outputs at the same time. To allay this concern, we conduct two "placebo" tests respectively for these two proxies. First, we use non-AI, non-data-related patent applications as a "placebo" for non-AI, data-related patent applications. If data-related patents only capture the general innovation capacity instead of reflecting the quantities of data, then other patents would also show strong explanatory power as a mediator. We do not expect other patents to play no role, because many AI firms have multiple business lines. A firm's innovation in other technologies is related to its innovation in AI due to spillover effects. Second, we use jobs with low education requirements (high school degree or below) as a "placebo" for data-related jobs. This is motivated by the fact that 88% of data-related jobs require at least a bachelor's degree, and many jobs with low education requirements are necessary jobs whose growth in the number of positions is associated with the growth of the firm. However, these jobs are orthogonal to AI research, which is mainly conducted by workers with at least an undergraduate degree (Babina, Fedyk, He and Hodson, 2022).[30]

[29]We use the stata code written by Hicks and Tingley (2011). This method is used in the literature, such as Doyle (2020), Andrabi, Daniels and Das (2021), and Fenske, Gupta and Yuan (2022).

[30]Non-data-related jobs with high education requirements could also contribute to AI innovation.

Table A.8 displays the estimates using the cumulative number of other patent applications as the mediator. The first two columns show that investments by large technology firms increase non-AI, non-data-related patent applications by AI startups, relative to other investments. This could stem from the spillover effects of AI innovation and large technology firms' expertise in other high technologies. It also is consistent with the findings in Cockburn, Henderson and Stern (2018), who point out that AI will foster other innovation due to faster accumulation of knowledge. Column (3) to column (6) are analogous to those in Table 1, but we additionally control for the cumulative number of other patent applications. Around 16% of the total effects are mediated when the outcome variable is the cumulative number of AI patent applications and around 9% when the outcome variable is the cumulative number of software products. As a comparison, the mediation effects of data-related patents are more than three times as large as the mediation effects of other patents. This evidence confirms that data is crucial for AI technology outputs.

Table A.9 shows the results using the cumulative number of online job postings with low education requirements as the mediator. As we can observe in column (1) and column (2), investments by large technology firms bring more lower-quality jobs, although it is not statistically significant. From column (3) to column (6), we can see that lower-quality jobs basically play no role in AI innovation. The mediation effects are close to zero across specifications. This evidence boosts our confidence that data-related patents and data-related jobs capture the quantities of data, and that having a larger input of data increases startups' AI technology outputs, after they build connections with data-rich large technology firms.

Table 5: Mechanism: Data-related Patent Applications

Dependent Var.:	Data Patent Applications		AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)	(5)	(6)
Invest×TechInvest×Post	4.867***	3.592***	0.229**	0.173*	3.413***	2.388**
	(0.911)	(0.935)	(0.102)	(0.103)	(1.115)	(1.055)
Invest×Post	0.817***	0.155	0.261***	0.219***	5.089***	4.201***
	(0.283)	(0.307)	(0.035)	(0.036)	(0.428)	(0.426)
Data Patent Applications			0.072***	0.071***	0.369***	0.335***
			(0.002)	(0.002)	(0.022)	(0.022)
County×Period FE	No	Yes	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes	No	Yes
Period FE	Yes	No	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Mediation Effect			0.350	0.251	1.781	1.191
Direct Effect			0.226	0.169	3.375	2.353
Proportion Mediated			60.6%	59.5%	34.6%	33.7%
Observations	100401	94625	100401	94625	100401	94625
R^2	0.664	0.715	0.727	0.754	0.720	0.765
Mean Dependent Var.:	3.456	3.456	0.926	0.926	7.574	7.574

Notes: The first outcome variable is the cumulative number of non-AI data-related patent applications. Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table 6: Mechanism: Data-related Online Job Postings

Dependent Var.:	Data Jobs		AI Patent Applications		AI Patent Applications	
	(1)	(2)	(3)	(4)	(5)	(6)
Invest×TechInvest×Post	4.187***	2.951***	0.888***	0.697***	0.593***	0.503***
	(0.562)	(0.573)	(0.146)	(0.150)	(0.137)	(0.140)
Invest×Post	1.891***	1.516***	0.447***	0.302***	0.314***	0.202***
	(0.184)	(0.188)	(0.050)	(0.052)	(0.048)	(0.050)
Data Jobs					0.070***	0.066***
					(0.004)	(0.004)
County×Period FE	No	Yes	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes	No	Yes
Period FE	Yes	No	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Mediation Effect					0.319	0.224
Direct Effect					0.431	0.352
Proportion Mediated					42.5%	39.0%
Observations	78596	75188	78596	75188	78596	75188
R^2	0.762	0.795	0.705	0.735	0.718	0.744
Mean Dependent Var.:	1.951	1.951	1.153	1.153	1.153	1.153

Notes: The first outcome variable is the cumulative number of data-related online job postings. Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table 7: Mechanism: Data-related Online Job Postings

Dependent Var.:	Data Jobs		Software Products		Software Products	
	(1)	(2)	(3)	(4)	(5)	(6)
Invest×TechInvest×Post	4.187***	2.951***	3.997***	2.535***	1.667**	1.153
	(0.562)	(0.573)	(0.879)	(0.873)	(0.842)	(0.839)
Invest×Post	1.891***	1.516***	3.902***	3.388***	2.850***	2.678***
	(0.184)	(0.188)	(0.364)	(0.383)	(0.364)	(0.381)
Data Jobs					0.557***	0.468***
					(0.028)	(0.029)
County×Period FE	No	Yes	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes	No	Yes
Period FE	Yes	No	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Mediation Effect					3.030	1.885
Direct Effect					1.854	1.374
Proportion Mediated					62.3%	57.9%
Observations	78596	75188	78596	75188	78596	75188
R^2	0.762	0.795	0.825	0.847	0.835	0.853
Mean Dependent Var.:	1.951	1.951	9.014	9.014	9.014	9.014

Notes: The first outcome variable is the cumulative number of data-related online job postings. Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

4.2 Potential Alternative Mechanisms

4.2.1 The Growth of the Startup

Do large technology firms promote the growth of AI startups by providing more money than other investors, thus stimulating more AI innovation? We examine this possibility using the time-variant registered capital of startups in our mediation analysis. As a firm grows, it is required to update its registered capital in the State Administration of Industry and Commerce in China.[31] If large technology firms invest larger amounts of money and if this is the main driver of more AI innovation, we would expect registered capital to increase more and it will explain a considerable fraction of the total effect.

As shown in Table A.10, the first two columns indicate that large technology firms facilitate higher growth of AI startups that they invest in, compared to the impact that other types of investors have in AI startups. However, the mediation effect of the registered capital only accounts for around 2% and 4% of the total effects respectively for the two outcome variables, which is weak compared to the mediation effect of data. It demonstrates that large technology firms provide more money and foster the growth of AI startups, but this is not the main reason for their sharp increase in AI innovation.

4.2.2 Government Contracts

Beraja, Yang and Yuchtman (2022) suggest that government contracts are important for product innovation by Chinese facial recognition firms, due to access to government data. If large technology firms help AI startups earn government contracts, it could also fuel their innovation, since Chinese local government possesses huge amounts of data. Ex-ante, it is ambiguous whether large technology firms could facilitate the connections between AI startups and local government. On the one hand, endorsement from large technology firms might enhance AI startups' credibility to undertake government projects. On the other hand, the vast majority of large technology firms are not state-owned. Other state-owned investors might be better intermediaries to link startups to government contracts.

We formally examine this potential mechanism on the basis of Equation 3.1. The first two columns of Table A.11 display the estimates when we use the cumulative number of government contracts as the outcome variable. It shows that investments by large technology firms lead to fewer government contracts, compared to investments by other firms. The formal mediation analysis from column (3) to column (6) suggests that the number of government contracts plays a limited role as a mediator. This finding rules out the possibility that large technology firms promote innovation by AI startups through government data. Large technology firms are not better at helping AI startups to earn government contracts, compared to other investors.

[31] The update could be lagged for a year.

5 Robustness Analysis

5.1 Negative Weights

We use a staggered difference-in-differences with two-way fixed effects including multiple firms and periods, which means that negative weights could bias our estimation. According to De Chaisemartin and d'Haultfoeuille (2020), this method estimates weighted sums of the treatment effect in each firm and period, and if the weights are negative, the estimates could be biased, even with a sign opposite to the real treatment effect. To address the possible spurious identification (if any) caused by negative weights, we employ the method proposed by Borusyak, Jaravel and Spiess (2021). Specifically, we proceed in three steps. First, we estimate the unit fixed effects, time fixed effects, and the coefficients of other control variables using the non-treated observations.[32] Second, we compute the treatment effect for each treated observation by extrapolating the model from the first step to treated observations. Third, we further compute the average treatment effect on the treated observations based on the estimates in the second step.

We present our results from re-estimating Equation 3.1 using this method in Table A.16 and results from re-estimating Equation 3.4 in Table A.17. The re-calculated estimates are less vulnerable to potential bias from negative weights. In both difference-in-differences and triple differences models, the new estimates are positive and significant and the magnitudes are fairly close to those of the corresponding baseline estimates, suggesting that negative weights are not a big concern in our estimation.

5.2 Other Measures of AI Innovation

We use AI patent applications to construct our key outcome variable for measuring innovation, due to the inherent lag in processing patent applications and the fact that the vast majority of AI innovation has occurred in recent years. To test the robustness of our results, we use the cumulative number of granted AI patents to construct an alternative measure of the outcome variable. As an alternative measure of product innovation, we use trademarks to test the robustness of results using software products as an outcome variable.[33] Trademarks as firms' intellectual property assets allow for differentiation of a firm's products from others and some studies have used them as an indicator of product innovation (Mendonça, Pereira and Godinho, 2004; Faurel, Li, Shanthikumar and Teoh, 2015; Heath and Mace, 2020; Babina, Fedyk, He and Hodson, 2021).

As shown in Table A.18, the DID estimates suggest that receiving investments from large technology firms leads to a significantly larger increase in the cumulative number of granted patents and trademarks. The DDD estimates in Table A.19 provide further evidence that the effect of investments by large technology firms is more pronounced, relative to investment by

[32] Never-treated or not-yet-treated.

[33] We cannot distinguish AI trademarks and non-AI trademarks.

other types of firms. The magnitudes are economically important compared to mean outcome variables, consistent with our baseline results.

5.3 Definitions of AI Startups

In our baseline definition, AI startups are defined as firms with at least one AI-related patent application and established between 2010 and 2020. This definition ensures the coverage of all AI-inventing startups over the period. Since the definition of AI startups is related to one of our outcome variables, estimation bias caused by the endogenous inclusion of AI startups is a concern. For example, some startups might be included in our sample because they receive investments from large technology firms. Without these investments, they might have no AI-related patent applications and ought to be excluded from the sample.

To allay this concern we test an alternative definition. We include only AI startups with at least one AI-related patent application and that did not receive an investment from a large technology firm by a certain year. This limits our enquiry to the effects of the investment at the intensive margin: conditional on the startups that already have AI-related patent applications by a certain year, whether the investments by large technology firms in a later year are more effective at increasing innovation. We first set the cutoff year at 2013 because there are few observations prior to 2013, due to few AI patent applications. In this way, we only consider startups with at least one AI-related patent application and without receiving investments from large technology firms by 2013. Then we use different cut off years from 2014, 2015,..., until 2019.[34]

The DID estimates and the DDD estimates are displayed in Table A.20 and Table A.21 respectively. Figure 6 plots the DID estimates with different cut off years. The solid red line represents the baseline DID estimates in Table 1. As one can see, as we change the cutoff year, the estimates are either close to or higher than our baseline estimate. When the cut off year is prior to 2017, the coefficients are positive but not significant in some scenarios, which is mainly due to a large confidence interval caused by reduced observations. For instance, for 2013 we only have 1,207 observations, since there are only a small number of AI startups in that year, which represent a tiny fraction of the baseline observations (94,625). Figure 7 shows an analogous pattern of DDD estimates. This evidence indicates that among a pre-determined sample of AI startups, those receiving investments from large technology firms innovate more, relative to those receiving investments from other firms. This is consistent with our baseline results.

5.4 Functional Forms

Our baseline estimation uses the cumulative number of AI technology outputs up to the end of each period as outcome variables, which better exploits the development of startups'

[34] There is a trade-off to set the cutoff year: an earlier cutoff year is associated with fewer AI startups since AI patent applications increase dramatically in recent years; a later cutoff year is associated with a shorter post-investment period.

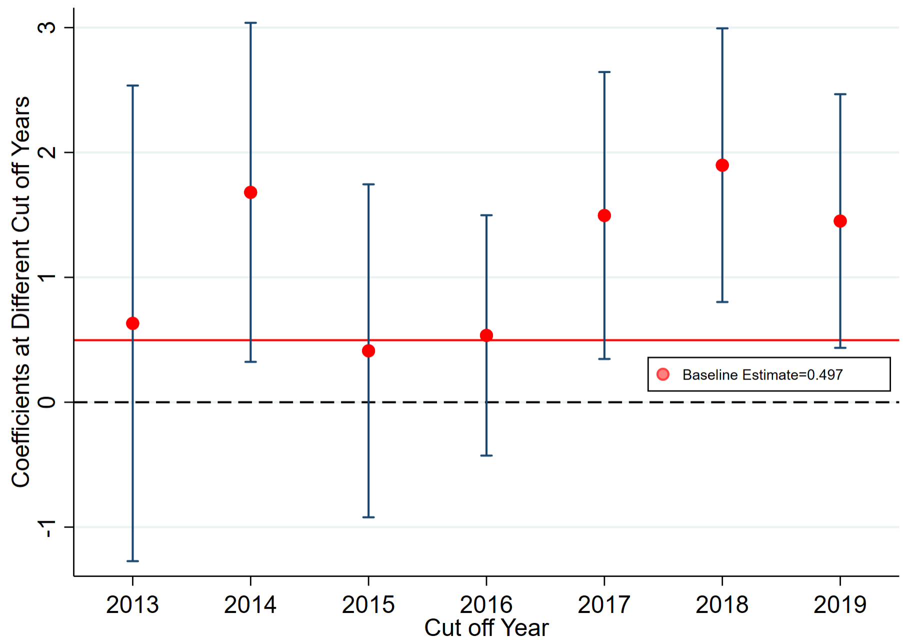


Figure 6: DID Coefficients at Different Cut off Years

Note: The solid red line marks the baseline DID estimate (0.497).

innovation outputs. To mitigate the concern that our results might be sensitive to particular functional forms, we conduct two sets of robustness checks. First, we transform the outcome variables based on three common functional forms: the flow number of AI-related patent applications and software products, outcome variables in logarithms (plus one), and the inverse hyperbolic sine transformation. Second, since the outcome variables are count variables, we also conduct additional analyses using the negative binomial regression and the Poisson regression, which are commonly employed to tackle count variables. For each functional form, we report the DID estimates and the DDD estimates respectively.

Table A.22 and Table A.23 show the coefficients of interest when the outcome variables are the flow number of AI patent applications and the flow number of registered software products. Table A.24 and Table A.25 display the estimates when the outcome variables are in logarithm (plus one). Table A.26 and Table A.27 present the results using the inverse hyperbolic sine transformation. Table A.28 and Table A.29 display the coefficients on the basis of the negative binomial model. Table A.30 and Table A.31 show the results using the Poisson regressions. Across tables, we can observe that investments by large technology firms positively impact AI technology outputs by startups, and that these investments are more effective compared to investments by other firms. The results suggest that our findings are robust to alternative functional forms.

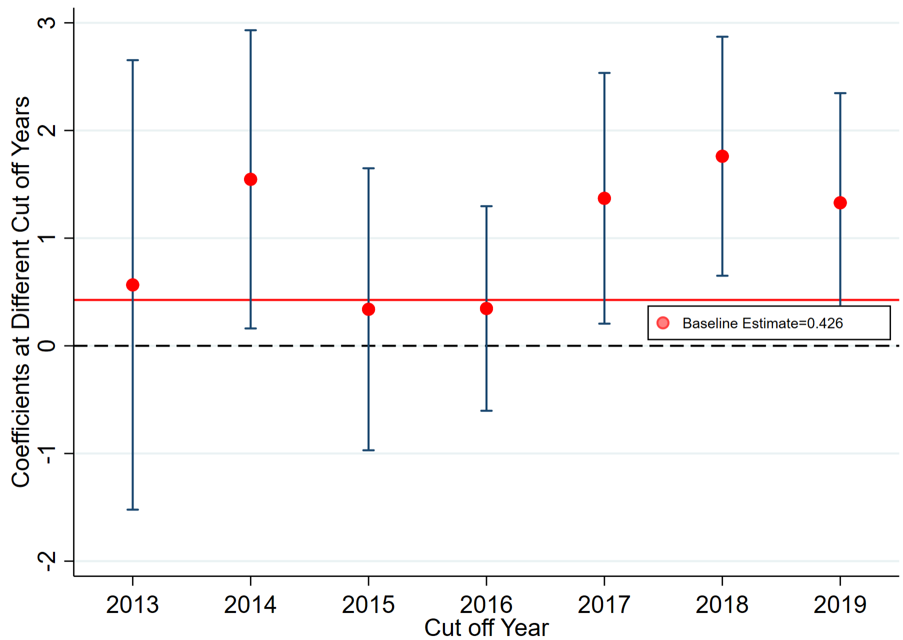


Figure 7: DDD Coefficients at Different Cut off Years

Note: The solid red line marks the baseline DDD estimate (0.426).

5.5 Innovation Potential

We have included various fixed effects in specifications in Section 3 to absorb the sources of selection of investments by large technology firms. In Section 4, we find that firm growth measured by registered capital has little explanatory power in AI innovation. For robustness, we further control for three variables capturing innovation potential: the initial number of AI patent applications, the initial number of registered software products, and the number of prior investment rounds (if any) before the first investment by large technology firms. We further have the three variables interacted with period fixed effects, respectively, and have the interactions included in Equation 3.1 and Equation 3.4. If our baseline results are driven mainly by the innovation potential of AI startups, the estimated coefficients of interest should change substantially, compared to the previous results.

The DID estimates and the DDD estimates are displayed in Table A.32 and Table A.33, respectively. We observe a slight reduction in the magnitude of DID estimates but the DDD estimates are very similar to those in the baseline triple differences estimation. Moreover, both tables show a positive and significant effect of investments by large technology firms. These estimates suggest that the bias caused by omitted innovation potential is less of a concern.

5.6 Joint Investments and the Amount of Investment

Two features of investments could confound our estimation. First, there are joint investments in the same startup made by several firms, although the majority of the investments are made by a single investor. Second, the monetary value of investment varies across investors.[35] The amount of money in each investment should also be reflected in the change in registered capital. As reported in Table A.10, we find registered capital plays a weak role in predicting AI technology outputs. For robustness, we further control for joint investments and the monetary value of investments interacted with the post dummy in the baseline DID and DDD estimation.

Table A.34 and Table A.35 show the estimates taking joint investments into consideration. As shown, joint investments are positively correlated with AI patent applications and software products. Conditional on joint investments, investments by large technology firms are still more effective at increasing AI innovation. Table A.36 and Table A.37 display the estimates when we control for the amount of investment interacted with the post dummy. Investments with a larger monetary value lead to a greater increase in both AI patent applications and software products. However, we still find positive effects of investments by large technology firms with similar magnitudes as our baseline estimation. These results suggest that our findings are driven by large technology firms, instead of by joint investments, and that large technology firms bring more than money to the table.

5.7 Additional Fixed Effects

In Equation 3.1 and Equation 3.4, we have included a variety of fixed effects to account for firm characteristics associated with large technology firms' selection and with county-specific shocks. For robustness, we further control for county×industry×period fixed effects and county×industry×firm type×period fixed effects. These fixed effects can absorb shocks and economic changes at a more granular level and ensure we are comparing firms with similar characteristics.

Table A.38 and Table A.39 present estimates considering additional fixed effects. Odd columns control for county×industry×period fixed effects with firm type×period fixed effects. Even columns further control for county×industry×firm type×period fixed effects, because there could be policy shocks or local economic environment changes that affect certain types of firms in some industries. For AI startups within the same county, the same industry and the same firm type, those receiving investments from large technology firms innovate more after the investments compared to those invested in by other firms. This evidence further mitigates the concern about omitted variable bias and lends additional confidence to our findings.

6 Conclusion

This paper empirically investigates the role of external financing in fostering innovation with non-rival inputs. We argue that investments from incumbents with data advantages stimulate innovation by startups receiving the investments through transferring data, which can be

[35] Around 60% of the investment events do not reveal the monetary value. We do not control for it in our baseline estimation to avoid dropping useful observations.

used simultaneously across multiple firms. We test the argument using unique data on Chinese AI startups. Using a staggered difference-in-differences strategy, we find that investments by large technology firms with data advantages substantially accelerate innovation by AI startups they invest in, as measured by AI-related patent applications and software products, compared to startups that do not receive such investments. Using a triple-differences strategy, we further find that the impact of investments by large technology firms is stronger than those by other firms without data advantages, including other CVC and independent VC/PE. Finally, we provide novel evidence that these results are driven mainly by AI startups acquiring data after building connections with large technology firms.

Even though data is crucial for AI innovation, whether large technology firms are willing to share data with AI startups remains uncertain. Our evidence suggests that AI startups receiving investments from large technology firms acquire data for developing AI technology. This evidence at least partially mitigates the concern that incumbents might stifle target AI startups' innovation, but raises another concern that data advantages are concentrated in large technology firms and firms with connections to them. More effective policies are needed to incentivize the sharing of the non-rival data across more firms, to achieve higher social welfare. The design of these incentives and how much data should be shared are interesting research topics for the future.

References

Acemoglu, Daron, Alex He, and Daniel le Maire, "Eclipse of rent-Sharing: The effects of managers' business education on wages and the labor share in the US and denmark," *National Bureau of Economic Research*, 2022.

_ **and Pascual Restrepo**, "Artificial intelligence, automation, and work," in "The economics of artificial intelligence: An agenda," University of Chicago Press, 2018, pp. 197–236.

_ **, David Autor, Jonathon Hazell, and Pascual Restrepo**, "AI and jobs: Evidence from online vacancies," *National Bureau of Economic Research*, 2020.

_ **, Suresh Naidu, Pascual Restrepo, and James A Robinson**, "Democracy does cause growth," *Journal of Political Economy*, 2019, *127* (1), 47–100.

Aghion, Philippe, Benjamin F Jones, and Charles I Jones, "Artificial intelligence and economic growth," in "The economics of artificial intelligence: An agenda," University of Chicago Press, 2018, pp. 237–282.

Alderucci, Dean, Lee Branstetter, Eduard Hovy, Andrew Runge, and Nikolas Zolas, "Quantifying the impact of AI on productivity and labor demand: Evidence from US census microdata," in "Allied Social Science Associations—ASSA 2020 Annual Meeting" 2020.

Andrabi, Tahir, Benjamin Daniels, and Jishnu Das, "Human capital accumulation and disasters: Evidence from the Pakistan earthquake of 2005," *Journal of Human Resources*, 2021, pp. 0520–10887R1.

Athey, Susan, "The impact of machine learning on economics," in "The economics of artificial intelligence: An agenda," University of Chicago Press, 2018, pp. 507–547.

Babina, Tania, Anastassia Fedyk, Alex Xi He, and James Hodson, "Artificial intelligence, firm growth, and product innovation," *Firm Growth, and Product Innovation (November 9, 2021)*, 2021.

_ **,** _ **,** _ **, and** _ , "Firm investments in artificial intelligence technologies and changes in workforce composition," *Available at SSRN*, 2022.

Barro, Robert J and Xavier Sala i Martin, "Public finance in models of economic growth," *The Review of Economic Studies*, 1992, *59* (4), 645–661.

Beraja, Martin, Andrew Kao, David Y Yang, and Noam Yuchtman, "AI-tocracy," *National Bureau of Economic Research*, 2021.

_ **, David Y Yang, and Noam Yuchtman**, "Data-intensive innovation and the state: evidence from AI firms in China," *Review of Economic Studies*, 2022.

Bessen, James E, Stephen Michael Impink, Lydia Reichensperger, and Robert Seamans, "The business of AI startups," *Boston Univ. School of Law, Law and Economics Research Paper*, 2018, (18-28).

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting event study designs: Robust and efficient estimation," *arXiv preprint arXiv:2108.12419*, 2021.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson, "Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics," in "The economics of artificial intelligence: An agenda," University of Chicago Press, 2018, pp. 23–57.

Chaisemartin, Clément De and Xavier d'Haultfoeuille, "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 2020, *110* (9), 2964–96.

Chemmanur, Thomas J, Elena Loutskina, and Xuan Tian, "Corporate venture capital, value creation, and innovation," *The Review of Financial Studies*, 2014, *27* (8), 2434–2473.

Cockburn, Iain M, Rebecca Henderson, and Scott Stern, "The impact of artificial intelligence on innovation: An exploratory analysis," in "The economics of artificial intelligence: An agenda," University of Chicago Press, 2018, pp. 115–146.

Cunningham, Colleen, Florian Ederer, and Song Ma, "Killer acquisitions," *Journal of Political Economy*, 2021, *129* (3), 649–702.

David, Hjjoep, "Why are there still so many jobs? The history and future of workplace automation," *Journal of Economic Perspectives*, 2015, *29* (3), 3–30.

Doyle, Orla, "The first 2,000 days and child skills," *Journal of Political Economy*, 2020, *128* (6), 2067–2122.

Eicher, Theo S, "Interaction between endogenous human capital and technological change," *The Review of Economic Studies*, 1996, *63* (1), 127–144.

Faurel, Lucile, Qin Li, Devin M Shanthikumar, and Siew Hong Teoh, "CEO incentives and product development innovation: Insights from trademarks," in "in" AAA 2015.

Fenske, James, Bishnupriya Gupta, and Song Yuan, "Demographic shocks and women's labor market participation: evidence from the 1918 influenza pandemic in India," *The Journal of Economic History*, 2022, *82* (3), 875–912.

Ford, Martin, *Rise of the robots: technology and the threat of a jobless future*, Basic Books, 2015.

Frey, Carl Benedikt and Michael A Osborne, "The future of employment: How susceptible are jobs to computerisation?," *Technological Forecasting and Social Change*, 2017, *114*, 254–280.

Fujii, Hidemichi and Shunsuke Managi, "Trends and priority shifts in artificial intelligence technology invention: A global patent analysis," *Economic Analysis and Policy*, 2018, *58*, 60–69.

Fulghieri, Paolo and Merih Sevilir, "Organization and financing of innovation, and the choice between corporate and independent venture capital," *Journal of Financial and Quantitative Analysis*, 2009, *44* (6), 1291–1321.

Furman, Jason and Robert Seamans, "AI and the economy," *Innovation Policy and the Economy*, 2019, *19* (1), 161–191.

Goldfarb, Avi, Joshua Gans, and Ajay Agrawal, *The economics of artificial intelligence: An agenda*, University of Chicago Press, 2019.

Grennan, Jillian and Roni Michaely, "Artificial intelligence and high-skilled work: Evidence from analysts," *Swiss Finance Institute Research Paper*, 2020, (20-84).

Guadalupe, Maria, Olga Kuzmina, and Catherine Thomas, "Innovation and foreign ownership," *American Economic Review*, 2012, *102* (7), 3594–3627.

Haucap, Justus, Alexander Rasch, and Joel Stiebale, "How mergers affect innovation: Theory and evidence," *International Journal of Industrial Organization*, 2019, *63*, 283–325.

Heath, Davidson and Christopher Mace, "The strategic effects of trademark protection," *The Review of Financial Studies*, 2020, *33* (4), 1848–1877.

Hellmann, Thomas and Manju Puri, "Venture capital and the professionalization of start-up firms: Empirical evidence," *The Journal of Finance*, 2002, *57* (1), 169–197.

Hicks, Raymond and Dustin Tingley, "Causal mediation analysis," *The Stata Journal*, 2011, *11* (4), 605–619.

Igna, Ioana, Francesco Venturini et al., "The determinants of AI innovation across European firms," in "University of Perugia Working Paper" 2021.

Imai, Kosuke, Luke Keele, and Teppei Yamamoto, "Identification, inference and sensitivity analysis for causal mediation effects," *Statistical science*, 2010, *25* (1), 51–71.

IPO, UK, "A worldwide overview of AI patents and patenting by the UK AI sector," 2019.

Ivanov, Vladimir I and Fei Xie, "Do corporate venture capitalists add value to start-up firms? Evidence from IPOs and acquisitions of VC-backed companies," *Financial Management*, 2010, *39* (1), 129–152.

Jones, Charles I and Christopher Tonetti, "Nonrivalry and the economics of data," *American Economic Review*, 2020, *110* (9), 2819–58.

Keisner, C Andrew, Julio Raffo, Sacha Wunsch-Vincent et al., *Breakthrough technologies–Robotics, Innovation and Intellectual Property*, Vol. 30, WIPO, 2015.

Lerner, Josh, "Assessing the contribution of venture capital," *the RAND Journal of Economics*, 2000, *31* (4), 674–692.

_ **and Ramana Nanda**, "Venture capital's role in financing innovation: What we know and how much we still need to learn," *Journal of Economic Perspectives*, 2020, *34* (3), 237–61.

Mendonça, Sandro, Tiago Santos Pereira, and Manuel Mira Godinho, "Trademarks as an indicator of innovation and industrial change," *Research Policy*, 2004, *33* (9), 1385–1404.

Mihet, Roxana and Thomas Philippon, "The economics of Big Data and artificial intelligence," in "Disruptive Innovation in Business and Finance in the Digital World," Emerald Publishing Limited, 2019.

Ornaghi, Carmine, "Mergers and innovation in big pharma," *International Journal of Industrial Organization*, 2009, *27* (1), 70–79.

Phillips, Gordon M and Alexei Zhdanov, "R&D and the incentives from merger and acquisition activity," *The Review of Financial Studies*, 2013, *26* (1), 34–78.

Romer, Paul M, "Endogenous technological change," *Journal of Political Economy*, 1990, *98* (5, Part 2), S71–S102.

—, "Mathiness in the theory of economic growth," *American Economic Review*, 2015, *105* (5), 89–93.

Roy, Vincent Van, Daniel Vertesy, and Giacomo Damioli, "AI and robotics innovation: A sectoral and geographical mapping using patent data," *GLO Discussion Paper*, 2019.

Seru, Amit, "Firm boundaries matter: Evidence from conglomerates and R&D activity," *Journal of Financial Economics*, 2014, *111* (2), 381–405.

Susskind, Daniel, *A world without work: Technology, automation and how we should respond*, Penguin UK, 2020.

Tricot, Roland, "Venture capital investments in artificial intelligence: Analysing trends in VC in AI companies from 2012 through 2020," 2021.

Veldkamp, Laura and Cindy Chung, "Data and the aggregate economy," *Journal of Economic Literature*, 2019.

West, Darrell M, *The future of work: Robots, AI, and automation*, Brookings Institution Press, 2018.

WIPO, "WIPO technology trends 2019: Artificial intelligence," *Geneva: World Intellectual Property Organization*, 2019.

Zhang, Daniel, Saurabh Mishra, Erik Brynjolfsson, John Etchemendy, Deep Ganguli, Barbara Grosz, Terah Lyons, James Manyika, Juan Carlos Niebles, Michael Sellitto et al., "The AI index 2022 annual report," *arXiv preprint arXiv:2103.06312*, 2021.

A Appendices

A.1 Data and Variables

Table A.1: IPC and CPC Codes of AI-related Patents

IPC Codes:				
G06F19/24	G06N3	G06N5	G06N7/02	G06N7/04
G06N7/06	G06N20	GO6T1/40	G16B4/20	G16B4/30
G16C20/70				
CPC Codes:				
H04L2025/03464	G06F19/24	G10H2250/151	G01N33/0034	A61B5/7267
H04N21/4662	G10H2250/311	G06F19/707	G01N2201/1296	B29C66/965
H04N21/4663	G10K2210/3024	G06F2207/4824	G01S7/417	B29C2945/76979
H04N21/4665	G10K2210/3038	G06K7/1482	G05B13/027	B60G2600/1876
B60G2600/1878	G05B13/0275	G06N3/004	G10L25/30	H04N21/4666
B60G2600/1879	G05B13/028	G06N3/02	G11B20/10518	H04Q2213/054
E21B2041/0028	G05B13/0285	G06N3/12	H01J2237/30427	H04Q2213/13343
F02D41/1405	G05B13/029	G06N5	H02P21/0014	H04Q2213/343
F03D7/046	G05B13/0295	G06N7	H02P23/0018	H04R25/507
F05B2270/707	G05B2219/33002	G06N20	H03H2017/0208	Y10S128/924
F05B2270/709	G05D1/0088	G06N99/005	H03H2222/04	Y10S128/925
F05D2270/707	G06F11/1476	G06T3/4046	H04L25/0254	Y10S706
F05D2270/709	G06F11/2257	G06T9/002	H04L25/03165	F16H2061/0081
G06F11/2263	G06T2207/20081	H04L41/16	F16H2061/0084	G06F15/18
G06T2207/20084	H04L45/08	G01N29/4481	G06F17/16	G08B29/186
H04L2012/5686				

Notes: Information from IPO (2019).

Table A.3: Summary Statistics before Investments

	With Investments by Tech Giants				Without Investments by Tech Giants			
Variables	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
AI patent	0.80	1.80	0.00	9.00	2.23	2.13	1.00	9.00
Software	8.83	13.21	0.00	72.00	14.16	20.24	0.00	117.00
Data patent	3.28	9.44	0.00	82.00	6.26	14.19	0.00	82.00
Other patent	13.23	51.11	0.00	428.00	20.75	56.29	0.00	428.00
Estab year	2015	2.43	2010	2020	2016	2.62	2010	2020
Capital	3.50	9.90	0.01	68.93	3.35	9.35	0.01	68.93
Observations:	314				9461			

Notes: We report the summary statistics of the cumulative number of AI-related patent applications, registered software products, non-AI data-related patent applications, other (non-AI, non-data-related) patent applications, establishment year and registered capital of the startups. The startups are split into two groups. The first four columns display the summary statistics of startups receiving investments by large technology firms. The last four columns display summary statistics of startups without receiving investments by large technology firms. For the first four columns, we report summary statistics before the first investment. For the last four columns, we report summary statistics at the end of 2020. These outcome variables are winsorized at the 1st and the 99th percentile.

Table A.2: List of Keywords Related to AI

人工智能 (Artificial Intelligence)	机器学习 (Machine Learning)
深度学习 (Deep Learning)	无监督学习 (Unsupervised Learning)
自然语言处理 (Natural Language Processing)	神经网络 (Neural Network)
增强式学习 (Reinforcement Learning)	随机森林 (Random Forest)
对象侦测 (Object Detection)	符号推理 (Symbolic Reasoning)
模式分析 (Pattern Analysis)	贝叶斯信念网络 (Bayesian Belief Networks)
知识表达与推理 (Knowledge Representing)	神经计算 (Neuromorphic Computing)
协作系统 (Collaborative Systems)	自动分类 (Automatic Classification)
分层控制系统 (Layered Control Systems)	语音识别 (Speech Recognition)
语言识别 (Language Recognition)	语音生成 (Speech Generation)
文本语音转换 (Text-to-speech Conversion)	语音转换文本 (Speech-to-text Conversion)
机器翻译 (Machine Translation)	人脸识别 (Face Recognition)
表情识别 (Expression Recognition)	面部识别 (Facial Recognition)
模式识别 (Pattern Recognition)	图像识别 (Picture Recognition)
计算机视觉 (Computer Vision)	电脑视觉 (Computer Vision)
自动驾驶 (Autonomous Drive)	无人汽车 (Autonomous Vehicle)
智能驾驶 (Intelligent Drive)	智能机器人 (Intelligent Robot)
人型机器人 (Humanoid Robot)	机器智能 (Machine Intelligence)
智能搜索 (Intelligent Search)	最优搜索 (Optimal Search)
无人机 (Drone)	虚拟现实 (Virtual Reality)

Notes: We use Chinese keywords in the search. English translations are in parentheses.

Table A.4: DID: Instrumental Variables with More Fixed Effects

Second-Stage Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	2.320***	1.445***	19.402***	14.089***
	(0.583)	(0.473)	(4.500)	(3.354)
First-Stage Dependent Var.:	TechInvest×Post			
	(1)	(2)	(3)	(4)
RatioTechInvest t-1	0.322***	0.358***	0.322***	0.358***
	(0.042)	(0.039)	(0.042)	(0.039)
RatioTechInvest t-2	0.172***	0.126***	0.172***	0.126***
	(0.041)	(0.035)	(0.041)	(0.035)
RatioTechInvest t-3	0.251***	0.242***	0.251***	0.242***
	(0.052)	(0.046)	(0.052)	(0.046)
County×Industry×Period FE	Yes	No	Yes	No
County×Firm Type×Period FE	No	Yes	No	Yes
Founding Year×Period FE	Yes	Yes	Yes	Yes
Firm Type×Period FE	Yes	No	Yes	No
Industry×Period FE	No	Yes	No	Yes
Period FE	No	No	No	No
Firm FE	Yes	Yes	Yes	Yes
Observations	77071	90720	77071	90720
R^2	0.742	0.731	0.742	0.731
KP F-statistics	55.1	70.9	55.1	70.9

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.5: DDD: Instrumental Variables with More Fixed Effects

Second-Stage Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	1.972***	1.344***	12.449***	9.162***
	(0.525)	(0.440)	(3.743)	(2.905)
Invest×Post	0.181***	0.213***	3.693***	4.233***
	(0.046)	(0.041)	(0.476)	(0.428)
First-Stage Dependent Var.:	Invest×TechInvest×Post			
	(1)	(2)	(3)	(4)
Invest×RatioTechInvest t-1	0.322***	0.358***	0.322***	0.358***
	(0.042)	(0.039)	(0.042)	(0.039)
Invest×RatioTechInvest t-2	0.172***	0.126***	0.172***	0.126***
	(0.041)	(0.035)	(0.041)	(0.035)
Invest×RatioTechInvest t-3	0.251***	0.242***	0.251***	0.242***
	(0.052)	(0.046)	(0.052)	(0.046)
County×Industry×Period FE	Yes	No	Yes	No
County×Firm Type×Period FE	No	Yes	No	Yes
Founding Year×Period FE	Yes	Yes	Yes	Yes
Firm Type×Period FE	Yes	No	Yes	No
Industry×Period FE	No	Yes	No	Yes
Period FE	No	No	No	No
Firm FE	Yes	Yes	Yes	Yes
Observations	77071	90720	77071	90720
R^2	0.742	0.731	0.742	0.731
KP F-statistics	67.9	81.8	67.9	81.8

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

A.2 IV: Exclusion Restrictions

A.3 Mechanisms

Table A.6: IPC and CPC Codes of Data-related Patents

IPC/CPC Code	Descriptions of Patents
G06F	Electric digital data processing
G06Q	Data processing systems or methods
G06T	Image data processing
G06E	Optical computer devices
G06J	Hybrid computing arrangements
G06K	Graphical data reading
H04W8	Network data management

Notes: There could be an overlap between data-related patents and AI-related patents. We exclude AI-related patents to obtain non-AI data-related patents.

Table A.7: Keywords in the Titles of Data-related Online Job Postings

数据处理 (Data processing)	数据整理 (Data processing)
数据分析 (Data analysis)	数据管理 (Data management)
数据工程 (Data engineering)	数据服务 (Data service)
数据模型 (Data model)	数据统计 (Data statistics)
数据开发 (Data development)	数据网络 (Data network)
数据需求 (Data demand)	数据研发 (Data research)
数据前端 (Data front-end)	数据采集 (Data collection)
数据收集 (Data collection)	数据挖掘 (Data mining)
数据考核 (Data assessment)	数据测试 (Data assessment)
数据预测 (Data prediction)	数据资源 (Data resource)
数据运营 (Data operation)	数据设计 (Data design)
大数据 (Big data)	App 数据 (App data)
数字分析 (Digital analysis)	数字经济 (Digital economy)
算法 (Algorithm)	云数据 (Cloud data)
云计算 (Cloud computing)	数理统计 (Mathematical statistics)
数理分析 (Mathematical analysis)	数理模型 (Mathematical model)

Notes: We use Chinese keywords to define data-related jobs. English translations are in parentheses.

A.4 Robustness Analysis

Table A.8: Placebo: Other Patent Applications

Dependent Var.:	Other Patent A.		AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)	(5)	(6)
Invest×TechInvest×Post	6.730**	6.698**	0.518***	0.359***	4.946***	3.252***
	(3.167)	(3.215)	(0.114)	(0.119)	(1.178)	(1.104)
Invest×Post	-1.070	-0.297	0.330***	0.233***	5.432***	4.269***
	(1.114)	(1.099)	(0.039)	(0.040)	(0.437)	(0.436)
Other Patent A.			0.010***	0.010***	0.039***	0.051***
			(0.001)	(0.001)	(0.007)	(0.007)
County×Period FE	No	Yes	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes	No	Yes
Period FE	Yes	No	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Mediation Effect			0.063	0.066	0.257	0.332
Direct Effect			0.514	0.355	4.905	3.216
Proportion Mediated			10.9%	15.8%	5.0%	9.4%
Observations	100401	94625	100401	94625	100401	94625
R^2	0.753	0.807	0.670	0.709	0.699	0.753
Mean Dependent Var.:	13.931	13.931	0.926	0.926	7.574	7.574

Notes: The first outcome variable is the cumulative number of non-AI, non-data-related patent applications. Invest is a dummy indicating the firm was ever invested by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.9: Placebo: Jobs with Low education requirements

Dependent Var.:	Low-edu Jobs		AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)	(5)	(6)
Invest×TechInvest×Post	1.659	1.489	0.887***	0.697***	3.996***	2.535***
	(2.341)	(2.453)	(0.146)	(0.150)	(0.878)	(0.873)
Invest×Post	0.361	0.076	0.447***	0.302***	3.902***	3.388***
	(0.561)	(0.598)	(0.050)	(0.052)	(0.364)	(0.383)
Low-education Jobs			0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.002)	(0.002)
County×Period FE	No	Yes	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes	No	Yes
Period FE	Yes	No	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Mediation Effect			0.000	0.000	0.001	0.000
Direct Effect			0.751	0.577	4.896	3.269
Proportion Mediated			0.0%	0.0%	0.0%	0.0%
Observations	78596	75188	78596	75188	78596	75188
R^2	0.775	0.786	0.705	0.735	0.825	0.847
Mean Dependent Var.:	15.434	15.434	1.153	1.153	9.014	9.014

Notes: The first outcome variable is the cumulative number of online job postings requiring a high school degree or below. Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.10: Potential Alternative Mechanism: Amount of Investment

Dependent Var.:	Registered Capital		AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)	(5)	(6)
Invest×TechInvest×Post	0.982**	0.578	0.568***	0.443***	5.006***	3.594***
	(0.430)	(0.434)	(0.120)	(0.123)	(1.188)	(1.110)
Invest×Post	0.510***	0.102	0.338***	0.258***	5.567***	4.403***
	(0.180)	(0.179)	(0.039)	(0.040)	(0.429)	(0.426)
Registered Capital			0.017***	0.015***	0.248***	0.256***
			(0.003)	(0.003)	(0.037)	(0.036)
County×Period FE	No	Yes	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes	No	Yes
Period FE	Yes	No	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Mediation Effect			0.017	0.009	0.239	0.143
Direct Effect			0.564	0.431	4.956	3.504
Proportion Mediated			2.9%	2.0%	4.6%	3.9%
Observations	97928	92184	97928	92184	97928	92184
R^2	0.823	0.845	0.657	0.697	0.700	0.752
Mean Dependent Var.:	2.888	2.888	0.926	0.926	7.574	7.574

Notes: The first outcome variable is the registered capital of startups. Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.11: Potential Alternative Mechanism: Government Contract

Dependent Var.:	Gov Contract		AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)	(5)	(6)
Invest×TechInvest×Post	-1.650	-1.566	0.584***	0.429***	5.261***	3.638***
	(2.128)	(2.175)	(0.117)	(0.121)	(1.179)	(1.110)
Invest×Post	2.008	2.098	0.317***	0.226***	5.327***	4.190***
	(1.772)	(1.565)	(0.040)	(0.041)	(0.431)	(0.433)
Gov Contract			0.001***	0.002***	0.032***	0.030***
			(0.000)	(0.000)	(0.007)	(0.008)
County×Period FE	No	Yes	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes	No	Yes
Period FE	Yes	No	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Mediation Effect			-0.002	-0.003	-0.055	-0.049
Direct Effect			0.580	0.425	5.220	3.601
Proportion Mediated			-0.4%	-0.7%	-1.1%	-1.4%
Observations	100401	94625	100401	94625	100401	94625
R^2	0.367	0.461	0.654	0.695	0.702	0.752
Mean Dependent Var.:	4.114	4.114	0.926	0.926	7.574	7.574

Notes: The first outcome variable is number of government contracts. Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.12: DID: 20 Largest Technology Firms

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.743***	0.510***	7.202***	4.366***
	(0.176)	(0.179)	(1.503)	(1.575)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	101138	95365	101138	95365
R^2	0.651	0.692	0.687	0.743
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.13: DDD: 20 Largest Technology Firms

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.658***	0.465***	5.823***	3.553**
	(0.177)	(0.180)	(1.537)	(1.577)
Invest×Post	0.340***	0.238***	5.530***	4.282***
	(0.039)	(0.041)	(0.436)	(0.438)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	101138	95365	101138	95365
R^2	0.652	0.693	0.692	0.745
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.14: DID: 100 Largest Technology Firms

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.644***	0.454***	6.989***	4.738***
	(0.114)	(0.119)	(1.107)	(1.056)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100373	94598	100373	94598
R^2	0.652	0.694	0.691	0.745
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.15: DDD: 100 Largest Technology Firms

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.527***	0.382***	5.037***	3.454***
	(0.116)	(0.120)	(1.158)	(1.075)
Invest×Post	0.327***	0.237***	5.409***	4.263***
	(0.040)	(0.041)	(0.435)	(0.436)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100373	94598	100373	94598
R^2	0.654	0.694	0.695	0.748
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.16: DID: Based on the Method by Borusyak et al. (2021)

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.730***	0.490***	7.407***	5.645***
	(0.108)	(0.114)	(1.050)	(1.049)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100225	99439	100225	99439
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.17: DDD: Based on the Method by Borusyak et al. (2021)

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.582***	0.394***	4.866***	3.879***
	(0.109)	(0.115)	(1.089)	(1.065)
Invest×Post	0.337***	0.247***	5.788***	4.546***
	(0.040)	(0.042)	(0.438)	(0.453)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100225	99439	100225	99439
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.18: DID: Granted Patents and Trademarks

Dependent Var.:	Granted Patents		Trademarks	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.166***	0.118***	28.239***	24.903***
	(0.040)	(0.042)	(3.002)	(2.966)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
R^2	0.650	0.687	0.744	0.778
Mean Dependent Var.:	0.171	0.171	9.502	9.502

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.19: DDD: Granted Patents and Trademarks

Dependent Var.:	Granted Patents		Trademarks	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.120***	0.090**	25.526***	23.196***
	(0.041)	(0.043)	(3.034)	(2.989)
Invest×Post	0.126***	0.092***	7.445***	5.533***
	(0.015)	(0.016)	(0.697)	(0.748)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
R^2	0.652	0.688	0.746	0.779
Mean Dependent Var.:	0.171	0.171	9.502	9.502

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.20: DID: Coefficients at Different Cut off Years

Dependent Var.:	AI Patent Applications						
Cutoff Year:	2013	2014	2015	2016	2017	2018	2019
TechInvest×Post	0.632	1.681**	0.412	0.535	1.495**	1.898***	1.451***
	(0.954)	(0.687)	(0.678)	(0.490)	(0.586)	(0.559)	(0.518)
County×Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Founding Year×Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Type×Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period FE	No	No	No	No	No	No	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1207	2571	5518	12689	26286	47223	69281
R^2	0.834	0.857	0.835	0.809	0.772	0.737	0.716
Mean Dependent Var.:	1.606	1.605	1.629	1.620	1.470	1.283	1.105

Notes: We only consider startups with at least an AI patent application and without receiving investments from large technology firms by a certain year as AI startups. We change the cut off year to be from 2013 to 2019. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.21: DDD: Coefficients at Different Cut off Years

Dependent Var.:	AI Patent Applications						
Cutoff Year:	2013	2014	2015	2016	2017	2018	2019
Invest×TechInvest×Post	0.566	1.546**	0.339	0.347	1.370**	1.761***	1.329**
	(1.046)	(0.701)	(0.666)	(0.484)	(0.594)	(0.566)	(0.519)
Invest×Post	0.247	0.326	0.252	0.501***	0.407***	0.380***	0.322***
	(0.581)	(0.258)	(0.193)	(0.155)	(0.104)	(0.068)	(0.052)
County×Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Founding Year×Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Type×Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period FE	No	No	No	No	No	No	No
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1207	2571	5518	12689	26286	47223	69281
R^2	0.834	0.858	0.835	0.810	0.773	0.738	0.716
Mean Dependent Var.:	1.606	1.605	1.629	1.620	1.470	1.283	1.105

Notes: We only consider startups with at least an AI patent application and without receiving investments from large technology firms by a certain year as AI startups. We change the cut off year to be from 2013 to 2019. Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.22: DID: Number of Outcomes during Each Period

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.226***	0.170***	0.850***	0.653***
	(0.036)	(0.037)	(0.172)	(0.178)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
R^2	0.221	0.284	0.310	0.360
Mean Dependent Var.:	0.218	0.218	1.389	1.389

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.23: DDD: Number of Outcomes during Each Period

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.177***	0.137***	0.523***	0.433**
	(0.036)	(0.038)	(0.176)	(0.181)
Invest×Post	0.133***	0.105***	0.896***	0.714***
	(0.011)	(0.012)	(0.064)	(0.068)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
R^2	0.223	0.285	0.313	0.362
Mean Dependent Var.:	0.218	0.218	1.389	1.389

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.24: DID: Outcome Variables in Logarithm (Plus One)

Dependent Var.:	ln(1+AI Patent Applications)		ln(1+Registered Software)	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.175***	0.120***	0.521***	0.318***
	(0.030)	(0.031)	(0.049)	(0.047)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
R^2	0.690	0.731	0.773	0.828
Mean Dependent Var.:	0.443	0.443	1.216	1.216

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.25: DDD: Outcome Variables in Logarithm (Plus One)

Dependent Var.:	ln(1+AI Patent Applications)		ln(1+Registered Software)	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.137***	0.097***	0.303***	0.186***
	(0.031)	(0.032)	(0.051)	(0.048)
Invest×Post	0.104***	0.075***	0.597***	0.429***
	(0.011)	(0.012)	(0.023)	(0.023)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
R^2	0.691	0.732	0.780	0.831
Mean Dependent Var.:	0.443	0.443	1.216	1.216

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.26: DID: Inverse Hyperbolic Sine Transformation

Dependent Var.:	IHS(AI Patent Applications)		IHS(Registered Software)	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.223***	0.153***	0.602***	0.362***
	(0.039)	(0.040)	(0.059)	(0.056)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
R^2	0.689	0.731	0.769	0.825
Mean Dependent Var.:	0.569	0.569	1.502	1.502

Notes: Outcome variables are transformed using an inverse hyperbolic sine function. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.27: DDD: Inverse Hyperbolic Sine Transformation

Dependent Var.:	IHS(AI Patent Applications)		IHS(Registered Software)	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.174***	0.123***	0.341***	0.204***
	(0.039)	(0.040)	(0.061)	(0.056)
Invest×Post	0.134***	0.098***	0.717***	0.511***
	(0.015)	(0.015)	(0.028)	(0.027)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
R^2	0.690	0.731	0.776	0.828
Mean Dependent Var.:	0.569	0.569	1.502	1.502

Notes: Outcome variables are transformed using an inverse hyperbolic sine function. Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.28: DID: Negative Binomial Regression

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.670***	0.463***	0.615***	0.401***
	(0.115)	(0.128)	(0.106)	(0.119)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
Pseudo R^2	0.000	0.000	0.000	0.000
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: Negative binomial regression. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.29: DDD: Negative Binomial Regression

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.547***	0.389***	0.482***	0.308**
	(0.119)	(0.131)	(0.116)	(0.126)
Invest×Post	0.415***	0.289***	0.673***	0.509***
	(0.059)	(0.063)	(0.058)	(0.064)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
Pseudo R^2	0.001	0.001	0.001	0.000
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: Negative binomial regression. Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.30: DID: Poisson Regression

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.744***	0.518***	0.921***	0.617***
	(0.136)	(0.151)	(0.171)	(0.191)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: Poisson regression. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.31: DDD: Poisson Regression

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.624***	0.445***	0.678***	0.456**
	(0.139)	(0.153)	(0.178)	(0.196)
Invest×Post	0.356***	0.244***	0.743***	0.554***
	(0.055)	(0.057)	(0.076)	(0.078)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: Poisson regression. Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.32: DID: Controlling for Growth Potential

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.439***	0.286**	2.373**	2.471**
	(0.110)	(0.114)	(0.943)	(0.960)
Growth Potential	Yes	Yes	Yes	Yes
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100231	94451	100231	94451
R^2	0.656	0.698	0.703	0.753
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.33: DDD: Controlling for Growth Potential

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.581***	0.431***	4.955***	3.484***
	(0.117)	(0.122)	(1.224)	(1.166)
Invest×Post	0.312***	0.223***	5.169***	4.175***
	(0.040)	(0.041)	(0.430)	(0.431)
Growth Potential	Yes	Yes	Yes	Yes
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100240	94461	100240	94461
R^2	0.656	0.697	0.706	0.754
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.34: DID: Controlling for Joint Investments

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.528***	0.351**	4.643***	2.656**
	(0.182)	(0.177)	(1.324)	(1.345)
Joint×Post	0.263	0.224	3.893*	3.448*
	(0.235)	(0.236)	(2.062)	(1.995)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100401	94625	100401	94625
R^2	0.652	0.693	0.691	0.745
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.35: DDD: Controlling for Joint Investments

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.491***	0.336***	3.993***	2.523**
	(0.123)	(0.126)	(1.213)	(1.174)
Invest×Post	0.269***	0.180***	4.707***	3.667***
	(0.045)	(0.046)	(0.485)	(0.478)
Joint×Post	0.179**	0.176**	2.401**	2.076**
	(0.084)	(0.088)	(0.944)	(0.940)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	100389	94613	100389	94613
R^2	0.653	0.694	0.696	0.748
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.36: DID: Controlling for the Amounts of Investments

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.640***	0.441*	7.990***	5.633***
	(0.244)	(0.245)	(1.913)	(1.895)
Amount×Post	0.118***	0.115***	0.267	0.211
	(0.036)	(0.042)	(0.210)	(0.200)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	97890	92104	97890	92104
R^2	0.652	0.695	0.689	0.745
Mean Dependent Var.:	0.919	0.919	7.439	7.439

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.37: DDD: Controlling for the Amounts of Investments

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.739***	0.495**	6.483***	4.088**
	(0.249)	(0.243)	(2.010)	(1.843)
Invest×Post	0.249***	0.186**	8.212***	6.467***
	(0.085)	(0.090)	(1.039)	(1.082)
Amount×Post	0.084***	0.101***	0.223	0.304*
	(0.032)	(0.032)	(0.231)	(0.162)
County×Period FE	No	Yes	No	Yes
Founding Year×Period FE	No	Yes	No	Yes
Firm Type×Period FE	No	Yes	No	Yes
Industry×Period FE	No	Yes	No	Yes
Period FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
Observations	79394	73729	79394	73729
R^2	0.660	0.708	0.689	0.749
Mean Dependent Var.:	0.900	0.900	6.612	6.612

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.38: DID: Additional Fixed Effects

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
TechInvest×Post	0.595***	0.565***	5.459***	5.726***
	(0.133)	(0.139)	(1.198)	(1.254)
County×Ind×Period FE	Yes	No	Yes	No
County×Ind×Type×Period FE	No	Yes	No	Yes
Founding Year×Period FE	Yes	Yes	Yes	Yes
Firm Type×Period FE	Yes	No	Yes	No
Period FE	No	No	No	No
Firm FE	Yes	Yes	Yes	Yes
Observations	78820	76045	78820	76045
R^2	0.724	0.733	0.778	0.784
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.

Table A.39: DDD: Additional Fixed Effects

Dependent Var.:	AI Patent Applications		Software Products	
	(1)	(2)	(3)	(4)
Invest×TechInvest×Post	0.532***	0.495***	4.387***	4.631***
	(0.134)	(0.140)	(1.221)	(1.278)
Invest×Post	0.229***	0.263***	3.909***	4.078***
	(0.046)	(0.047)	(0.483)	(0.496)
County×Ind×Period FE	Yes	No	Yes	No
County×Ind×Type×Period FE	No	Yes	No	Yes
Founding Year×Period FE	Yes	Yes	Yes	Yes
Firm Type×Period FE	Yes	No	Yes	No
Period FE	No	No	No	No
Firm FE	Yes	Yes	Yes	Yes
Observations	78820	76045	78820	76045
R^2	0.725	0.733	0.779	0.786
Mean Dependent Var.:	0.926	0.926	7.574	7.574

Notes: Invest is a dummy indicating the firm was ever invested in by a large technology firm or other firms. TechInvest is a dummy indicating the firm was ever invested in by a large technology firm. Post is a dummy indicating the periods since the first investment. Standard errors are clustered at the firm level. * p<0.1; ** p<0.05; *** p<0.01.