



# Artificial intelligence, firm growth, and product innovation

Tania Babina <sup>a,b</sup>, Anastassia Fedyk <sup>c,\*</sup>, Alex He <sup>d</sup>, James Hodson <sup>e,f</sup>

<sup>a</sup> Columbia University, New York, NY, USA

<sup>b</sup> NBER, USA

<sup>c</sup> University of California, Berkeley, Berkeley, CA, USA

<sup>d</sup> University of Maryland, College Park, MD, USA

<sup>e</sup> AI for Good Foundation, Berkeley, CA, USA

<sup>f</sup> Institut Jožef Stefan, Ljubljana, Slovenia

## ARTICLE INFO

Dataset link: <https://data.mendeley.com/datasets/s26kxvsn7/2>

JEL classification:

D22

E22

J23

J24

L11

O33

Keywords:

Artificial intelligence

Intangible capital

Technological change

Product innovation

Superstar firms

Industry concentration

## ABSTRACT

We study the use and economic impact of AI technologies. We propose a new measure of firm-level AI investments using employee resumes. Our measure reveals a stark increase in AI investments across sectors. AI-investing firms experience higher growth in sales, employment, and market valuations. This growth comes primarily through increased product innovation. Our results are robust to instrumenting AI investments using firms' exposure to universities' supply of AI graduates. AI-powered growth concentrates among larger firms and is associated with higher industry concentration. Our results highlight that new technologies like AI can contribute to growth and superstar firms through product innovation.

## 1. Introduction

Technological change is a key driver of investment opportunities and economic growth (Romer, 1990; Aghion and Howitt, 1992; Kogan et al., 2017). The past decade has seen a new technological shift: substantial developments in artificial intelligence (AI) technologies and their wide-spread commercial application (Furman and Seamans, 2019). As a prediction technology, AI allows firms to learn better and faster from vast quantities of data, with the potential to significantly improve business decision-making. As such, AI can be a general purpose technology that generates growth through increased productivity and product innovation across a wide range of sectors (Aghion et al., 2017; Agrawal et al., 2019).<sup>1</sup> Yet it remains an open question whether AI can transform

economies and spur economic growth, as lackluster aggregate productivity growth over the past decade has led to concerns that the benefits of AI may be over-hyped or take a much longer time to materialize (Mihet and Philippon, 2019; Brynjolfsson et al., 2019). To date, the lack of comprehensive data on firm-level AI adoption has posed the key challenge to understanding the adoption patterns and the economic impact of AI technologies (Seamans and Raj, 2018).

In this paper, we propose a new measure of investments in AI technologies based on firms' AI-skilled human capital. The heavy reliance of AI on human expertise makes the human-capital-based approach particularly well-suited in this setting. We take advantage of a unique combination of datasets that capture both the *stock* of and the *demand* for AI-skilled employees among U.S. firms: resume data from Cognism

\* Corresponding author at: 545 Student Services Building #1900, Berkeley, CA 94720, USA.

E-mail addresses: [tania.babina@gsb.columbia.edu](mailto:tania.babina@gsb.columbia.edu) (T. Babina), [fedyk@berkeley.edu](mailto:fedyk@berkeley.edu) (A. Fedyk), [axhe@umd.edu](mailto:axhe@umd.edu) (A. He), [hodson@ai4good.org](mailto:hodson@ai4good.org) (J. Hodson).

<sup>1</sup> In a 2018 Deloitte survey of executives at companies investing in AI, 70% anticipate that AI will fundamentally transform their companies and industries within the next five years.

<https://doi.org/10.1016/j.jfineco.2023.103745>

Received 19 May 2022; Received in revised form 27 October 2023; Accepted 27 October 2023

Available online 23 November 2023

0304-405X/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Inc., which offer job histories for 535 million individuals globally, and job postings data from Burning Glass, which capture 180 million job vacancies. Our new AI measure allows us to analyze the patterns of AI adoption and examine its potential benefits for the adopting firms and industries. Our main takeaway is that firms that invest more in AI experience higher growth through increased product innovation, which can be seen in increased trademarks, product patents, and updates to firms' product portfolios. Our results suggest that, so far, the first-order effect of AI has been in empowering growth through product innovation, consistent with AI reducing the costs of product development.

Our work offers several innovations over the existing literature. First, we introduce a novel measure of firm-level investments in AI technologies. Our detailed data and measure allow us to study the impact of AI technologies on firms, whereas other studies focus on the impact of AI on labor (Acemoglu et al., 2022b) and tend to look at the occupation or aggregate level (e.g., Felten et al., 2019). We provide novel evidence that AI investments are associated with firm growth and explore the mechanisms by which this growth can accrue. Second, we are able to measure AI adoption for a broad sample of AI-using firms across a wide range of industries, which complements recent work that focuses on AI-inventing firms (Alderucci et al., 2020). Our broad industry coverage allows us to examine the implications of AI investments for aggregate trends such as industry growth and concentration. Third, in the absence of administrative U.S. firm-worker matched data with individual workers' occupations, our Cognism resume data provide unique coverage of U.S. jobs with detailed job descriptions while representing more than 64% of full-time U.S. employment as of 2018.<sup>2</sup> This enables us to compare AI labor demand identified from job postings with the stock of AI workers identified from resumes. Finally, our rich data on firms' human capital allow us to measure and control for confounding factors, such as the use of non-AI information technologies, and capture the use of external AI solutions and software (e.g., IPSoft Amelia).

Even with our detailed data, identifying firms' AI investments is challenging due to the multifaceted nature of AI applications.<sup>3</sup> We circumvent this challenge by proposing a new data-driven approach to identify AI-related jobs, which does not depend on pre-specified lists of keywords. Instead, our algorithm learns the AI-relatedness of each job empirically from the detailed skills of the job postings. First, we measure the AI-relatedness of each skill in the job postings data, based on that skill's co-occurrence with the core AI skills—machine learning, computer vision, and natural language processing. Second, we obtain a measure of AI-relatedness of each job posting by averaging the AI-relatedness of all skills required by the job posting. Finally, we leverage the most AI-related skills identified from the job postings data to classify AI workers in the less structured resume data. For each employee, we consider whether skills with the highest AI-relatedness (e.g., “deep learning”) appear either in the job title, in the job description, or in any publications, patents, or awards received during that job. This gives us a classification of each employee of each firm at each point in time. We aggregate both the resume data and the job postings data to the firm level and match to public firms in the Compustat database. Encouragingly, the two measures of AI investments, although based on two independent datasets, are highly correlated and yield consistent results.

We confirm that our human-capital-based measures of AI investments display intuitive properties. First, we confirm that our AI measure

does not pick up general data-related skills, only those that are specifically associated with AI implementation. Second, we manually inspect large samples of AI-classified jobs and confirm that our classification picks up highly AI-skilled positions. Third, given that we mainly rely on required skills to identify AI-related terms, we validate our measure by confirming that the job postings with the highest AI-relatedness measures skew heavily towards highly AI-specific job titles. Fourth, we provide detailed case studies of specific applications of AI within several firms. Fifth, we confirm that AI-investing firms also increase research and development (R&D) expenditures, consistent with increased experimentation with applying the new AI technologies. Finally, we enrich our baseline measure by incorporating the use of external AI solutions and software, confirming that this augmented measure yields very similar results.

We begin our analysis by describing key patterns in AI investments. In both employee resume and job postings datasets, the fraction of AI jobs has increased dramatically over time, growing more than seven-fold from 2010 to 2018. The share of AI jobs is highest in the technology sector, but the rate of increase in AI investments over time is similar across sectors. At the firm level, growth in AI investments is more pronounced among ex ante larger firms and firms with higher cash holdings. Looking at the local labor market conditions, we observe that higher-wage and more educated areas experience faster growth in AI-skilled hiring.

We next address the fundamental question of whether AI investments are associated with higher firm growth. As is standard in settings with slow-moving processes like technological change (e.g., Acemoglu and Restrepo, 2020), our primary specification is a long-differences regression of changes in firm outcomes from 2010 to 2018 on changes in firm-level AI-skilled human capital, measured by the share of AI workers. This strategy is especially well-suited for our setting, where AI investments accumulate gradually over time and generate effects that may not be immediate. We include a rich set of controls: industry fixed effects and firm-, industry-, and commuting-zone-level characteristics as of 2010. We document a strong and consistent pattern of higher growth among firms that invest more in AI: a one-standard-deviation increase in the resume-based measure of AI investments over the 8-year period corresponds to a 19.5% increase in sales, a 18.1% increase in employment, and a 22.3% increase in market valuation. The results are ubiquitous across major industry sectors (e.g., manufacturing, finance, and retail), supporting the idea that AI is a general purpose technology.

While the long-differences specification controls for time-invariant firm characteristics, we perform several tests to address concerns about omitted variables or reverse causality. First, we exploit firm-level panel data to examine firm growth dynamically in each year around AI investments using a standard distributed lead-lag model (Aghion et al., 2020). We find no pre-trends in firm growth prior to AI investments, confirming that AI-investing firms are not on differential growth trends, and an increase after a lag of two to three years, suggesting that the effects of AI are not immediate. Second, the results are robust to the inclusion of controls for past firm and industry growth and future growth opportunities proxied by Tobin's  $q$ . Third, we confirm that our results reflect specifically investments in AI, rather than other technologies: the effects of AI investments remain unchanged when controlling for contemporaneous firm-level investments in robotics, non-AI information technologies, and non-AI data analytics.

To further address concerns regarding unobserved shocks driving both firm growth and AI investments, we use a novel instrumental variables (IV) strategy: we instrument for firm-level AI investments using variation in firms' ex-ante exposure to the subsequent supply of AI talent from universities that are historically strong in AI research. The core idea is that the scarcity of AI-trained labor is one of the most important constraints to firms' AI adoption (e.g., CorrelationOne, 2019), and universities that are historically strong in AI research have been able to train more AI-skilled graduates in recent years, enabling firms that historically hired from those universities to more readily recruit

<sup>2</sup> For comparison, while the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program provides firm-worker matched data and worker wages, it does not include any information on workers' occupations or their jobs (Abowd et al., 2009; Haltiwanger et al., 2014). Moreover, a typical project using the LEHD data does not have access to all states due to administrative reasons: for example, Babina (2020) has access to about 40% of employment and Babina and Howell (2023)—60%.

<sup>3</sup> Within a single firm, e.g., Caterpillar Inc., AI can have use cases ranging from improving machinery via computer vision to offering a new product line of Internet-of-Things-style analytics to machine operators.

AI talent. To construct the instrument, we compile two new datasets on (i) the ex-ante strength of AI research in each university and (ii) firm-university hiring networks prior to 2010 to measure firms' exposure to AI-strong universities. Consistent with commercial interest in AI becoming widespread only since 2012, we show that firms' connections to AI-strong universities in 2010 were not driven by the need to hire AI-skilled workers and do not correlate with firm growth before 2010. The instrument has a strong first stage, and we show that the instrumented firm-level increase in AI investments robustly predicts firm growth between 2010 and 2018. We verify that these results are not driven by other characteristics of AI-strong universities such as strength in general computer science or overall university ranking.

We next explore the mechanisms through which AI can generate firm growth. We provide a theoretical framework in which AI can lead to firm growth through two non-mutually-exclusive channels: (i) product innovation and (ii) process innovation and reduction in operating costs. According to the first channel, AI can reduce the costs of product innovation, which improves the quality of existing products and allows firms to create new products (Klette and Kortum, 2004a; Hottman et al., 2016). Theoretically, AI can potentially reduce the costs of product innovation in several ways. First, since product development involves lengthy experimentation with uncertain benefits (Braguinsky et al., 2021), the ability of AI algorithms to quickly learn from large datasets can reduce the uncertainty of experimentation in product development and make the process of learning about promising projects more efficient. For example, at Moderna, AI algorithms have been leveraged in the development of the first COVID-19 vaccine in just 65 days, a process that would previously take years. Second, AI algorithms themselves can constitute improved products (e.g., AI-powered trading platforms). Third, AI can contribute to increased product scope by improving firms' ability to learn about customer preferences and tailor product offerings to customer tastes (Mihet and Philippon, 2019). Empirically, we find that firms with larger AI investments see increased product innovation, reflected in more product patents (i.e., patents focusing on product innovation, see Ganglmair et al., 2021) and trademarks (Hsu et al., 2021).

The second channel through which AI can stimulate growth is by increased process innovation, which would lower operating costs and improve productivity for existing products—for example, by replacing human labor for some tasks (Agrawal et al., 2019; Acemoglu and Restrepo, 2019) or by increasing operational efficiency through more efficient processes and better forecasting of the inputs into the production process (Basu et al., 2001; Farboodi and Veldkamp, 2021). Empirically, we do not find support for this second channel. AI investments are not associated with changes in sales per worker, total factor productivity, or process patents (i.e., patents focusing on process innovation). Several previous technologies have displayed labor effects consistent with task-based models of automation (e.g., Acemoglu and Restrepo, 2018). In the case of AI, detailed case studies of firms' use of AI reveal the breadth of AI applications, and empirically the labor-replacing effect does not appear to be the main driver in our analysis. Instead, the relationship between AI investments and firm growth appears to be driven by product innovation, which allows firms to expand by creating more products, thereby expanding firm scale. AI investments can help overcome capacity constraints by allowing firms to deploy more capital to produce additional products, but this comes with corresponding increases in costs.

Our final set of results speak to potential aggregate effects of AI on industry dynamics. First, we estimate the relationship between firm AI investments and firm growth within groups of firms by initial size and find that the positive relationship between AI investments and firm growth is much stronger among ex-ante larger firms, consistent with the theories where AI can increase inequality by favoring large firms with more data, which is a crucial input to AI implementation (Mihet and Philippon, 2019; Farboodi et al., 2019). We then test whether AI-fueled firm-level growth translates into industry-level growth. It is possible that the positive effects on AI-investing firms are offset or even

dominated by negative spillovers to competitors within the industry, and Basu et al. (2006) show that the use of technology can be contractionary at the aggregate level if input use declines. Nevertheless, we find that industries that invest more in AI experience an overall increase in sales and employment within the sample of Compustat firms. Finally, AI investments are associated with increased industry concentration, consistent with our finding that AI favors ex-ante larger firms with more data. This suggests that AI investments can affect industry dynamics by reinforcing winner-take-most dynamics.

Overall, we document that AI is strongly associated with higher firm growth, and this growth comes mainly from firms' use of AI technologies for product innovation. This mechanism reflects the nature of AI as a prediction technology. Predictions are essential for firms' decision-making across all aspects of operations (Farboodi and Veldkamp, 2022) and particularly in product development (Cockburn et al., 2018), which requires experimentation and learning about promising projects and customer preferences (Braguinsky et al., 2021). The ability to perform better predictions with AI can create new business opportunities. In this context, our paper offers micro-level evidence and helps to unpack the black box of where "new projects" and investment opportunities come from: new technologies like AI, which allow firms to learn better and faster, can expand the investment opportunity frontier by decreasing firms' product development costs.

### Related literature

Our paper provides one of the first pieces of systematic evidence on how investments in artificial intelligence relate to firm and industry outcomes. Recent work makes progress in examining the impact of AI technologies on firm activities in various specific settings: robo-advising (D'Acunto et al., 2019), fintech innovation (Chen et al., 2019), loan underwriting (Jansen et al., 2020; Fuster et al., 2020), financial analysts (Grennan and Michaely, 2019; Abis and Veldkamp, 2023; Cao et al., 2021), and entrepreneurship (Gofman and Jin, 2022). Acemoglu et al. (2022b) use Burning Glass job postings data to study the effect of exposure to AI technologies (based on firms' occupation structure) on labor demand.<sup>4</sup> Our comprehensive data on firm employees and data-driven approach allow us to measure actual AI investments across a wide range of industries and shed light on how AI can stimulate economic growth as a general purpose technology (Goldfarb et al., 2023).<sup>5</sup> Our empirical evidence supports this view and offers an additional insight: the mechanism through which AI fuels growth is by empowering product innovation, which has been considered a key driver of growth (e.g., Hottman et al., 2016; Argente et al., 2021). Importantly, in the case of AI, labor replacement does not appear to be the main mechanism driving firm-level effects, unlike the task-based model of automation that has been applied to previous technologies (e.g., Acemoglu and Restrepo, 2018). Instead, our results point to the main use of AI being product innovation, a less explored mechanism in the literature on technology adoption. As a prediction technology (Agrawal et al., 2019), AI creates new business opportunities by enabling firms to learn better and faster from big data. As such, our results support Cockburn et al. (2018), who argue that AI technologies can spur innovation by allowing for faster accumulation of knowledge. The product innovation channel also helps to explain the results in Rock (2019), who shows that the launch of Google's TensorFlow expedited the gain in market valuations associated with firms' exposure to AI while having null effects on productivity. Finally, our results are consistent with recent work by Hirvonen et al. (2022), who show that manufacturing robot adoption in Finland benefits firm growth mostly through increased product innovation.

<sup>4</sup> In follow-up work (Babina et al., 2022), we examine the direct relationship between firm AI investments and changes in workforce composition.

<sup>5</sup> Our firm-level measures of AI investments based on human capital are complementary to recent work that measures technology adoption using survey data (e.g., Brynjolfsson and McElheran, 2016; Acemoglu et al., 2022a).

Methodologically, our paper offers a new approach to measure firms' intangible capital based on human capital, with a specific application to capturing investments in AI. Despite ongoing efforts to measure intangibles in the U.S. at the national level (Corrado et al., 2016), most firm-level measures of intangible capital use cost items such as R&D and SG&A (e.g., Eisfeldt and Papanikolaou, 2013; Peters and Taylor, 2017; Crouzet and Eberly, 2019; Eisfeldt et al., 2020). Our methodology offers a new measure of intangibles related to technology use that is consistent across firms and sectors and can be applied to measure various forms of intangible assets, especially those based on human expertise. For example, while we focus on AI investments, we are also able to measure firm investments in robotics, non-AI information technology, and non-AI data analytics. More broadly, our method contributes to the growing literature that uses textual analysis to measure intangibles such as human capital and innovation. For example, Hoberg and Phillips (2016) analyze text of 10-K filings to create measures of firms' product portfolios, Kogan et al. (2019) construct occupation-specific indicators of technological change using patent text, Fedyk and Hodson (2023) use textual analysis to measure firms' focus on technical skills, Argente et al. (2020) map patent text to products, Babina et al. (2023a) use patent text to measure technological entrepreneurship, and Bloom et al. (2021) identify technologies using textual analysis of patents, job postings, and earnings calls.

Finally, we contribute to the recent literature on industry concentration and superstar firms (e.g., Gutiérrez and Philippon, 2017; Covarrubias et al., 2019; Grullon et al., 2019; Autor et al., 2020). Previous literature documents that larger firms adopt more IT and Internet technologies and benefit more from them (Forman, 2005; Brynjolfsson et al., 2008, 2023; Bessen, 2020). Our findings suggest that new technologies like AI have similar scale advantages and may be an important driver of the superstar firms phenomenon. This supports the hypothesis that intangible assets propel growth of the largest firms and contribute to increased industry concentration (e.g., Crouzet and Eberly, 2019). In particular, AI appears to reduce the costs of product development that are especially high for large firms (Akcigit and Kerr, 2018), allowing these firms to scale more easily. Finally, our evidence is also consistent with the lack of productivity growth among superstar firms documented in Gutiérrez and Philippon (2019).

## 2. Artificial intelligence: background and mechanisms

According to the Organisation for Economic Co-operation and Development (2019), an AI system is defined as a “*Machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions.*” We provide a brief overview of the current commercial use and key features of AI, followed by a discussion on economic mechanisms through which AI investments might lead to higher firm growth across a broad range of industries.

### 2.1. Artificial intelligence: a brief overview

Commercial applications and investments in AI have increased exponentially over the past decade. While there are no systematic data on AI investments by firms, recent estimates hover around \$140 billion globally per year.<sup>6</sup> There has also been an expansion of AI investments across industry sectors. While the tech sector was an early adopter of AI, surveys of executives indicate widespread adoption of AI technologies by firms in all industries (see [here](#) for a survey by McKinsey).

Academic research in AI has flourished for decades since John McCarthy coined the term in 1955 (McCarthy et al., 1955).<sup>7</sup> The recent explosion of commercial interest in AI in the private sector is driven by supply-side factors: rapid accumulation of data, decreasing costs of

computation, and advances in methodologies, including deep learning (Hodson, 2016). In terms of commercial applications, three key areas of AI have captured the bulk of private sector investments: machine learning, natural language processing, and computer vision.<sup>8</sup> These core techniques are united by their ability to perform high-skilled, non-routine tasks, such as prediction, detection, and classification (Agrawal et al., 2019). Their main distinction from traditional methods of data analysis consists of these techniques' ability to learn from vast quantities of high-dimensional data (including text, speech, and image data; Hauptmann et al., 2015) and significantly improve the accuracy of predictions. For example, the ImageNet challenge in 2012 led to an almost halving of image recognition error rates (relative to traditional methods), which launched large corporate interest in the computer vision space.<sup>9</sup>

AI has several key economic properties. First, AI is a prediction technology, and predictions are at the heart of decision-making under uncertainty—faced by firms in all aspects of their operations. As a result, the ability to perform better predictions with AI can create new business opportunities. Second, economists have argued that AI is a general purpose technology (GPT) and can be leveraged across different business segments and sectors to solve a wide range of business problems. Examples of GPT include the steam engine, electricity, and the Internet. Third, investments in AI center around human expertise, with complementary investments in computing technology and data infrastructure. This differs from technologies that require mainly capital investments, such as industrial robots (Benmelech and Zator, 2022). As such, AI is an intangible asset, reflecting the broader shift towards intangible capital (Mihet and Philippon, 2019). The fourth key feature of AI technologies is that they are information goods with non-rival uses: new algorithms are usually published openly and can be used simultaneously by many firms. However, the extent to which AI can benefit firms depends on who owns big data—the key input to AI technologies (Fedyk, 2016; Jones and Tonetti, 2020).

### 2.2. Artificial intelligence and firm growth: mechanisms

It is an open question whether and how investments in AI technologies benefit firms. On the one hand, as a potential general purpose technology, AI might spur economic growth. On the other hand, current attention to AI may be over-hyped (Mihet and Philippon, 2019), or AI may still be too early in the adoption cycle to have a meaningful impact on firm growth (Brynjolfsson et al., 2021).

In Online Appendix A1, we present a model with multi-product firms, and outline how AI can lead to firm growth either through process innovation or product innovation. Below, we discuss intuitions and predictions for these two non-mutually-exclusive channels.

**AI as a Driver of Product Innovation.** AI can lead to firm growth by reducing the costs of product innovation. Product innovation and the expansion of product varieties is an important mechanism for firm growth (Klette and Kortum, 2004a; Hottman et al., 2016). Product innovation can increase the product appeal and demand for existing products or enable firms to expand their product offerings. Braguinsky et al. (2021) point out that product variety and product appeal are endogenously determined through experimentation by firms, and AI can potentially facilitate the accumulation of knowledge through experimentation and reduced costs of product innovation (Bustamante et al., 2020). According to surveys of executives, enhancement of existing

<sup>8</sup> See [here](#) for a survey by Deloitte in 2018. While our focus is on AI technologies rather than automation technologies like ATMs and industrial robots, our measure does incorporate relevant recent robotics technologies (e.g., autonomous vehicles, vision-guided robots) that are highly related to computer vision and machine learning technologies.

<sup>9</sup> See [ImageNet Large Scale Visual Recognition Challenge 2012 \(ILSVRC2012\)](#).

<sup>6</sup> See the PitchBook AI & ML Emerging Tech Report 2021 [here](#).

<sup>7</sup> A brief history of AI research can be found [here](#).

products and services and the creation of new ones is the top use of AI to date (see [here](#) for a survey by Deloitte).

As a prediction technology, AI can potentially affect product innovation in several ways. First, the ability of AI algorithms to quickly analyze large datasets and learn about the underlying relationships from data can potentially reduce the uncertainty of experimentation and make the learning process more efficient, which leads to more experimentation and creation of new products (Cockburn et al., 2018). In practice, recent years show a number of ways in which AI has enabled or sped up the product innovation process. For example, AI can shorten the drug development life cycle. At Moderna, AI algorithms have contributed to the development and the production of the first dose of the COVID-19 vaccine in just 65 days, a process that would previously take years.<sup>10</sup>

Second, AI algorithms can help innovate on the quality of existing products and services by building AI models directly into products. For example, in Online Appendix A2, we offer detailed case studies of the applications of AI, which include examples such as the AI-driven trading platform DeepX at JPMorgan (which allows for faster and cheaper execution of trades) and “smart” machinery at Caterpillar (which improves machine safety and flexibility).

Third, AI can also improve product appeal by helping firms learn about customer preferences more efficiently and therefore better tailor product and service offerings to customers’ tastes and needs. When firms launch new products or expand their product variety, they face uncertainty regarding what customers want and how customer preferences might change. Using AI to analyze customer data can potentially enable firms to overcome this hurdle, providing “the right product on a hyper-individualized basis” (Hodson, 2016) and overcoming frictions in firms’ demand accumulation processes (Foster et al., 2016; Argente et al., 2021). For example, data on individual behaviors, such as web browsing and location history and other digital footprints, can enable better approximations of parameters entering individual demand functions than pure demographic information, leading to more heterogeneity in products tailored to customers with different tastes (Mihet and Philippon, 2019).

**AI as a Driver of Process Innovation and Lower Operating Costs.** AI can also lead to firm growth by reducing the costs of process innovation. Process innovation improves firms’ productivity in producing their existing products, and many prior technological innovations have aimed at lowering operating costs and improving productivity (e.g., Basu et al., 2001; Cardona et al., 2013; Acemoglu et al., 2020).

In theory, AI technologies can stimulate process innovation and productivity improvements in at least two ways. First, AI can potentially replace human labor for some tasks (Agrawal et al., 2019), cutting per-unit labor costs. Specifically, the ability of AI to aid in the decision-making process and in solving complex cognition problems has led to concerns that AI can disrupt high-skill and high-wage occupations, in contrast to previous waves of technology adoption (Webb, 2020). Second, AI can increase operational and production efficiency through better forecasting (Mihet and Philippon, 2019). Tanaka et al. (2020) present a model of firm input choice under uncertainty and costly adjustment, where forecast errors result in under- or over-investment. AI can potentially help reduce forecast errors and optimize input decisions by firms.

The potential to use AI-based forecasting for streamlining firms’ existing operations can be seen in our data. The case studies in Online

Appendix A2 highlight how AI-enabled forecasting improves firm operations across a variety of industries: for example, AI workers at JPMorgan Chase model default of non-performing loans; Caterpillar leverages AI for inventory management; and UnitedHealth uses AI to support efficient medical billing.

These two mechanisms have different empirical predictions. Product innovation predicts creation of new products, improvements in product quality, and expansion of product portfolios, whereas process innovation does not affect firms’ product portfolios. In terms of productivity, process innovation leads to lower operating costs and higher productivity, but the effect of product innovation on productivity is ambiguous. Studies on previous general purpose technologies mostly find positive effects on productivity (e.g., Fiszbein et al., 2020; Acemoglu et al., 2020), and some also show a positive effect on product innovation (e.g., Bartel et al., 2007). We will empirically examine the channels through which AI affects firm growth in Section 6.

### 3. Data

We propose a new measure of firms’ investments in AI based on their intensity of AI-skilled hiring. AI-skilled labor is a key input to AI implementation. Other inputs to AI, such as data and computing infrastructure, are complementary to AI-skilled human capital, so our human-capital-based measure allows us to capture the relative intensity of AI investments across firms.

A central challenge in the literature on the economic impact of AI is the dearth of firm-level data on AI investments. We overcome this challenge by leveraging rich datasets on firms’ employee profiles and job postings simultaneously measuring firms’ *stock* of AI workers and *demand* for AI workers. We detail each dataset and describe our sample construction process.

#### 3.1. Employment profiles from Cognism

We use employee resumes to measure the actual *stock* of AI workers at each firm. We leverage a novel dataset of approximately 535 million individual profiles provided by Cognism, an aggregator of employment profiles for lead generation and client relationship management services. Cognism obtains the resumes from a variety of sources, including publicly available online profiles, collaborations with recruiting agencies, third party resume aggregators, human resources databases of partner organizations, and direct user-contributed data.<sup>11</sup> These data are introduced and described in detail in Fedyk and Hodson (2023). While the data slightly over-represent high-skilled employees, they cover approximately 64% of the entire U.S. workforce as of 2018 and offer a representative breakdown across industries. For each employment record listed by the individual, we see the start and end dates, the job title, the company name, and the job description. Individuals may also list their patents, awards, and publications. Cognism’s AI Research department leverages techniques from machine learning and natural language processing, including named entity disambiguation and graph-based modeling methods, to further enrich the resume data by normalizing job titles and occupations, associating employees with functional divisions and teams within each firm, and identifying institutions, degrees, and majors from education records.<sup>12</sup>

We match employer names in the Cognism data to the company names in the Compustat data. Fedyk and Hodson (2023) provide further details on the procedure as applied to the resume data. The matching

<sup>10</sup> See [here](#) and [here](#), where Dave Johnson, Moderna’s VP of Informatics, Data Science, and AI, explains how Moderna was able to develop a COVID vaccine so quickly: “We very purposely designed all this infrastructure that we think of as an AI factory, in order to rapidly deliver algorithms from concept to production, to enable our scientists to leverage the power of AI in their daily jobs. [...] That allows our scientists to design novel mRNA constructs, use AI algorithms to optimize them, and then order them from our high throughput preclinical scale production line.”

<sup>11</sup> The processing of all profiles is compliant with the applicable GDPR and CCPA regulations.

<sup>12</sup> The data snapshot is from July of 2021. Following Tambe et al. (2020), we only use the years through 2018, because the lag in workers updating their resumes could otherwise add noise to our measures.

of individual resumes to firm entities is performed dynamically to account for acquisitions and divestitures. Of the 657 million US-based person-firm-year employment records between 2007 and 2018, 120 million (18%) are matched to U.S. public firms (rather than private firms or non-commercial sectors). This is consistent with approximately 26% of overall U.S. employment being accounted for by publicly listed firms (Davis et al., 2006). The sample of 120 million person-firm-years matched to U.S. public firms is comprised of 19 million distinct individual employees.

### 3.2. Job postings from Burning Glass

The second dataset we use covers over 180 million job postings in the United States in 2007 and 2010–2018. The dataset is provided by Burning Glass Technologies (BG in short) and draws from a rich set of sources. BG examines more than 40,000 online job boards and company websites, aggregates the job postings data, parses them into a systematic, machine-readable form, and creates labor market analytic products. The company employs a sophisticated deduplication algorithm to avoid double counting vacancies that post on multiple job boards. BG data contain detailed information for each job posting, including job title, job location, occupation, and employer name. Importantly, the job postings are tagged with thousands of specific skills standardized from the open text in each job opening. The main advantages of the BG dataset are the breadth of its coverage and the rich detail of the individual job postings. The dataset captures the near-universe of jobs posted online and covers approximately 60–70% of all vacancies posted in the U.S., either online or offline. Hershbein and Kahn (2018) provide a detailed description of the BG data and show that their representativeness is stable over time at the occupation level. Burning Glass job postings complement our main Cognism resume dataset in two ways. First, we take advantage of the detailed taxonomy of skills in the job postings data to empirically identify highly AI-relevant skills, as we detail in Section 4.1. Second, Burning Glass job postings data are widely accessible to the academic community; by validating them against Cognism resume data, we show that, in absence of matched employer-employee data, job postings can serve as a valid proxy for firms' technological investments.

We focus on jobs with non-missing employer names and at least one required skill. About 65% of job postings have employer information and 93% of job postings require at least one skill.<sup>13</sup> We also drop job postings that are internships. We then match the employer firms in the remaining job postings to Compustat firms. This step is necessary to aggregate job postings to the firm level and merge with other firm-level variables. We perform a fuzzy matching between firm names in BG and Compustat after stripping out common endings such as “Inc” and “L.P.” For observations that do not match exactly on firm name, we manually assess the top ten potential fuzzy matches by looking at the firm name, industry, and location. Out of 112 million job postings with non-missing employer names and skills, 42 million (38%) are matched to Compustat firms. This slightly over-represents employees of publicly listed firms, which constitute just over one fourth of U.S. employment in the non-farm business sector (Davis et al., 2006).

### 3.3. Additional data sources

We merge the Cognism resume data and the Burning Glass job postings data with several additional data sources. We collect commuting-zone-level wage and education data from the Census American Community Surveys (ACS), industry-level wages and employment data from the Census Quarterly Workforce Indicators (QWI), and academic publications from the Open Academic Graph (described in detail in Ap-

<sup>13</sup> Job postings with missing employer names are primarily those listed on recruiting websites that mask the employers' identities.

pendix A). Firm-level operational variables (e.g., sales, employment, market value) come from Compustat.

## 4. Methodology and descriptive evidence

We use the Cognism resume data to construct a human-capital-based measure of firm-level AI investments. To do this, we first leverage job postings data to learn the most AI-related skills directly from the data; we then focus on the most empirically-AI-related skills and identify them in the resume data. Finally, we aggregate the worker-level data to the firm level by calculating the share of each firm's employees who are AI-skilled.

### 4.1. AI investments from job postings (Burning Glass)

We take advantage of the detailed information on required skills in the job postings data to propose a new data-driven methodology for identifying AI-related skills. Other work relies on pre-specified lists of key terms,<sup>14</sup> which is likely to suffer from both Type I (incorrectly labeling tangentially-related employees as AI-related) and Type II (missing real AI skills that did not make the initial dictionary) errors due to the arbitrariness of the list of keywords. This is especially relevant in a quickly-evolving domain such as AI, where new emerging skills can easily be missed. Our methodology circumvents these challenges by learning the AI-relatedness of each of approximately 15,000 unique skills directly from the job postings data, based on their empirical co-occurrence (within required lists of skills across job postings) with unambiguous core AI skills. We then aggregate the skill-level measure to the job level by generating a continuous measure of AI-relatedness for each job posting, from which we can classify jobs into AI-skilled and non-AI-skilled.

To measure the AI-relatedness of each skill, we calculate the skill's co-occurrence with Artificial Intelligence (AI) and its three main sub-fields: machine learning (ML), natural language processing (NLP), and computer vision (CV).

$$w_s^{AI} = \frac{\# \text{ of jobs requiring skill } s \text{ and (ML, NLP, CV or AI in required skills or in job title)}}{\# \text{ of jobs requiring skill } s}$$

Intuitively, this measure captures how correlated each skill  $s$  is with the core AI skills. For example, the skill “Tensorflow” has a value of 0.9, which means that 90% of job postings with Tensorflow as a required skill also require one of the core AI skills or contain one of the core AI skills in the job title. Hence, a “Tensorflow” requirement in a job posting is highly indicative of that job being AI-related. On the other hand, the AI-relatedness measure of the skill “Microsoft Office” is only 0.003. We list the skills with the highest AI-relatedness measures in Online Appendix Table A1.

We define the job-posting-level AI-relatedness measure  $\omega_j^{AI}$  for a given job posting  $j$  as the mean skill-level measure  $w_s^{AI}$  across all skills required by job posting  $j$ . We transform the continuous AI measure into a binary indicator by defining each job posting  $j$  as AI-related if the measure  $\omega_j^{AI}$  is above 0.1, a threshold that captures the full range of AI-related technical jobs while minimizing false positives based on manual inspection of the data. The firm-level measure  $Share_{f,t}^{AI}$  is then defined as the fraction of job postings by firm  $f$  in year  $t$  that are AI-related (i.e.  $\omega_j^{AI} > 0.1$ ).<sup>15</sup> We use a discrete classification for ease of interpretability

<sup>14</sup> For example, Hershbein and Kahn (2018) classify jobs as requiring cognitive abilities if any listed skills include at least one of the following terms: “research,” “analy-,” “decision,” “solving,” “math,” “statistic,” or “thinking.” Similar bag-of-words approaches with pre-specified search terms are used to identify AI-related employees (e.g., Alekseeva et al., 2021; Acemoglu et al., 2022b).

<sup>15</sup> Throughout our empirical analyses, we focus on jobs that are matched to Compustat firms. Online Appendix Figure A5 plots the share of all job postings and the share of AI-related job postings that are matched to Compustat in each

and consistency with the resume-based measure in Section 4.2, but we show in Section 5.1 that the results are robust to: (i) applying alternative cut-offs (e.g., 0.05 and 0.15) and (ii) using the continuous measure  $\omega_j^{AI}$  aggregated to the firm level. This job-postings-based measure of AI investments provides a secondary measure to our main measure based on the stock of employees, obtained from the resume data as described in the next subsection.

Online Appendix Table A2 provides examples of AI-related and non-AI-related job postings. For each job, the continuous AI measure is the average AI-relatedness of all required skills. Our measure enables us to capture a wide range of AI-related jobs, from data scientists to speech recognition scientists to autonomous vehicle engineers. While many AI-related jobs are data scientists and similar data-analysis-related jobs, our measure differentiates data-analysis jobs specifically related to AI (job postings numbered 6–10) from data-analysis jobs that are not specific to AI and that focus on more traditional statistical methods (job postings numbered 11–15). In addition, we further ensure that our measure is not picking up general programming or statistics skills not specific to AI by showing (in Section 5.1) the robustness of our results to manually refining our measure. In particular, we screen out skills that represent general programming languages (e.g., Python) or statistics (e.g., linear regression) and only keep skills that relate specifically to AI, including AI methodology or algorithms (e.g., supervised learning) and AI software (e.g., Tensorflow). This process, curated by the AI-trained personnel at the AI for Good Foundation, categorizes the 700 skills that have an AI-relatedness measure above 0.05 and are required in at least 50 job postings into “narrow” and “broad” AI skills. This refinement mainly leaves out skills with relatively lower AI-relatedness measures and empirically has little effect on the results.<sup>16</sup>

#### 4.2. AI investments from resumes (cognism)

For our main measure of firms’ AI investments, we identify the employees in the Cognism resume data whose job positions directly involve AI. We begin with the set of 67 keywords in Online Appendix Table A1, which are skills with the highest skill-level AI-relatedness measures based on job postings data. We search for these terms in each employment record of each individual in the resume data to see whether: (i) that job (role and description) directly includes any of the identified AI terms; (ii) the individual obtained any patents during that year or the two following years (to account for the time lag between the work and the patent grant) with these AI terms; and (iii) the individual has any publications or awards during that year or the following year that include the identified AI terms. If any of these conditions are met, then that person at that firm in that year is classified as an AI-related employee. For example, jobs with titles such as “senior *machine learning* developer” or job descriptions such as “develop chatbots using Python with *Tensorflow* and *deep learning* models” are identified as AI jobs.

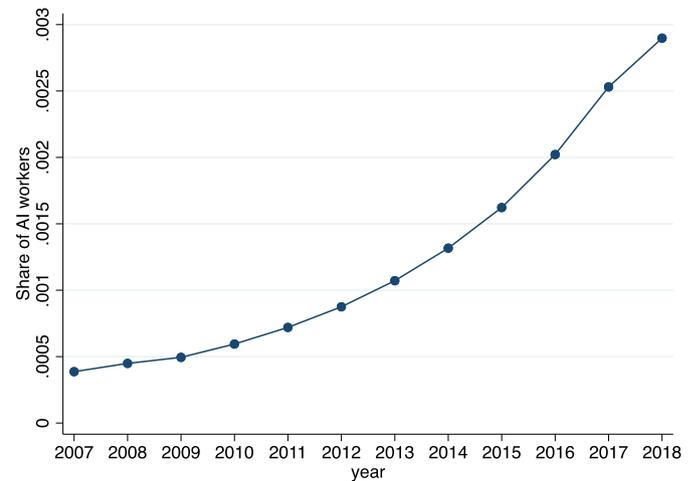
After classifying each individual at each point in time, we use the number of AI-related employees and the number of total employees at each firm in each year to compute the percentage of employees of that firm in that year who are classified as AI-related. Given that our empirical analyses focus on U.S.-listed firms, our firm-level measure focuses on the employees who are based in the U.S.

#### 4.3. Summary statistics and validation

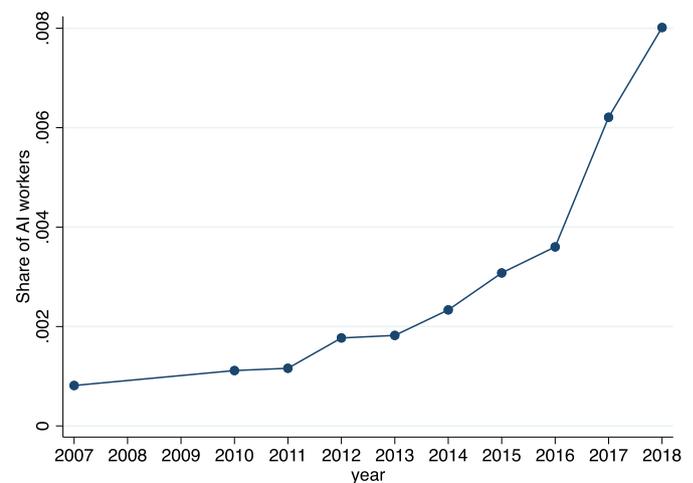
We examine both of our constructed measures of AI investments, confirm that they display intuitive properties, and discuss how our

year. Although publicly listed firms constitute 38% of all job postings, they account for approximately half of all AI-related job postings. This suggests that, on average, publicly-listed firms hire more AI workers than private firms.

<sup>16</sup> For example, among the 50 skills with the highest AI-relatedness measures, 49 are classified as narrow AI skills (the single exception is “statsmodels,” a Python package for general statistical analysis).



(a) Resume data (Cognism)



(b) Job posting data (Burning Glass)

**Fig. 1.** Time Series of AI Investments. This figure shows the time series of the two measures of AI investments for U.S. public firms. Panel (a) shows the fraction of all employees in a given year who are classified as holding AI-related positions in the Cognism resume data from 2007 to 2018. Panel (b) reports the fraction of job postings in the Burning Glass data (with job-level continuous AI measure above 0.1) for 2007 and 2010–2018.

resume data help address potential limitations of measuring AI investments through job postings. Validating our novel measure is challenging, given the lack of existing firm-level measures of AI investments. However, we show that our measure displays a number of intuitive properties, captures specifically AI investments, and does not suffer from biases such as firms investing in AI by acquiring AI startups.

First, we document that both measures—based on resumes and job postings—display a natural rise over time, increasing more than seven-fold from 2010 to 2018. Panel (a) of Fig. 1 shows a rapid increase in the fraction of employees whom we classify as AI-related: this fraction starts at 0.04% in 2007 and reaches 0.29% in 2018. Panel (b) shows analogous patterns in the job postings data: the fraction of AI-skilled job postings starts out at 0.1% in 2010, rises rapidly over time (with the increase speeding up from 2014 to 2018), and peaks at 0.8% in 2018. There is substantial heterogeneity in the growth in AI-skilled labor across individual firms, which provides the variation needed to examine the relationship between AI investments and firm outcomes. For the entire sample of public firms, while a median firm sees an increase of 0% (0%) in the resume-based (job-postings-based) measure,

this increase is 0.35% (1.33%) at the 90th percentile, 0.62% (2.99%) at the 95th percentile, and 2.22% (8.11%) at the 99th percentile.

It is helpful to put the incidence of AI-skilled workers among U.S. employees into perspective. While, as expected, AI workers constitute a relatively small fraction of total employment, skyrocketing demand for AI skills and correspondingly high salaries that they command—on the order of millions of dollars for prominent AI-researchers (Gofman and Jin, 2022)—suggest that AI-skilled workers are similar to other specialized, high-skilled, high-wage jobs. For example, in terms of the technological and innovative nature of their work, AI-skilled workers could be compared to inventors. Inventors also tend to be highly paid and represent around 0.13–0.24% of the U.S. workforce, which is similar in prevalence to AI workers.<sup>17</sup> Overall, while AI workers form a small fraction of the overall workforce, it is helpful to contextualize their impact against that of executives (Bertrand and Schoar, 2003) and patent inventors (Kline et al., 2019), both of whom are similarly small, high-skilled groups of employees that can nonetheless disproportionately affect firm outcomes.

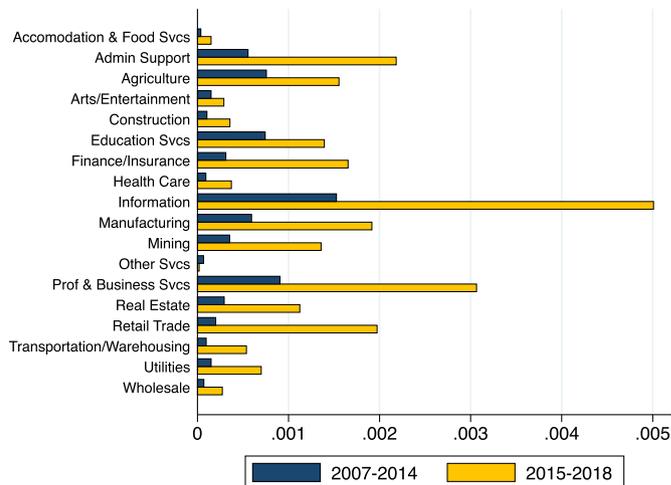
Second, we document that the increase in AI jobs displays an intuitive distribution across industries. Panel (a) of Fig. 2 plots the average share of AI-related workers in the resume data for public firms in each of the 2-digit NAICS sectors, separately for the years 2007–2014 and 2015–2018. Panel (b) repeats the same analysis for the share of AI-related job postings. The figure highlights that the share of AI-skilled jobs (job postings) is highest in the “Information” sector, growing from 0.15% (0.57%) in the early years of 2007–2014 to 0.50% (1.68%) in the later period of 2015–2018. However, almost all sectors see a meaningful increase, supporting the notion that AI is a general purpose technology (Goldfarb et al., 2023). The ability of our measures to pick up AI investments in a broad cross-section of economic sectors highlights a key advantage of our human-capital-based approach.

Third, AI investments correlate positively with increased R&D expenditures. For example, changes in the resume-based share of AI workers from 2010 to 2018 display a correlation of 0.27 with changes in log R&D expenditures over the same time period, controlling for industry fixed effects. The pattern of AI-investing firms increasing research and development (R&D) expenditures supports the notion that AI investments involve a great deal of experimentation with applying the new technology (Braguinsky et al., 2021).

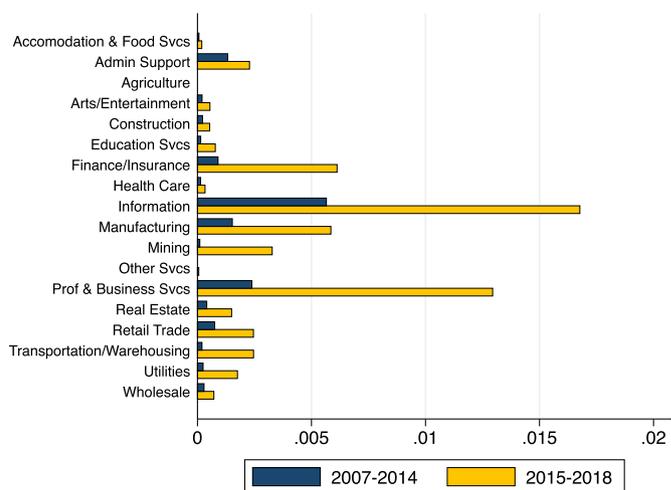
Fourth, digging deeper into the skills and jobs with the highest AI-relatedness measures according to our methodology, we observe that our measure is indeed capturing the essence of AI-skilled hiring by firms. The required skills in job postings with the highest AI-relatedness measures, presented in Online Appendix Table A1, are highly AI-specific skills, such as “Tensorflow” and “Random Forests,” while general data-analysis-related skills have low AI-relatedness measures: for example, the measure is equal to 0.04 for “Data Modeling” and 0.03 for “Quantitative Analysis.” Similarly, Online Appendix Table A3 shows that the job titles associated with the highest job-level measures of AI-relatedness are all very relevant postings such as “Artificial Intelligence Engineer” (average AI-relatedness measure of 0.497), “Senior Data Scientist - Machine Learning Engineer” (0.394), and “AI Consultant” (0.369). Since we do not use information contained in job titles of job postings to identify AI-related skills and jobs, these patterns provide additional validation that our measure captures relevant AI positions.

As a further validation, it is worth noting the geographic locations of the identified AI jobs. We aggregate firm-level AI hiring of Compustat firms to the commuting zone level and link the commuting-zone-level changes in the share of AI workers from 2010 to 2018 to 2010 commuting zone characteristics from the Census American Community Survey.

<sup>17</sup> These estimates come from the USPTO patent data (Babina et al., 2023a), where 0.13% is the share of U.S. workers who file patents in a given year, and 0.24% is the share of U.S. workers who file patents over a three-year period.



(a) Resume data (Cognism)



(b) Job posting data (Burning Glass)

**Fig. 2.** AI Investments by Industry Sector. This figure presents the average share of AI jobs at the industry level, based on the sample of U.S. public firms. For each sector (based on 2-digit NAICS industry codes), we compute the fraction of AI-related employees in the Cognism resume data (in Panel (a)) and the average share of AI-related job postings in the Burning Glass data (with job-level continuous AI measure above 0.1) across all job postings (in Panel (b)). The statistics are computed for firms in each industry sector across two sub-periods: 2007–2014 and 2015–2018. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Online Appendix Figure A6 shows that there is significant variation in AI investments across commuting zones. Online Appendix Figure A7 shows a strong positive relationship between the change in the share of AI workers from 2010 to 2018 and the average wage or the fraction of college-educated workers at the commuting zone level in 2010. These patterns are intuitive, given that AI employees tend to be high-skilled technologically-oriented workers, and contrast with investments in robotics, which concentrate in areas with larger shares of manufacturing employment (Acemoglu et al., 2020).

Finally, in Table 1, we observe a high correlation between our measure of AI investments based on the novel resume data and on the job postings data. The resume data provide several advantages over the job postings data for measuring firms’ AI investments. First, the resume data measure actual hiring of AI-skilled labor. For example, if a firm is unable to fill AI-related vacancies, the job postings measure will overstate that

**Table 1**

Correlations between Job-posting-based and Resume-based AI Measures. This table reports, for each year from 2010 to 2018, the Spearman rank correlations between three pairs of firm-level variables: (i) the absolute number of AI job postings in Burning Glass against the absolute number of AI employees in resumes from Cognism; (ii) the fraction of employees classified as AI-related in the two datasets; and (iii) the fraction of AI employees in Cognism against the average continuous AI measure in Burning Glass. Panel 1 shows raw correlations, and Panel 2 displays correlations conditional on the 2-digit NAICS industry sector fixed effects and the baseline controls all measured as of 2010: firm-level characteristics (log sales, cash/assets, firm age, and log number of resumes and job postings), log industry wage, and characteristics of the commuting zones where the firms are located (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers). All correlations are computed over the cross-section of firms with at least 20 U.S. jobs in the Cognism resume data in each year of the sample.

Panel 1: Raw Correlations			
Year	Correlations between: Numbers of AI jobs	Fractions of AI Jobs	Cognism fraction & BG continuous measure
2010	0.320	0.272	0.374
2011	0.341	0.288	0.390
2012	0.338	0.291	0.388
2013	0.424	0.363	0.447
2014	0.468	0.410	0.484
2015	0.474	0.405	0.496
2016	0.503	0.421	0.499
2017	0.564	0.474	0.531
2018	0.574	0.484	0.538
Panel 2: Correlations Conditional on Baseline Controls			
Year	Correlations between: Numbers of AI jobs	Fractions of AI Jobs	Cognism fraction & BG continuous measure
2010	0.825	0.650	0.470
2011	0.822	0.622	0.476
2012	0.801	0.583	0.487
2013	0.784	0.569	0.498
2014	0.757	0.526	0.551
2015	0.729	0.467	0.513
2016	0.702	0.475	0.501
2017	0.687	0.507	0.510
2018	0.670	0.502	0.513

firm's investments in AI. Second, human capital on-boarded through acquisitions is captured by the resume data, where employees of acquisition targets are counted as employees of the acquirer subsequent to the acquisition.<sup>18</sup> For these reasons, we focus on the resume-based measure of AI investments in our baseline specifications. Nevertheless, we find high correlations between the two measures and consistent results throughout the remainder of the paper, which alleviates some of the concerns about job postings data in measuring firms' AI talent. This consistency suggests that, in the absence of matched employer-employee data, our methodology offers a good proxy for firms' actual AI hiring using the more widely accessible job postings data.

#### 4.4. Firm-level determinants of AI investments

We consider the firm-level determinants of AI investments and document that larger firms and firms with higher cash reserves tend to invest in AI more aggressively.

Our focus is on understanding the *use* of AI technologies by a wide range of firms, rather than the invention of new AI tools. For that reason, we exclude firms in the tech sector from our main empirical analyses in this and the following sections.<sup>19</sup> Our main regression sample is comprised as follows. In 2010, there are 3,735 U.S.-listed public firms that have non-missing industry codes, positive sales and employ-

<sup>18</sup> Industry reports estimate that 90% of firms' investments in AI are internal, with only 10% coming from acquisitions (Bughin et al., 2017).

<sup>19</sup> We exclude firms in two 2-digit NAICS sectors: 51 ("Information") and 54 ("Professional, Scientific, and Technical Services"). In later analyses, we confirm that the main effects of AI spurring firm-level growth are also present in these industries. A complementary analysis of the impact of AI on specifically AI-inventing firms is provided by Alderucci et al. (2020).

ment, and are not in the tech sector. Among these firms, 2,668 are matched to Cognism,<sup>20</sup> and we further restrict to firms with at least 20 U.S. jobs in both 2010 and 2018 to ensure good coverage of the firm's workforce, which leaves us with 1,993 firms.

In Table 2, we examine which ex ante firm characteristics predict future growth in firm-level AI investments. For each measure of AI investments, we estimate the following specification:

$$\Delta ShareAIWorkers_{i,[2010,2018]} = \beta FirmVariable_{i,2010} + SectorFE + \epsilon_i, \quad (1)$$

where  $\Delta ShareAIWorkers_{i,[2010,2018]}$  denotes the change in the share of firm  $i$ 's AI-related employees from 2010 to 2018. All regressions include 2-digit NAICS industry fixed effects ( $SectorFE$ ). Here and throughout all subsequent analyses, the  $\Delta ShareAIWorkers_{i,[2010,2018]}$  variables are standardized to have a mean of zero and a standard deviation of one to aid in economic interpretation.  $FirmVariable_{i,2010}$  represents one of the ex ante firm characteristics of interest measured as of 2010: log firm sales in column 1, the ratio of cash to total assets (Cash/Assets) in column 2, the ratio of R&D expenditures to sales (R&D/Sales) in column 3, revenue total factor productivity (TFP)<sup>21</sup> in column 4, log markup measured as the log of the ratio of sales to cost of goods sold following De Loecker et al. (2020) in column 5, Tobin's Q defined as market value

<sup>20</sup> Firms that are not matched to Cognism tend to be either ADRs that do not have U.S. employees or smaller firms with few employees.

<sup>21</sup> We use standard methodology to calculate revenue TFP as the residual from regressing log real sales on log employment and log capital, controlling for firm fixed effects and year fixed effects:  $\log y_{it} = \mu_i + \mu_t + \alpha'_s \log(l_{it}) + \alpha'_k \log(k_{it-1}) + \epsilon_{it}$ . The regression is estimated using OLS separately for each industry. The capital stock is constructed using the perpetual inventory method. The TFP measure is specific to Cobb-Douglas production functions, while sales per worker measure labor productivity for more general production functions.

**Table 2**

Firm-level Determinants of Resume-based AI Investments. This table reports the coefficients from regressions of cross-sectional changes in AI investments by U.S. public firms (in non-tech sectors) from 2010 to 2018 on the following ex-ante firm characteristics measured in 2010: log sales in column 1, cash/assets in column 2, R&D/sales in column 3, revenue TFP in column 4, log markup measured following De Loecker et al. (2020) in column 5, Tobin's Q in column 6, market leverage in column 7, return on assets (ROA) in column 8, and firm age in column 9. The dependent variable is the growth in the share of AI workers from 2010 to 2018 using the resume data from Cognism. Regressions are weighted by the number of Cognism resumes in 2010. All specifications control for the 2-digit NAICS industry sector fixed effects. The dependent variable is normalized to have a mean of zero and a standard deviation of one. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta$ Share of AI Workers, 2010–2018									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log Sales 2010	0.109*** (0.038)									0.175*** (0.033)
Cash/Assets 2010		3.986*** (1.242)								3.233*** (0.832)
R&D/Sales 2010			3.734 (2.313)							1.062 (1.745)
Revenue TFP 2010				1.386 (1.038)						0.567 (0.532)
Log Markup 2010					0.418 (0.312)					0.147 (0.177)
Tobin's Q 2010						0.335** (0.153)				0.158 (0.118)
Market Leverage 2010							-0.973 (0.690)			0.213 (0.301)
ROA 2010								1.409** (0.656)		0.164 (1.127)
Firm Age 2010									-0.003 (0.004)	-0.004** (0.002)
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Adj R-Squared	0.151	0.309	0.172	0.148	0.142	0.215	0.129	0.129	0.118	0.418
Observations	1,731	1,731	1,731	1,731	1,731	1,731	1,731	1,731	1,731	1,731

of assets divided by book value of assets in column 6, market leverage measured as total debt divided by market value in column 7, return on assets (ROA) measured as the ratio of net income plus interest expense to assets in column 8, and firm age in column 9. Column 10 includes all variables in a multivariate specification. We winsorize all continuous variables at 1% and 99% to limit the influence of outliers, although we confirm in untabulated analyses that, empirically, our results are little changed by the winsorization. To account for differences in precision in the measurement of AI investments across firms with different numbers of available observations, the estimating equation is weighted by each firm's number of resumes in 2010.<sup>22</sup>

The results reported in Table 2 highlight that ex ante larger firms experience higher growth in AI investments. A one-standard-deviation increase in log sales in 2010 (which equals 2.1) corresponds to the share of AI workers increasing by 23% of the standard deviation from 2010 to 2018, significant at the 1% level. In addition, firms with higher starting Cash/Assets also see greater investments in AI, in both univariate and multivariate regressions. Younger firms also see greater investments in AI, but the relationship is not robust as we only observe this in the multivariate regression. By contrast, R&D/Sales, revenue TFP, markups, firm valuation (Tobin's Q), market leverage, and return on assets do not robustly predict future AI investments. In all further regressions, we control for the ex-ante firm characteristics that predict firm AI adoption—size, cash/assets, and firm age. Online Appendix Table A4 shows that the patterns for firm-level demand for AI talent measured with Burning Glass job postings data are consistent with the results using Cognism resume data, reinforcing the high correlations documented in Table 1.

<sup>22</sup> Since the numbers of worker resumes are correlated with firm size, this weighting scheme also roughly weights firms in accordance to their contribution to the economy.

## 5. AI investments and firm growth

We next document that firms investing in AI technologies grow faster in sales, employment, and market value. We consider and rule out alternative explanations for this result, including reverse causality (e.g., firms on faster growth trajectories invest more in AI) and omitted variables (e.g., concurrent investments in other technologies or demand shocks drive both firm growth and AI investments). Finally, we introduce a novel instrumental variable strategy to address remaining concerns.

### 5.1. Long-differences results

We begin the analysis by examining whether firms that invest in AI see faster growth from 2010 to 2018. As is standard in settings with slow-moving processes, such as technological progress (e.g., Acemoglu and Restrepo, 2020), our primary specification is a long-differences regression of changes in firm outcomes from 2010 to 2018 on changes in AI investments proxied by the share of AI workers. This strategy is especially well-suited for our setting because AI investments are gradual over time (with 70% of firms onboarding AI workers over a span of multiple years), with effects that may not be immediate. By taking first differences in independent and dependent variables, the long-differences specification ensures that time-invariant firm characteristics do not drive the results.<sup>23</sup> In Table 3, we report the estimates from the following regression:

<sup>23</sup> A potential concern with the long-differences specification is that it requires AI-investing firms to be present at the beginning (2010) and the end of the sample (2018), which might introduce a sample selection bias. In Section 5.2, we use the full panel dataset, which does not condition on firms being present over the entire sample period, and find similar effects. Moreover, the industry-level analysis in Section 7 shows that the inclusion of entering and exiting firms has little effect on our results, which would not be the case if the composition of firms changed in an important way for our analysis.

**Table 3**

AI Investments and Firm Growth: Long-differences Estimates Using the Resume-based AI Measure. This table reports the coefficients from long-differences regressions of firm growth from 2010 to 2018 on the contemporaneous firm-level changes in AI investments among U.S. public firms (in non-tech sectors). We consider three measures of firm growth: changes in log sales (columns 1 and 2), changes in log employment (columns 3 and 4), and changes in log market value (columns 5 and 6). The main independent variable is growth in the share of AI workers (based on the resume data) from 2010 to 2018, standardized to mean zero and standard deviation of one. Regressions are weighted by the number of Cognism resumes in 2010. All specifications control for the 2-digit NAICS industry sector fixed effects. Columns 2, 4, and 6 also include the baseline controls all measured as of 2010: firm-level characteristics (log sales, cash/assets, firm age, and log number of resumes), log industry wage, and characteristics of the commuting zones where the firms are located (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers). Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta$ Log Sales		$\Delta$ Log Employment		$\Delta$ Log Market Value	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta$ Share AI Workers	0.202*** (0.069)	0.195*** (0.069)	0.239** (0.097)	0.181** (0.086)	0.240** (0.093)	0.223** (0.086)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
Adj R-Squared	0.221	0.422	0.237	0.405	0.247	0.364
Observations	1,052	1,052	1,052	1,052	1,009	1,009

$$\Delta FirmVariable_{i,[2010,2018]} = \beta \Delta ShareAIWorkers_{i,[2010,2018]} + Controls'_{i,2010} \gamma + SectorFE + \epsilon_i, \quad (2)$$

where the main independent variable,  $\Delta ShareAIWorkers_{i,[2010,2018]}$ , captures the change in the share of AI workers based on the resume data at firm  $i$  from 2010 to 2018, standardized to have a mean of zero and a standard deviation of one. As in Section 4.4, this analysis focuses on firms in non-tech sectors.  $SectorFE$  are 2-digit NAICS industry fixed effects. In columns 1, 3, and 5 we include only industry fixed effects to examine the unconditional relationship between changes in AI investments and firm growth. In columns 2, 4, and 6, we include a rich set of controls that are all measured at the start of the sample period in 2010: (i) the initial firm-level characteristics that predict changes in AI investments in Section 4.4 (log sales, cash/assets, firm age) and the log of the firm's total Cognism employment<sup>24</sup>; (ii) characteristics of the commuting zones (CZ) where the firms are located (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers); and (iii) the log industry-average wage.<sup>25</sup> Out of the 1,993 non-tech firms in the sample in Table 2, 1,472 firms have positive sales and employment in 2018, which are necessary to calculate the dependent variables. We further restrict the sample to firms with non-missing control variables throughout, to keep the sample composition stable. This results in a sample of 1,052 firms. The results of the regressions without controls are similar when estimated on the entire available sample. Summary statistics on key variables for the main regression sample are provided in Online Appendix Table A5.

In columns 1 and 2 of Table 3, the dependent variable is the firm-level change in log sales from 2010 to 2018. Changes in AI investments are associated with a significant and economically meaningful increase in sales growth: a one-standard-deviation increase in the share of AI workers over an eight-year period corresponds to an additional 19.5%

<sup>24</sup> We control for log employment to address the concern that the share of AI workers may be more volatile in firms with fewer total workers. This control ensures that the variation in the share of AI workers is between firms with similar total employment in Cognism but different numbers of AI workers.

<sup>25</sup> When firms span multiple commuting zones, we use the commuting zone with the most BG job postings, which restricts the sample in the Cognism regression analysis to firms that are also matched to the Burning Glass data. The results are similar in magnitude and economic significance if we only include firm-level controls enumerated in list (i).

growth in sales. In columns 3 and 4, we find a positive association with employment growth of a similar magnitude to the relationship with sales. This suggests that AI is not yet displacing firms' workforces, at least on net, although we do not rule out the reallocation of labor across different job functions or tasks.<sup>26</sup> Columns 5 and 6 show that firms investing in AI also see increases in their stock market valuations: a one-standard-deviation increase in the share of AI workers is associated with a 22–24% increase in the firm's market value.<sup>27</sup> It is worth noting that the inclusion of firm-level, location-level, and industry-level controls in even columns (all measured at the start of the sample period in 2010) generally has little effect on the estimated coefficients. This makes it unlikely that the results are driven by ex-ante omitted firm characteristics (Altonji et al., 2005).

The magnitude of the effects in Table 3 is economically meaningful—on the order of about 2% increase in *annual* sales growth per one-standard-deviation increase in the share of AI workers. Our results provide initial evidence that hiring AI-skilled labor can have a strong positive relationship with firm growth. In this context, our results are consistent with prior evidence that certain key, high-skilled employees—including chief executives, inventors, and entrepreneurs—can have a disproportionate effect on firm outcomes. It's also important to caution that the correct interpretation of our results is not that investing in AI workers, without changing any other inputs, will lead to an additional 2% in annual sales growth. Instead, the main mechanism through which AI appears to stimulate firm growth is product innovation (as we show in Section 6), and the adoption of AI likely necessitates investments in other inputs for product development. For this reason, we do not interpret the results as directly measuring the return on investment (ROI) of AI investments, but rather as broader evidence of how AI technologies can help stimulate firm growth.

The positive relationship between increases in AI investments and firm growth is ubiquitous across different sectors of the economy, reinforcing the notion that AI is a general purpose technology. Online Appendix Table A7 displays the results from regressing changes in log

<sup>26</sup> In untabulated analyses, we confirm that the results are similar when using changes in employee counts in the Cognism resume data for the outcome variable instead of Compustat employment.

<sup>27</sup> Market value is defined as total assets (*at*), minus the book value of common equity (*ceq*), plus the market value of common equity (*prcc\_c* times *csho*). Online Appendix Table A6 shows that AI investments are also associated with increases in unadjusted and risk-adjusted buy-and-hold returns.

**Table 4**

Heterogeneous Relationship between AI Investments and Firm Growth by Initial Firm Size Using the Resume-based AI Measure. This table reports the coefficients from long-differences regressions of firm growth from 2010 to 2018 on contemporaneous changes in AI investments among U.S. public firms (in non-tech sectors), separately for each tercile of initial firm size. Firms in each 2-digit NAICS industry sector are divided into terciles based on employment in 2010. We consider three measures of firm-level growth for the dependent variable: changes in log sales (columns 1 and 2), changes in log employment (columns 3 and 4), and changes in log market value (columns 5 and 6). The main independent variable is the growth in the share of AI workers (based on the resume data) from 2010 to 2018, standardized to mean zero and standard deviation of one. Regressions are weighted by the number of Cognism resumes in 2010. All specifications control for the 2-digit NAICS industry sector interacted with initial firm size tercile fixed effects. Columns 2, 4, and 6 also include the baseline controls all measured as of 2010: firm-level characteristics (log sales, cash/assets, firm age, and log number of resumes), log industry wage, and characteristics of the commuting zones where the firms are located (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers). Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. We report the t-statistics and p-values from the t-tests for the difference between the coefficient of  $\Delta \text{Share AI Workers} \times \text{Size Tercile 1}$  and the coefficient of  $\Delta \text{Share AI Workers} \times \text{Size Tercile 3}$  at the bottom of the table. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta$ Log Sales		$\Delta$ Log Employment		$\Delta$ Log Market Value	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta$ Share AI Workers*Size Tercile 1	0.046** (0.023)	-0.001 (0.021)	0.041** (0.018)	-0.021 (0.031)	0.057* (0.030)	0.008 (0.051)
$\Delta$ Share AI Workers*Size Tercile 2	0.219*** (0.054)	0.183*** (0.050)	0.217*** (0.048)	0.157*** (0.055)	0.210*** (0.045)	0.170*** (0.050)
$\Delta$ Share AI Workers*Size Tercile 3	0.223*** (0.077)	0.204*** (0.076)	0.260** (0.105)	0.182** (0.091)	0.260** (0.102)	0.242** (0.099)
NAICS2*Size tercile FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
Adj R-Squared	0.248	0.419	0.253	0.414	0.240	0.347
Observations	1,044	1,044	1,044	1,044	1,002	1,002
T-test statistic	3.7	6.8	3.7	5.3	4.4	8.8
T-test p value	0.054	0.010	0.054	0.022	0.038	0.003

sales and log employment on the change in the share of AI workers, separately for the largest 2-digit NAICS sectors: (i) Manufacturing, (ii) Wholesale and Retail Trade, (iii) Finance, and (iv) the remaining non-tech sectors. While we exclude tech sectors from our main analysis, we find that AI also has a positive relationship with growth for tech firms (see Online Appendix Table A8). Overall, we observe that investments in AI are associated with economically significant increases in firms' operations, and these effects are meaningful across key economic sectors.

However, the benefits from AI investments are not evenly distributed across the firm size distribution. Table 4 shows the relationship between changes in AI investments and firm growth, across terciles of firms by employment in 2010 (within the firm's 2-digit NAICS sector), controlling for initial size and sector-by-size-tercile fixed effects. The relationship between firm AI investments and firm growth is monotonically increasing in the firm's initial size. The stronger positive relationship between changes in AI investments and growth among the ex ante larger firms is consistent with big data and AI technologies having scale effects that favor large firms, which accumulate large amounts of data as a by-product of their economic activity (Farboodi et al., 2019; Farboodi and Veldkamp, 2022). Akcigit and Kerr (2018) highlight that larger firms face constraints on their ability to scale due to higher costs of new product innovation. The results in Table 4 suggest that AI may provide a channel through which large firms can combat barriers to innovation and scale by leveraging their data assets. For example, biotech firms that have accumulated large troves of proprietary samples of molecular compounds are able to leverage AI tools to obtain an advantage over competitors.<sup>28</sup> On the other hand, the benefits of AI are not limited to a few prominent firms: Online Appendix Table A9 shows that dropping firms in the top 1% or 5% of the size distribution has little effect on the full-sample results.

#### 5.1.1. Robustness

We show that our results are robust to using alternative constructions of the AI measure and address several identification concerns regarding the effects of AI investments on firm growth.

**Alternative measures of AI investments.** We use the resume-based share of AI workers as our main measure of AI investments because the resume data address two important potential measurement concerns regarding job postings data: (i) the job-postings-based measure captures only firms' demand for AI talent and not their actual ability to hire; and (ii) firms might obtain AI expertise through acquisitions, which would not be reflected in job postings but is captured in the Cognism resume data, which reflect actual employees, including those onboarded through acquisitions.

Nevertheless, in Online Appendix Tables A10–A12, we show that our results remain similar when measuring AI using the Burning Glass job postings data following the methodology discussed in Section 4. This confirms that the job-postings-based measure of AI investments, which is highly correlated with the resume-based measure, can also be used to assess firm-level benefits of AI. The results are robust to using the share of job postings with continuous AI-relatedness measure above various cutoffs or the firm-level average continuous AI-relatedness measure.

We next address concerns regarding the skewness of investments in AI. Instead of using standardized measures of AI investments, we also consider dummy variables indicating whether firms' AI investments are in the top 10 percent or 25 percent of the distribution. Online Appendix Table A13 shows that the top 25% of firms in terms of AI investments grow their sales by 31% more than other firms and the top 10% of firms grow their sales by 54% more than other firms between 2010 and 2018.

Finally, while our measure is centered on internal AI investments, our rich resume data allow us to also consider whether firms' use of external AI solutions might affect the interpretation of our results. Even external AI software requires internal data management and implementation guidance by AI-skilled workers to be effective (Fedyk, 2016), and industry reports underscore that AI-skilled labor is the most critical input to successful deployment of AI programs. Nevertheless, we process individual job descriptions and job titles in our resume data for any mention of external AI software (including IBM Watson, IPSoft Amelia,

<sup>28</sup> See [here](#) for the PitchBook AI & ML Emerging Tech Report 2021.

Symphony, AyasdiAI, Salesforce Einstein, and about a hundred other key AI-powered solutions) to construct a proxy for firms' reliance on external AI solutions.<sup>29</sup> In Online Appendix Table A14, we confirm that our results are robust to directly including this proxy in our overall measure of AI investments.

**Confounding factors.** The estimates above may not reflect the effects of AI investments if contemporaneous changes at AI-investing firms lead to both increased investment in AI and higher firm growth. In Section 5.3, we use an instrumental variable strategy to address omitted variable bias. Below, we discuss the main confounding factors and provide evidence that our estimates are not driven by these confounders.

First, we address the possibility that investments in AI are correlated with investments in other technologies (e.g., IT). We leverage our detailed data to develop measures of investments in non-AI technologies that parallel the measure of AI investments: for each firm, we measure the percentage of job postings in each year requiring IT-, robotics- or data-related skills that are not specific to AI. In Online Appendix Table A15, we control for growth in: (i) investments in (non-AI) IT, (ii) investments in robots, (iii) investments in non-AI data skills (e.g., "Data Cleaning"), and (iv) investments in non-AI-related data analytics (e.g., "SAS"). Panel 1 uses the resume-based AI measure, and Panel 2 uses the job-postings-based AI measure. The estimated relationship between growth in AI investments and firm growth remains similar with the addition of these controls, confirming that the documented effects on firm growth are specifically driven by AI rather than by other technologies.

Second, AI-investing firms may happen to experience positive demand shocks or higher growth trajectories, leading to a positive bias in our estimates. In Online Appendix Table A16, we control for detailed industry fixed effects to absorb industry-specific shocks. The coefficient on the changes in AI investments remains stable, and the standard error increases as more granular industry controls absorb more of the variation. Moreover, in Online Appendix Tables A17 and A18, we confirm that the results are robust to controlling for (i) past industry-level and firm-level growth in the decade before our sample period (from 2000 to 2008), (ii) Tobin's Q as of 2010, which proxies for the firm's future growth opportunities, and (iii) state fixed effects, which control for growth opportunities and other potential omitted variables at the state level. Finally, Online Appendix Table A19 estimates a predictive regression of firm growth during the later part of our sample (2015–2020) on growth in AI investments during the earlier part of the sample (2010–2015).<sup>30</sup> The estimates are qualitatively and quantitatively similar to those in Table 3, with milder magnitudes corresponding to the shorter estimation period (growth from 2015 to 2020 rather than from 2010 to 2018), pointing against reverse causality driving our results. In the next two sections, we use a dynamic specification and an IV strategy to further address the potential bias from unobserved shocks.

## 5.2. Dynamic relationship between firm AI investments and firm growth

We augment our long-differences specification by estimating the dynamics of firm growth following AI investments. This analysis not only offers evidence against reverse causality concerns and AI-investing firms being on differential growth trajectories prior to AI investments, but also elucidates the lag between AI investments and their realized effects.

<sup>29</sup> The Cognism resume data are especially well-suited to capture the use of external technological solutions, given Cognism's emphasis on developing "technographic data" (defined by Cognism as "the technologies that the employee or company is using"). Cognism advertises these data for two purposes: (i) enhancing technology providers' targeted marketing of their products and (ii) improving individual firms' understanding of which technologies are used by their competitors.

<sup>30</sup> In these regressions, we can use firm growth estimated through 2020, because this specification does not require firm AI data (which end in 2018) to go beyond 2015.

We use firm-level panel data to estimate firm growth dynamically around AI investments in a distributed lead-lag model, which allows for continuous variation in the treatment variable (Stock and Watson, 2015; Aghion et al., 2020). This specification is especially well-suited to our setting, because firms tend to invest in AI on a continuous basis, rather than make lumpy investments in a single year, which precludes us from examining dynamic effects in a standard event-study framework with discontinuous treatment (e.g., before and after a lumpy investment).<sup>31</sup> The standard distributed lead-lag model is specified as:

$$Y_{it} = \sum_{k=-2}^5 \delta_k \Delta \text{ShareAIWorkers}_{i,t-k} + \mu_i + \lambda_{it} + \theta_{st} + \epsilon_{it} \quad (3)$$

where  $\Delta \text{ShareAIWorkers}_{i,t-k}$  is the annual change in the share of AI workers from year  $t-k-1$  to year  $t-k$ , normalized to have a mean of zero and a standard deviation of one, and  $Y_{it}$  is either log sales or log employment in year  $t$ . We include firm fixed effects  $\mu_i$  to absorb firm-specific time-invariant factors, and 2-digit NAICS industry-year fixed effects  $\lambda_{it}$  and state-year fixed effects  $\theta_{st}$  to control for industry-specific and state-specific trends. Each lead-lag coefficient  $\delta_k$  captures the cumulative response of the outcome variable in year  $t$  to AI investments in year  $t-k$ , holding fixed the path of AI investments in all other years. As such, specification (3) incorporates both immediate and delayed responses of firm size to firms' AI investments.<sup>32</sup> The estimated coefficients for the leads can be used as a pre-trend test: if firms investing in AI are on similar growth trends as other firms prior to AI investments,  $\delta_k$  with  $k < 0$  should be statistically indistinguishable from zero.<sup>33</sup>

Fig. 3 reports the coefficients from the lead-lag regressions. The top panel shows that sales increase following AI investments, but not immediately—it takes two to three years for firms to realize the benefits from AI investments. The cumulative effect of a one-standard-deviation increase in annual AI investments on log annual sales is 1.5%–2% and remains steady five years out. This is consistent with the long-differences estimates in Section 5.1, where a one-standard-deviation increase in AI investments is associated with a 19.5% increase in sales over eight years. The bottom panel shows that AI investments are associated with a similar increase in firms' employment. Importantly, there is no evidence of pre-trends in either outcome variable: conditional on the controls we include, firms that invest more in AI in any given year show comparable sales and employment paths in prior years and start diverging only afterwards. This provides additional evidence that our results are not capturing the reverse effect of firm growth on AI investments or the effect of omitted variables placing AI-investing firms on differential growth trajectories, helping to bolster a causal interpretation of our main results.

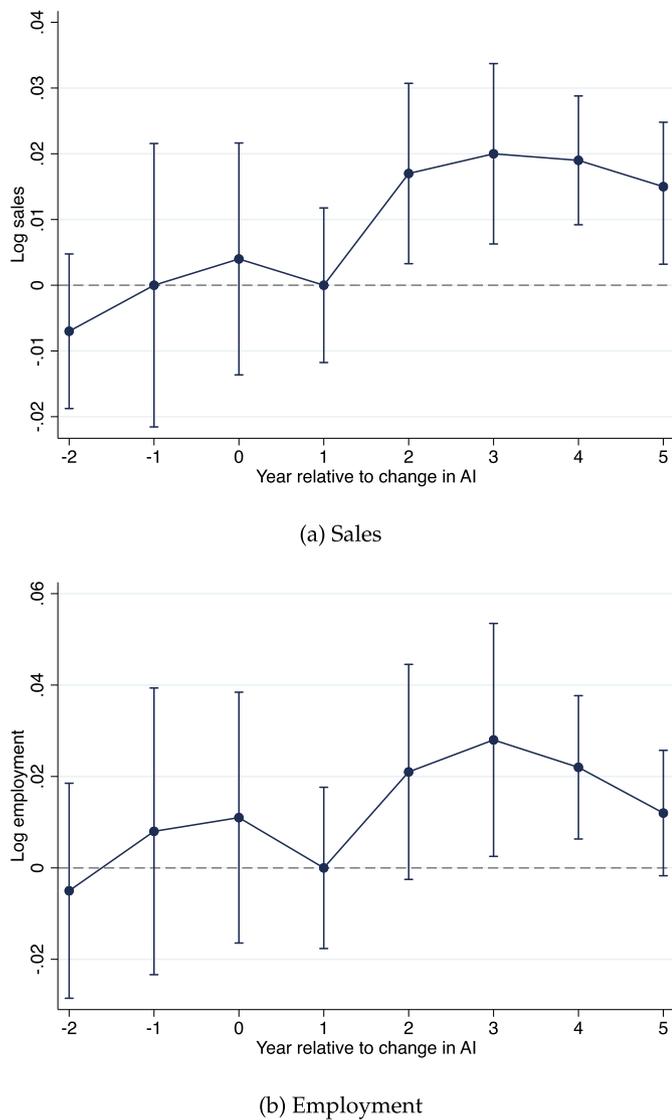
## 5.3. Instrumental variable estimates

In this section, we instrument firm-level changes in AI investments using firms' exposure to the supply of AI talent from U.S. universities. This helps to isolate variation in firms' AI investments that comes from the supply of AI labor, mitigating potential bias from demand shocks

<sup>31</sup> The percentage of AI-investing firms that only invest in a single year is 29.5%, compared to 70.6% for robotics (Humlum, 2019).

<sup>32</sup> For each firm-year observation of sales or employment between 2010 and 2016, we consider five lags and two leads, so that we estimate the cumulative impact of AI investments on firm growth from two years before the investments to five years after the investments. Since the data on AI investments end in 2018, we include only two leads to keep all firm-year observations up to 2016. We obtain similar results when including only one lead or no leads at all.

<sup>33</sup> It is worth noting that, given that the independent variables in this distributed lead-lag model are changes in continuous AI investments instead of period dummies as would be the case in a standard event-study framework, we cannot normalize the estimates to an exact zero for any given period.



**Fig. 3.** AI Investments and Firm Growth Over Time. This figure plots the coefficients from the distributed lead-lag model. The dependent variable is annual log sales in Panel (a) and log employment in Panel (b). The independent variable is the annual change in the share of AI workers in the Cognism resume data, standardized to have a mean of zero and a standard deviation of one. Regressions include firm-level sales (or employment) observations between 2010 and 2016 and control for firm fixed effects, 2-digit NAICS industry-by-year fixed effects, and state-by-year fixed effects. Regressions are weighted by the number of workers in the Cognism resume data. The vertical bars indicate 95% confidence intervals. Standard errors are clustered at the 5-digit NAICS level.

driving both firms' AI investments and growth. The scarcity of AI talent is a key constraint to firms' adoption of AI technologies, and the IV estimates are directly informative about the treatment effects of policies targeting the supply of AI labor and relaxing the constraints of AI adoption (e.g., funding university AI research and training AI-skilled human capital, Babina et al., 2023c). At the same time, academic research in AI has been ongoing for much longer than commercial interest in AI. As a result, firms' preexisting connections to AI-strong universities offer an arguably exogenous source of variation in firms' access to the supply of AI talent during the 2010s boom in commercial interest in AI.

In particular, we instrument firm-level changes in AI investments using variation in firms' ex-ante exposure to the supply of AI talent from universities that are historically strong in AI research. The core idea is that the scarcity of AI-trained labor is one of the most important constraints to firms' AI adoption (e.g., CorrelationOne, 2019), and

universities that are historically strong in AI research have been able to train more AI-skilled graduates in recent years, enabling firms that typically hire from those universities to more readily attract AI talent. This intuition is motivated by the evidence of the spillover effects from academic science to industrial research and technology adoption in the U.S. (Furman and MacGarvie, 2007). Since commercial interest in AI became widespread only around 2012, we argue (and offer empirical support) that firms' connections to AI-strong universities in 2010 were not driven by the need to hire AI-skilled workers, especially for the sample of non-tech firms that are the focus of this paper. To construct the instrument, we compile two datasets on: (i) the ex-ante strength of AI research in each university, and (ii) firm-university hiring networks. To the best of our knowledge, there is no comprehensive historical data on either of these two aspects. We briefly discuss the construction of both datasets below, with a more comprehensive discussion in Appendix A.

To identify universities strong in AI research before 2010, we use data from the Open Academic Graph (OAGv2), which provides the most comprehensive openly available repository of scholarly work since 1870 (Tang et al., 2008; Sinha et al., 2015). We match 689 research institutions in the National Science Foundation's Higher Education Research and Development Survey (HERDS) to researchers in the OAGv2 and work with the field experts at the AI for Good Foundation to identify AI-related publications. We classify each AI researcher based on the share of AI publications in that researcher's overall portfolio, and we classify universities as AI-strong if their number of AI researchers is at the top of the distribution over 2005–2009 (i.e., in the top 5%).

We construct the firm-university hiring networks by leveraging our resume data to observe the universities granting the degrees of each firm's employees.<sup>34</sup> For the firm-university hiring networks to provide the necessary variation for our instrumental variable strategy, different firms need to hire from different sets of universities, and these networks need to be persistent over time. Our data show evidence of both: each firm tends to concentrate its hiring in a small number of universities, and ex-ante networks (i.e., which universities each firm hired from before 2010) strongly predict the universities from which firms hire after 2010 (see Appendix Table A.1).

We define our instrument for each firm  $i$  as:  $IV_i = \sum_u s_{iu}^{2010} AI\ strong_u$ , where  $s_{iu}^{2010}$  is the share of STEM workers in firm  $i$  in 2010 who graduated from university  $u$ , and  $AI\ strong_u$  equals one if university  $u$  is identified as an AI-strong university based on pre-2010 publications.<sup>35</sup> We use pre-2010 publications to measure AI-strong universities because research in AI has flourished in universities long before 2010, while commercial use of AI began after 2010. Thus, post-2010 publications may be affected by the demand for AI from the firms that universities are connected to, whereas pre-2010 publications are an arguably exogenous measure of universities' AI talent.

A key concern with our instrument is that AI-strong universities may also be different in other ways. First, if universities that are strong in AI research are also strong in the broader field of computer science (CS), producing more CS-skilled graduates, this might affect firm outcomes through channels other than AI investments. Second, if AI-strong universities are ranked highly in general, then high-quality firms—who are

<sup>34</sup> Aggregated to the university-year level, our resume data cover, on average, 59% of all degrees conferred by each university according to data in the Integrated Postsecondary Education Data System (IPEDS), and the number of fresh graduates in the resume data is highly correlated with the total number of degrees conferred (correlation = 0.73) in the IPEDS data. Confirming the relevance of our measure of AI-strong universities, Appendix Fig. A.2 shows that the increase in AI-trained graduates during the 2010s was much more pronounced in ex-ante AI-strong universities than in non-AI-strong universities.

<sup>35</sup> We use firm-university hiring networks based on STEM workers to account for potential segmentation in firms' hiring networks, where business employees may be hired from different universities than technically-skilled employees. However, empirically, firm-university hiring networks constructed from all workers yield similar results.

**Table 5**

AI Investments and Firm Growth: IV Estimates Using the Resume-based AI Measure. This table estimates the relationship between AI investments and firm growth from 2010 to 2018 for U.S. public firms (in non-tech sectors), where firm AI investments are instrumented with ex-ante firm-level exposure to AI-skilled graduates from AI-strong universities (see the definition of the instrument in Section 5.3 and the details of instrument construction in Appendix A). The independent variable is the change in the share of AI workers from 2010 to 2018 based on the resume data. Regressions are weighted by the number of Cognism resumes in 2010. The independent variable and the instrument are standardized to mean zero and standard deviation of one. We consider changes in log sales in columns 1 to 4, log employment in columns 5 to 8, and log market value in columns 9 to 12. All specifications control for the 2-digit NAICS industry sector fixed effects and ex-ante exposure to universities that are strong in computer science research as well as top 10 universities. Columns 2–4, 6–8, and 10–12 also control for the baseline controls all measured as of 2010: firm-level characteristics (log sales, cash/assets, firm age, and log number of resumes), log industry wage, and characteristics of the commuting zones where the firms are located (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers). Columns 3–4, 7–8, and 11–12 additionally control for firm-level changes in log sales and log employment from 2000 to 2008. Columns 4, 8, and 12 add state fixed effects. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. The first-stage F-statistics of the instrument are reported for all specifications. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta$ Log Sales				$\Delta$ Log Employment				$\Delta$ Log Market Value			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\Delta$ Share AI Workers	0.360*** (0.101)	0.510*** (0.125)	0.467*** (0.137)	0.317** (0.152)	0.466*** (0.154)	0.712*** (0.224)	0.584*** (0.194)	0.330** (0.163)	0.449*** (0.129)	0.529*** (0.160)	0.470*** (0.172)	0.360* (0.184)
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
University Control	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Baseline Controls	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Control Pre-trend	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
State FE	N	N	N	Y	N	N	N	Y	N	N	N	Y
F Statistic	12.7	19.3	15.7	18.7	12.7	19.3	15.7	18.7	12.8	19.5	15.5	19.3
Observations	1,001	1,001	777	773	1,001	1,001	777	773	963	963	753	746

likely to grow regardless of AI—may hire from AI-strong universities purely by hiring from highly-ranked universities. To address these concerns, we control for firms' ex-ante exposure to CS-strong universities and top-ranked universities. In particular, we construct analogous measures of firms' ex-ante exposure to CS-strong and top-10 universities:  $\sum_u s_{iu}^{2010} CSstrong_u$  and  $\sum_u s_{iu}^{2010} Top10_u$ , where  $CSstrong_u$  is the average pre-2010 share of (non-AI) CS researchers at university  $u$ , and  $Top10_u$  equals one if a university is among the top 10 universities ranked by the U.S. News & World Report.

Another potential concern regarding our instrument is that firms that anticipated the surge in demand for AI may have started building their connections to AI-strong universities before 2010, making firm-university hiring networks in 2010 endogenous to firms' demand for AI-trained students. However, this runs counter to the lack of both commercial interest in AI by firms and AI-skilled graduates by universities prior to 2010 (see Appendix Fig. A.2). Moreover, we confirm empirically that firms connected to AI-strong universities in 2010 did not increase their share of hired fresh graduates from those universities from 2005 to 2010 (see Appendix Table A.2).

Appendix Table A.3 shows the first-stage results. We control for industry fixed effects and exposure to CS-strong and top-10 universities in all columns. We sequentially add (i) baseline controls (firm-, industry-, and commuting-zone-level controls) in column 2, (ii) pre-period firm sales and employment growth between 2000 and 2008 to address unobservable firm characteristics that might simultaneously drive firms' growth trajectories and their hiring of AI workers in column 3, and (iii) state fixed effects to control for local labor market characteristics that might drive both firms' AI hiring and their growth in column 4. The instrument has a strong first stage with F-statistics above 10 for all specifications and close to 20 when all controls are included. Online Appendix Figure A8 plots the reduced form relationship between the instrument and firm growth. For all three outcome variables (sales, employment, and market value), we see a strong positive relationship between firms' ex-ante exposure to AI-strong universities and subsequent growth. In contrast, Appendix Table A.4 shows that firms that are more exposed to AI-strong universities are not growing faster before 2010. This is consistent with the exclusion restriction that the instrument only affects firm growth through firms' AI investments after 2010.

Table 5 presents the 2SLS estimates. The results show a robust and significant effect of AI investments on sales (columns 1–4), employment

(columns 5–8), and market value (columns 9–12). When all controls are included, a one-standard-deviation increase in AI investments leads to a 32% increase in sales, a 33% increase in employment, and a 36% increase in stock market valuations. The magnitudes are about 50% larger than OLS estimates, which may be due to measurement error or a negative bias from firms with lower opportunity costs of innovation and lower growth prospects investing in AI.<sup>36</sup> It is important to note, however, that the OLS and IV coefficients are not statistically different. This suggests that the difference between the point estimates could also be driven by estimation error. We conduct a series of robustness checks of our IV results in Online Appendix Tables A20–A23, discussed in detail in Appendix A.

## 6. Mechanisms

We examine the drivers of AI-fueled firm growth by considering the two non-mutually-exclusive mechanisms detailed in Section 2. We document that AI-investing firms are able to significantly increase their product innovation and find no evidence of reductions in operating costs.

### 6.1. AI as a driver of product innovation

AI can contribute to firm growth via product innovation by (i) facilitating the creation of new and improved products and (ii) increasing product scope through improved tailoring of products to customer tastes. To explore this empirically, we need firm-level data on products and services, which are challenging to obtain, especially across different sectors. We overcome this challenge by using three proxies for firms' product innovation. We examine product innovation using our main long-differences specification from Equation (2).

First, we examine whether AI-investing firms experience increases in trademarks, which are registered whenever new products or services are ready for commercialization and therefore offer a good proxy for the creation of new products and services (Hsu et al., 2021). Columns 1 and 2 in Table 6 present the results looking at the changes in firms' USPTO trademarks against growth in their AI investments, showing that

<sup>36</sup> Consistent with this notion, controlling for log sales in 2010 increases the coefficient on AI investments, and the coefficient on the log sales control itself is negative.

**Table 6**

AI Investments and Product Innovation Using the Resume-based AI Measure. This table reports the coefficients from long-differences regressions of the changes in measures of product innovation from 2010 to 2018 on the contemporaneous changes in AI investments by U.S. public firms (in non-tech sectors). The dependent variables are the change in  $\log(1+\text{number of trademarks})$  in columns 1 and 2; the change in  $\log(1+\text{number of product patents})$  in columns 3 and 4; and the change in the product mix in columns 5 and 6. Product patents are patents with over 50% of the claims being product claims, following the categorization in Ganglmair et al. (2021). The change in the product mix is measured as the sum of annual changes from 2010 to 2018, where each annual change is the angle between the two word vectors indicating firms' product offerings in that year and the previous year (the word vectors are constructed as in Hoberg et al. (2014)). The main independent variable is the resume-based measure of the growth in the share of AI workers from 2010 to 2018, which is standardized to mean zero and standard deviation of one. Regressions are weighted by the number of Cognism resumes in 2010. All specifications control for the 2-digit NAICS industry sector fixed effects. Columns 2, 4, and 6 also include the baseline controls all measured as of 2010: firm-level characteristics (log sales, cash/assets, firm age, and log number of resumes), log industry wage, and characteristics of the commuting zones where the firms are located (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers). Columns 1 and 2 control for the log number of trademarks from 2009 to 2010, and columns 3 and 4 control for the log number of patents from 2005 to 2010. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta$ Log Number of Trademarks		$\Delta$ Log Number of Product Patents		Change in Product Mix	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta$ Share AI Workers	0.144** (0.065)	0.134* (0.078)	0.221*** (0.035)	0.239*** (0.031)	0.149*** (0.036)	0.059 (0.038)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
Observations	550	550	619	619	958	958

AI-investing firms significantly increase their trademark portfolios. A one-standard-deviation increase in the share of AI workers is associated with approximately 13% more trademarks.<sup>37</sup> Second, columns 3 and 4 reveal a similar relationship between AI investments and the number of product patents, which are patents specifically focusing on product innovations.<sup>38</sup> While trademarks are registered with the creation of new products, product patents reflect both new product creation and innovations in the quality of existing product lines. We find that a one-standard-deviation increase in the share of AI workers over eight years corresponds to about 23% increase in the number of product patents.

Finally, we build a measure of changes in firms' product mix based on the self-fluidity measure in Hoberg et al. (2014). Using firm 10K filings, Hoberg et al. (2014) take the cosine similarity between word vectors describing a firm's product offerings in two adjacent years to measure the extent to which the firm's product offerings changed in a given year. These changes reflect both the creation of new products and the tailoring of existing products to evolving consumer tastes.<sup>39</sup> In column 5, we find that growth in AI investments is associated with increased changes in firms' product mix, but the relationship becomes statistically insignificant when adding additional controls in column 6.

<sup>37</sup> The dependent variable is the change in  $\log(1 + \text{number of trademarks})$  from 2010 to 2018, so that the regression takes into account firms with zero trademarks in either 2010 or 2018. The results are also robust to using the inverse hyperbolic sine transformation (i.e.,  $\ln(x + \sqrt{1 + x^2})$ ). The regression sample is smaller than our baseline sample, because not all public firms file trademarks (we include firms with at least one trademark in 2009–2018).

<sup>38</sup> See Ganglmair et al. (2021) for the methodology to distinguish between product patents and process patents. The regression sample is smaller than our baseline sample, because not all public firms file patents, and we only include firms with at least one patent during 2005–2018. The dependent variable is the change in  $\log(1 + \text{number of product patents})$  from 2010 to 2018.

<sup>39</sup> We use the same word vectors as Hoberg et al. (2014) and construct our measure as follows: for each year, we calculate the angle between the two word vectors indicating firms' product offerings in that year and the previous year. For example, the measure equals 0 if the product offerings remain exactly the same and  $\pi/2$  if the product offerings change completely. We sum up the angle of each year over eight years from 2010 to 2018 to measure the total change in firms' product portfolios from 2010 to 2018.

Online Appendix Table A24 shows that the results on all three product innovation measures remain similar when using the job-postings-based AI measure.

Table 7 shows that the instrumented AI investments also have a positive relationship with the number of trademarks, the number of product patents, and the change in product offerings (although not always significant). Overall, the results point towards firms utilizing AI to expand product variety and customization, consistent with surveys of corporate executives, who highlight product improvement and creation as the main use of AI so far (see here).

## 6.2. AI as a driver of lower operating costs

We next test whether the increase in firm growth from AI investments could reflect AI technologies lowering firms' operating costs, increasing firm-level productivity, and improving process innovation. First, in columns 1 to 4 of Table 8 we look at costs directly by considering how growth in firms' AI investments relates to changes in costs of goods sold (COGS) and operating expenses. AI investments are associated with increases in costs that are similar in magnitude to increases in firm sales, suggesting that AI is not associated with lower average operating costs.

Second, columns 5 to 8 of Table 8 consider two measures of productivity: sales per worker (i.e., labor productivity) and revenue TFP. The relationship between AI investments and both productivity measures is consistently insignificant. The lack of growth in labor productivity is consistent with the results in Section 5 that AI investments predict similar increases in sales and employment, challenging the view that the primary effect of AI so far is to replace jobs.<sup>40</sup> Furthermore, in columns

<sup>40</sup> It is worth noting that both sales per worker and revenue TFP are revenue-based measures of productivity and may not fully reflect actual physical productivity. For example, sales per worker and revenue TFP may provide downward-biased estimates of actual productivity changes if quantities produced increase to such an extent that lower prices are charged (Foster et al., 2008; Garcia-Marin and Voigtländer, 2019; Caliendo et al., 2020). To consider this possibility, in untabulated analyses we find that there are no changes in AI-investing firms' markups.

**Table 7**

AI Investments and Product Innovation: IV Estimates Using the Resume-based AI Measure. This table estimates the relationship between AI investments and product innovation from 2010 to 2018 for U.S. public firms (in non-tech sectors), where firms' AI investments are instrumented with ex-ante firm-level exposure to AI-skilled graduates from AI-strong universities (see the definition of the instrument in Section 5.3 and the details of instrument construction in Appendix A). The independent variable is the change in the share of AI workers from 2010 to 2018 based on the resume data. The independent variable and the instrument are standardized to mean zero and standard deviation of one. Regressions are weighted by the number of Cognism resumes in 2010. We consider the change in  $\log(1+\text{number of trademarks})$  in columns 1 to 4, the change in  $\log(1+\text{number of product patents})$  in columns 5 to 8, and the change in the product mix in columns 9 to 12. Product patents are patents with over 50% of the claims being product claims, following the categorization in Ganglmair et al. (2021). The change in the product mix is measured as the sum of annual changes from 2010 to 2018, where each annual change is the angle between the two word vectors indicating firms' product offerings in that year and the previous year, following Hoberg et al. (2014). All specifications control for the 2-digit NAICS industry sector fixed effects and ex-ante exposure to universities that are strong in computer science research as well as top 10 universities. Columns 2–4, 6–8, and 10–12 also include the baseline controls all measured as of 2010: firm-level characteristics (log sales, cash/assets, firm age, and log number of resumes), log industry wage, and characteristics of the commuting zones where the firms are located (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers). Columns 3, 4, 7, 8, 11, and 12 additionally control for firm-level changes in log sales and log employment from 2000 to 2008. Columns 4, 8, and 12 add state fixed effects. Columns 1–4 control for the log number of trademarks from 2009 to 2010, and columns 5–8 control for the log number of patents from 2005 to 2010. Standard errors are clustered at the 5-digit NAICS industry level, and reported in parentheses. The first-stage F-statistics of the instrument are reported for all specifications. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta$ Log Number of Trademarks				$\Delta$ Log Number of Product Patents				Change in Product Mix			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\Delta$ Share AI Workers	0.223 (0.217)	0.519 (0.347)	0.640** (0.321)	0.648** (0.278)	0.242 (0.168)	0.260 (0.201)	0.481*** (0.175)	0.430* (0.238)	0.185 (0.286)	0.187 (0.315)	0.281 (0.322)	0.314 (0.325)
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
University Control	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Baseline Controls	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Control Pre-trend	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
State FE	N	N	N	Y	N	N	N	Y	N	N	N	Y
F Statistic	11.8	10.5	8.5	15.9	11.5	23.8	36.9	34.6	13.1	22.2	16.0	21.8
Observations	528	528	435	426	586	586	479	469	932	932	725	717

**Table 8**

AI Investments and Operating Costs Using the Resume-based AI Measure. This table reports the coefficients from long-differences regressions of changes in firm operating costs and firm productivity from 2010 to 2018 on contemporaneous changes in AI investments by U.S. public firms (in non-tech sectors). The main independent variable is the change in the share of AI workers (based on the resume data) from 2010 to 2018, standardized to mean zero and standard deviation of one. We look at two measures of operating costs: log COGS in columns 1 and 2 and log operating expenses in columns 3 and 4. We consider two measures of productivity: log sales per worker (columns 5–6) and revenue TFP (columns 7–8). Revenue TFP is the residual from regressing log revenue on log employment and log capital (constructed using the perpetual inventory method), with separate regressions for each industry sector. In columns 9 and 10, the dependent variable is the change in  $\log(1+\text{number of process patents})$ , where process patents are patents with over 50% of the claims being process claims, following the categorization in Ganglmair et al. (2021). Regressions are weighted by the number of Cognism resumes in 2010. All specifications control for the 2-digit NAICS industry sector fixed effects. Columns 2, 4, 6, and 8 also include the baseline controls all measured as of 2010: firm-level characteristics (log sales, cash/assets, firm age, and log number of resumes), log industry wage, and characteristics of the commuting zones where the firms are located (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers). Columns 9 and 10 also control for the log number of patents before 2010. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta$ Log COGS		$\Delta$ Log Operating Expense		$\Delta$ Log Sales per Worker		$\Delta$ Revenue TFP		$\Delta$ Log Number of Process Patents	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\Delta$ Share AI Workers	0.195*** (0.052)	0.177*** (0.053)	0.206*** (0.066)	0.191*** (0.065)	-0.036 (0.028)	0.013 (0.022)	-0.049 (0.046)	-0.003 (0.037)	-0.010 (0.039)	0.014 (0.054)
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y	N	Y	N	Y
Adj R-Squared	0.213	0.390	0.237	0.420	0.176	0.322	0.222	0.358	0.700	0.747
Observations	1,052	1,052	1,052	1,052	1,052	1,052	977	977	619	619

9 and 10, we bring another proxy for efficiency gains that complements revenue-based measures of productivity: process patents, which reflect process innovations and potential improvements in efficiency. We find a zero relationship between AI investments and process innovation, in contrast to the large increase in product patents documented in Table 6. Online Appendix Tables A25 and A26 show similarly insignificant relationships with productivity measures and process patents using the job-postings-based AI measure and the instrumented AI investments, respectively.

Overall, we find that AI technologies benefit firms through product innovations rather than through reductions in operating expenses

or improvements in productivity. This contrasts with previous general purpose technologies, such as electricity, which led to rapid productivity gains (Fizsbein et al., 2020). This juxtaposition of results is consistent with Acemoglu et al. (2022a), who use U.S. Census data and find no correlation between artificial intelligence and labor productivity, but find positive productivity effects for other technologies such as robotics and specialized software.

One potential explanation for the lack of productivity growth is the productivity J-curve proposed by Brynjolfsson et al. (2021). In particular, productivity growth from investing in general purpose technologies may be initially underestimated, because capital and labor are spent to

**Table 9**

AI Investments and Changes in Industry Growth and Concentration Using the Resume-based AI Measure. This table reports the coefficients from industry-level long-differences regressions of the changes in industry sales, employment, and concentration on contemporaneous changes in industry-level AI investments. All industry-level variables are calculated for all firms in Compustat (regardless of whether they are in our main regression sample in Table 3 or not). Each observation is a 5-digit NAICS industry, and (as in our main analysis) we exclude tech sectors. The independent variable is the change in the share of AI workers (based on the resume data) from 2010 to 2018, standardized to mean zero and standard deviation of one. Regressions are weighted by the total (industry-level) number of Cognism resumes. The dependent variables are the changes, from 2010 to 2018, in log total sales in columns 1 and 2, log total employment in columns 3 and 4, the Herfindahl-Hirschman Index (HHI) in columns 5 and 6, and the market share of the top firm in an industry in columns 7 and 8. All specifications control for the 2-digit NAICS industry sector fixed effects. Regressions in columns 2, 4, 6, and 8 also include industry-level controls for log total employment, log total sales, and log average wage in 2010. Standard errors are robust against heteroskedasticity and reported in parentheses. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta$ Log Sales		$\Delta$ Log Employment		$\Delta$ HHI		$\Delta$ Top Firm Market Share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta$ Share AI Workers	0.169*** (0.055)	0.173*** (0.050)	0.195*** (0.069)	0.201*** (0.067)	0.018*** (0.007)	0.012* (0.007)	0.022*** (0.007)	0.014* (0.007)
Industry Sector FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y	N	Y
Observations	275	275	275	275	275	275	275	275

accumulate unmeasured output in the form of intangible capital that complements the new technology. In Online Appendix Table A27, we examine the effect of changes in AI investments during the first half of the period (2010–2014) on productivity growth through 2018 and do not find any significant positive effect. Hence, even with a lag of a few years, AI investments are not yet associated with productivity improvements. Besides, while the productivity J-curve reflects forgone measurable output in the short run, we find a significant and positive effect on sales growth two to three years following AI investments. Our evidence suggests that at least so far, AI mostly stimulates firm growth through product innovation. As we discuss in Online Appendix A1, the effect of product innovation on productivity is theoretically ambiguous since firms may have higher or lower productivity in the new product lines. Our empirical results suggest that AI-investing firms are able to maintain the same level of productivity at a larger scale. These findings align with recent work documenting that investments in technologies in recent years are associated with increased scale of the firm but no productivity gains (Aghion et al., 2019; Curtis et al., 2021; Hirvonen et al., 2022).

**7. AI investments and industry-level outcomes**

To shed light on the potential aggregate effects of AI, we examine the relationship between industry-level variation in AI investments and: (i) industry growth; and (ii) industry concentration. While AI-investing firms grow faster, the industry-level gains may be zero-sum if the use of AI technologies creates a business-stealing effect on competitors (Bloom et al., 2013). For example, negative spillovers have been shown to dominate positive firm-level effects in the case of robotics, leading to an overall negative effect on aggregate employment (Acemoglu et al., 2020; Benmelech and Zator, 2022). Hence, signing the relationship between industry AI investments and industry growth is an empirical question. We estimate the following long-differences regression at the industry level:

$$\Delta \ln y_{j,[2010,2018]} = \gamma \Delta \text{ShareAIWorkers}_{j,[2010,2018]} + \text{Controls}'_{j,2010} \gamma + \text{SectorFE} + \epsilon_j, \tag{4}$$

where  $\Delta \ln y_{j,[2010,2018]}$  is the change in total sales or employment for all Compustat firms (including those that entered the sample after 2010 or exited before 2018) in 5-digit NAICS industry  $j$ , and  $\Delta \text{ShareAIWorkers}_{j,[2010,2018]}$  is the change in the share of AI workers among Compustat firms in industry  $j$  from 2010 to 2018. Analogously to the firm-level tests, the regressions are weighted by the total number of resumes in each industry in 2010.

Columns 1 to 4 of Table 9 show that AI investments are associated with a robust increase in employment and sales at the industry

level. Odd columns estimate the unconditional relationship (with 2-digit NAICS industry fixed effects only), and even columns add controls for log employment, log sales, and log average wages at the industry level in 2010. For example, with the full set of controls, a one-standard-deviation increase in the industry-level share of AI workers in the resume data is associated with a 17% increase in sales and a 20% increase in employment. Importantly, in Online Appendix Table A28, we show that the results remain similar when we restrict the sample to firms that are in the Compustat sample both in 2010 and 2018 (i.e., excluding entrants and exits). This indicates that sample selection issues are not driving our main results for publicly traded firms.<sup>41</sup> Due to data limitations, our analysis does not capture cross-industry spillovers, where firms from one industry steal business of firms in other industries. Nevertheless, our results suggest that within-industry business-stealing effect is unlikely to dominate the positive firm-level growth from AI.

We next examine whether the higher AI-fueled growth among larger firms is substantial enough to translate into increased industry concentration. We link industry-level growth in AI investments to contemporaneous changes in industry concentration from 2010 to 2018. Following Autor et al. (2020), we use the Herfindahl-Hirschman Index (HHI) to measure industry concentration. To examine winner-take-most dynamics, we also consider the fraction of sales accruing to the largest firm in each 5-digit NAICS industry among the Compustat firms. Columns 5 to 8 of Table 9 show a robust positive relationship between industry-level growth in AI investments and changes in industry concentration. Online Appendix Table A29 shows consistent results using the industry-level share of AI-related job postings in Burning Glass as the independent variable.

To shed light on industry-level growth and concentration beyond public firms, we use data from the Economic Census to calculate industry-level sales, employment, and concentration measures in Online Appendix Table A30. We look at 2012–2017 changes because the Economic Census is available every 5 years.<sup>42</sup> The increase in industry-level sales and employment is significant, but smaller than in the industry-level analysis constructed from the sample of Compustat firms in Ta-

<sup>41</sup> A caveat with these results is that the Compustat sample assigns each firm to a single main industry, even for firms that might have operations in several industries. This caveat is unlikely to affect the interpretation of our results, given that prior research using U.S. Census micro data shows that for a typical U.S. public firm, the large majority of its operations fall within one main industry (Babina, 2020).

<sup>42</sup> The independent variable is the industry-level change in the share of AI workers among Compustat firms, since we do not have the industry information for non-Compustat firms. This could induce measurement error and, therefore, this result should be interpreted with caution.

ble 9. On the other hand, the increase in HHI becomes larger, and there is a significant increase in the market share of the four largest firms for AI-investing industries. Since publicly-listed firms are on average larger than private firms, this further supports the interpretation that gains from AI technologies accrue mostly to large, established firms that have the requisite data and resources.

Since AI-investing industries may differ from other industries in other ways during the time period we look at, our estimates do not necessarily reflect the direct causal effect of AI investments on industries. However, our results suggest that, as a general purpose technology that can be applied across many industries, AI is associated with increased concentration across a broad range of industries by facilitating product innovation and expansion for the largest firms. These findings support the argument by Crouzet and Eberly (2019) that investments in intangible assets are responsible for the rise in industry concentration observed in the U.S. data.

## 8. Conclusion

In this paper, we study how firms invest in and benefit from one of the most important new technologies of the last decade—artificial intelligence. We introduce a novel measure of investments in AI technologies at the firm level using two detailed datasets on human capital: resume data from Cognism, which reveal the actual composition of each firm's workforce, and job postings from Burning Glass Technologies, which indicate each firm's demand for particular skills. Our unique measure allows us to examine both the determinants and the consequences of AI investments by firms across a wide range of sectors. We find a positive feedback loop between AI investments and firm size: AI investments concentrate among the largest firms, and as firms invest in AI, they grow larger, gaining sales, employment, and market share. This AI-fueled growth does not appear to stem from cost-cutting; instead, AI-investing firms expand through product innovation and increased product offerings.

Our findings highlight important differences between the adoption of AI technologies and the adoption of information technology (IT) in the 1980s and 1990s.<sup>43</sup> Much of the previous literature finds that IT investments were associated with economically large productivity increases but mixed results on firm growth measures such as market share. By contrast, we observe increased growth for AI-investing firms, along with increased product innovation, but no evidence (yet) of higher firm-level productivity. Our results also show higher AI adoption and larger gains from AI investments for larger firms, which contrasts with prior work on diffusion patterns for IT (Hobijn and Jovanovic, 2001). These differences underscore the distinctive features of AI relative to previous waves of IT: as a prediction technology, our results show that AI facilitates product innovation and creates new business opportunities by enabling firms to learn better and faster from big data. The use and applications of AI technology have quickly expanded in the 2010s and beyond. Our results speak to the early wave of AI adoption, and efficiency gains, if present, may be more backloaded. We expect the evolving effects of AI to be an exciting area for future research.

Our findings imply that the benefits from AI depend to a large extent on who owns big data—the key input to AI technologies (Fedyk, 2016). While data are non-rival (data can be used by any number of firms simultaneously), recent theoretical work suggests that, fearing creative destruction, firms may choose to hoard data they own, leading to inefficient use of nonrival data; and that giving the data property rights to consumers can generate allocations that are close to optimal (Jones and Tonetti, 2020). Recent empirical evidence suggests that shifting data ownership rights away from firms to consumers in financial services can incentivize firm entry and potentially break big firm data advantages

<sup>43</sup> See Dedrick et al. (2003) and Cardona et al. (2013) for reviews of that literature.

(Babina et al., 2023b). While our empirical work does not directly speak to the optimality of data ownership, our results suggest that—in the current status quo where firms own consumer data—AI contributes to the increase in industry concentration and the rise of “superstar” firms documented in recent work (Gutiérrez and Philippon, 2017; Autor et al., 2020). Further understanding how AI affects production processes, corporate strategies, and the organizational structure of firms and assessing the distributional impacts of AI technologies across firms and workers are fruitful avenues for future research.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

All data and code have been published here: <https://data.mendeley.com/datasets/s26kxvspn7/2>.

## Acknowledgements

The authors are grateful to the editor, Dimitris Papanikolaou, and two anonymous referees for their excellent suggestions. The authors also thank Daron Acemoglu, Philippe Aghion, David Autor, Ann Bartel, Erik Brynjolfsson, Jan Bena, Francesco D'Acunto, Nicolas Crouzet, Xavier Giroud, Maarten de Ridder, Larry Katz, Anton Korinek, David Levine, Max Maksimovic, Gustavo Manso, Filippo Mezzanotti, Thomas Philippon, David Sraer, Tano Santos, Bledi Taska, Laura Veldkamp, Shang-Jin Wei, Michal Zator, and seminar participants at AFFECT, Babson College, Colloquium on The Peril and Promise of Artificial Intelligence (AI) for Corporations, Columbia University, ConnectAI, EFA, Federal Reserve Board Collab Week on “Digitalization of the Economy”, Federal Reserve Bank of Atlanta, Federal Reserve Bank of Richmond Conference “Secular Trends in Macroeconomics and Firm Dynamics,” FIRS, FOM Conference, Frontiers in Economics for Ukraine, Jozef Stefan Institute AI lab, Labor and Finance Group Conference, Michigan State University, NBER Corporate Finance, NBER Economics of AI, NBER productivity seminar, NBER SI Macroeconomics and Productivity, Northwestern University, NYU WAPFIN, Society for Economic Dynamics, Tilburg University, Triangle Macro-Finance Workshop, Tulane University, Queen Mary University of London, Sao Paulo School of Economics, UC Berkeley, UNC Junior Roundtable, University of Illinois Chicago, University of Maryland, University of Oklahoma, University of Washington, and Vienna Graduate School of Finance for helpful comments. The authors thank Cognism Ltd. for providing the employment data; Burning Glass Technologies for providing the job postings data; Bernhard Ganglmair, W. Keith Robinson, and Michael Seeligson for sharing their data on patent (product vs. process) classification; and Gerard Hoberg and Gordon Phillips for providing data on changes in firms' product mix. Nikita Chuvakhin, Joanna Harris, Derek Ho, Gary Nguyen, and Hieu Nguyen provided excellent research assistance. Fedyk gratefully acknowledges financial support from the Clausen Center for International Business and Policy at UC Berkeley.

## Appendix A. Appendix on instrument construction

We instrument firm-level AI investments using variation in firms' ex-ante exposure to the supply of AI-trained graduates from universities that are historically strong in AI. The core idea is that the scarcity of AI-trained labor is one of the most important barriers to firms' AI adoption (e.g., CorrelationOne (2019)). Universities are a key source of skilled labor, and universities historically strong in AI research are able to train more AI-skilled graduates following the widespread rise of commercial interest in AI in the 2010s. This enables firms with more

ex-ante connections to AI-strong universities (e.g., via alumni networks) to more readily attract AI talent from those universities in the 2010s. It is important to note that while AI research flourished in universities long before 2010 (research in AI and machine learning goes back to the 1950s), commercial interest in AI applications started around 2012, driven by rapid accumulation of data, decreasing costs of computation, and methodological advances in applying techniques such as deep learning.<sup>44</sup> Moreover, universities did not set up specialized data science programs until the mid-2010s. For example, Columbia's Data Science Institute (described as a "trailblazer in the field"; see [here](#)) was established in 2012. Therefore, in 2010, firms' connections to AI-strong universities were not driven by the need to hire AI-skilled workers, but rather by other pre-existing connections such as alumni networks (e.g., the CEO having graduated from a particular university), especially for the sample of non-tech firms that are the focus of this paper.

To construct the instrument, we need two different datasets. The first is a measure of the strength of AI research in each university in the pre-period. The second, even more difficult to construct, is a measure of firm-university hiring networks in the pre-period. To the best of our knowledge, there is no comprehensive historical data on either of these two aspects. To construct the first measure, we group all universities into those that are ex-ante strong in AI research and those that are not, based on the number of researchers producing AI-related publications in each university before 2010. A key concern with this measure for our instrument is that AI-strong universities are likely to also be strong in the broader field of computer science (CS), producing more CS-skilled graduates, which might affect firm outcomes through channels other than AI investments. To address this concern, we also collect information on the number of CS researchers in each university in each year to be included as a control. To construct the second measure (firm-university hiring networks), we leverage our resume data to observe the alma maters of the firm's employees as of 2010. To validate the data, we also measure: (i) the number of fresh graduates in each year from each university hired by each firm to confirm that ex-ante firm-university networks predict ex-post hiring, and (ii) the number of AI-trained graduates from each university to validate our premise that ex-ante AI-strong universities produce more AI-trained graduates following the increase in commercial interest in AI.

**Data Construction.** First, to identify universities that are ex-ante strong in AI, we use data from the Open Academic Graph (OAGv2) to measure AI-related publications associated with each university. OAGv2 provides a unified view of two large-scale databases of academic paper metadata, abstracts, citations, and author links: (i) the Microsoft Academic Graph (part of the Microsoft Academic Service infrastructure in Sinha et al., 2015), and (ii) ArnetMiner (Tang et al., 2008). Together, these two datasets provide the most comprehensive openly available repository of scholarly work starting from the 1870s, allowing us to track research articles and faculty across the near-universe of academic and commercial institutions. The Open Academic Graph contains hundreds of millions of papers from 366M distinct author names and lists author affiliations where available. We use a keyword-based matching procedure to link 689 research institutions (or 99%) in the Higher Education Research and Development Survey (HERDS) data to faculty information in the OAGv2. HERDS data are collected by the National Science Foundation and cover all universities in the U.S. that have at least \$150,000 in R&D expenditures in each fiscal year. Our strict matching procedure requires that the full formal university name, or an official shortened variant thereof, be found in full form within the institutional affiliations in the OAGv2 paper metadata files, with only common "stop-words" (such as "and," "the," and "in") removed from both sides of the match. A manual review of the resulting linked data shows over 96% precision in matching author affiliations from the Open Academic Graph to HERDS data, with the remaining incorrect en-

tries manually adjusted to ensure full correctness. For each university matched to HERDS, we consider all publications in the Open Academic Graph in each year that have at least one co-author affiliated with that university.

We work with the field experts at the AI for Good Foundation to identify AI-related publications.<sup>45</sup> First, we identify a small set of "seed" journals and conference proceedings that explicitly include terms like "artificial intelligence" and "machine learning" in their title (e.g., *Journal of Machine Learning Research* and *Proceedings of the International Joint Conference on Artificial Intelligence*). Second, to identify potential additional AI-related journals and conference proceedings, we look at all other journals and proceedings that have published work by the authors of the papers in the seed journals and proceedings. We manually filter this broader set of journals and conference proceedings to the ones that focus predominantly on AI, leading to a final list of 355 journals and conference proceedings globally.

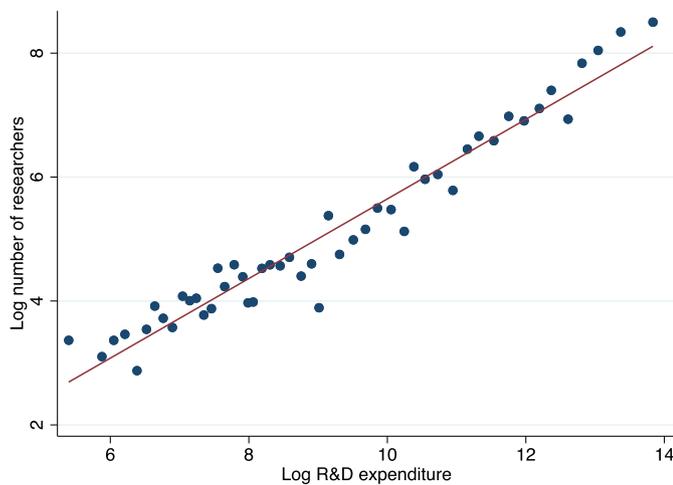
To make sure that our results are not driven by firms' exposure to broader (non-AI) CS-skilled workers, we control for firms' ex-ante exposure to CS-strong universities based on firms' hiring networks. In particular, we construct an analogous measure of computer science publications by starting with a set of seed journals and conference proceedings across different fields of computer science (those with the terms "compilers," "databases," "cryptography," "computation," "software," "programming," "informatics," "robotics," or "information security" in their titles) and then manually screening all other journals and conference proceedings that publish papers by the same authors. We exclude any journal or conference proceeding that we classify as AI-related, leaving a total of 796 non-AI computer science journals and conference proceedings.

After identifying the set of AI-related and CS-related journals and conference proceedings, we classify the focus area of each researcher  $r$  as either AI, computer science, or neither. If at least one third of all publications co-authored by  $r$  are in either AI or CS journals and conference proceedings, then  $r$  is considered a candidate researcher. If  $r$  is a candidate researcher *and* at least half of  $r$ 's AI/CS publications are in specifically AI journals and proceedings, then  $r$  is marked as an AI researcher. If more than half of  $r$ 's AI/CS publications are in non-AI computer science journals and proceedings, then  $r$  is considered a non-AI CS researcher. Finally, if more than two thirds of  $r$ 's overall publications are outside of the set of identified AI and CS journals and proceedings, then  $r$  is classified as a researcher in other (unrelated) fields.

At the university level, we compute the percentage of researchers in each year who are classified as AI researchers and the percentage of researchers who are classified as CS researchers. Researchers in other unrelated fields are included in the denominators of both measures. To reduce noise, we assume that each researcher is employed at the respective university in a non-publishing year if that researcher is employed at that university in both the following and the preceding year. For example, if researcher  $r$  is identified as affiliated with university  $u$  in both 2005 and 2007 but has no publications in 2006, then  $r$  is still considered to be affiliated with university  $u$  in year 2006. We then classify whether each university is AI-strong. We define a university as being strong in AI if it satisfies one of the following two criteria in at least one year between 2005 and 2009: (i) the number of AI researchers is in the top 5% of the distribution across all universities in a given year; or (ii) the number of AI researchers is in the top 10% of the distribution, and the share of AI researchers (the number of AI researchers divided by the number of other researchers in the OAGv2 data) is in the top 5% of the distribution across all universities in a given year. We use the second criterion because there are some smaller, tech-oriented colleges that could potentially have a large share of researchers in AI but do

<sup>44</sup> A brief history of AI research can be found [here](#).

<sup>45</sup> Learn more about the AI for Good Foundation [here](#).



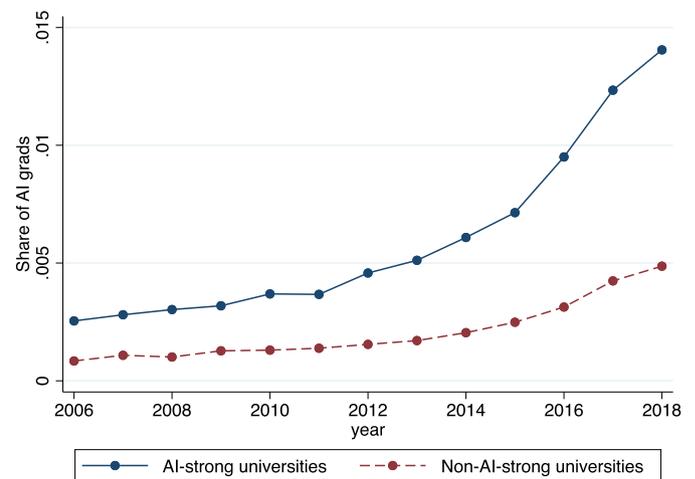
**Fig. A.1.** Correlation Between the Number of University Researchers and University R&D Expenditures. This figure is a binned scatterplot of the log number of researchers in each university against the log R&D expenditure in each university in 2010. Each dot represents roughly the same number of universities, and the solid line is the fitted regression line. The number of researchers in each university is the number of authors from that university with at least one publication in the OAGv2 data. The R&D expenditure of each university is from the NSF's HERDS data.

not necessarily have large departments. Our results are robust to using other cutoffs and earlier years.

We verify that the OAGv2 publication data provide a reliable measure of university research. In Fig. A.1, we plot the log number of (all) researchers in each university in the OAGv2 data against the log R&D expenditure in the HERDS data. We find a strong positive correlation of 0.83. Furthermore, the top universities we identify as AI-strong include top AI departments, such as Carnegie Mellon University, UCLA, Stanford University, UIUC, New York University, and University of Maryland College Park, but are *not* strongly correlated with the overall highest-ranked universities based on the U.S. News & World Report. For example, only 50% (39%) of the top 20 (top 50) universities are AI-strong universities, and among AI-strong universities, only 25% (56%) are ranked in the top 20 (top 50) universities in the U.S. News & World Report.

To construct the second ingredient for our instrument—firm exposure to AI-strong universities via the ex-ante firm-university hiring networks—we use our Cognism resume data. In these data, we observe the granting institutions of all degrees that workers list on their resumes. We disambiguate university names and match them to the HERDS data. We define an individual  $i$  as a graduate of university  $u$  if  $i$ 's resume lists at least one degree (undergraduate or graduate) from university  $u$ . We define an individual  $i$  as a *fresh graduate* from university  $u$  in year  $t$  if  $i$  joined a firm in year  $t$  and graduated from university  $u$  in year  $t$  or year  $t - 1$ . These data offer comprehensive coverage of universities; for example, in 2010, 668 of the 716 universities in the HERDS dataset have at least one fresh graduate in our resume data. Since the firms' hiring patterns might be different for STEM versus non-STEM workers (e.g., if a firm has a hiring relationship with an economics department for economic policy talent and with a business school for management talent), we also consider the firm-university hiring networks based specifically on STEM workers, in case such networks are more relevant for hiring AI workers. We define STEM workers as employees who have at least one degree with a major in either engineering (e.g., electrical, chemical, mechanical), physical sciences (e.g., math, physics, chemistry, computer science, statistics), or biological sciences (e.g., biology, pharmacology). We drop firms if the share of STEM workers with missing university information is above the 95th percentile.

We compare the coverage of our university graduates data with official statistics from universities and show that our resume data cover



**Fig. A.2.** Time Series of the Share of AI-trained Fresh Graduates from Ex-ante AI-strong Universities and Other Universities. This figure plots the average share of AI-trained fresh graduates out of all fresh graduates from 2006 to 2018, separately for ex-ante AI-strong universities and non-AI-strong universities. We define a university as an AI-strong university if it satisfies one of the following two criteria in at least one year between 2005 and 2009: (i) the number of AI researchers is in the top 5% of the distribution across all universities in a given year; (ii) the number of AI researchers is in the top 10% of the distribution, and the share of AI researchers is in the top 5% of the distribution across all universities in a given year. We define an individual  $i$  as a fresh graduate from university  $u$  in year  $t$  if individual  $i$  joined a firm in year  $t$  and graduated from university  $u$  in year  $t$  or year  $t - 1$ . An individual is considered an AI-trained fresh graduate in year  $t$  if the individual is a fresh graduate in year  $t$  and that individual's first job after graduation is an AI-skilled job. AI-skilled jobs are defined based on the methodology described in Section 4.2 and used throughout the paper.

a sizable proportion of university graduates in the U.S. In particular, we aggregate the data to university-year level by calculating the total number of fresh graduates from each university in each year. We compare these numbers with the total numbers of all degrees (bachelors, masters, and PhDs) conferred by each university in each year, using the Integrated Postsecondary Education Data System (IPEDS) data, which contain the total enrollment and the number of degrees conferred each year for all post-secondary institutions in the U.S. As of 2012 (the latest year of the IPEDS data), our resume data cover, on average, 59% of all fresh graduates at each university. The number of fresh graduates in the resume data is also highly correlated with graduates in the cross-section of universities (correlation = 0.73).

Finally, we use our Cognism resume data to measure the share of all fresh university graduates from each university who get AI-skilled jobs in each year between 2006 and 2018. These data allow us to validate our premise that ex-ante AI-strong universities are able to increase the supply of AI-skilled graduates following the increase in commercial applications in AI in the first half of the 2010s, discussed below.

**Instrument Validation.** We validate several core assumptions underlying the intuition behind our instrument. To begin with, we show that the increase in AI-trained graduates during the 2010s was much more pronounced in AI-strong universities than in non-AI-strong universities. Fig. A.2 plots the share of fresh graduates that are AI-trained from AI-strong and non-AI-strong universities from 2006 to 2018. In 2006, there were few AI graduates across the board, with the share of AI graduates below 0.3% for both AI-strong and non-AI-strong universities. Even in 2012, the share of AI graduates remained below 0.5% in both groups of universities. From 2012 to 2018, however, the share of AI graduates tripled (to about 1.5%) in AI-strong universities, while the share of AI graduates remained under 0.5% in non-AI-strong universities.

We then examine whether firm-university hiring networks provide the necessary variation for our instrumental variable strategy. First, our

**Table A.1**

Persistence of Firm-University Hiring Networks. This table reports the coefficients from regressing the share of each firm's fresh graduates hired from each university after 2010 on the pre-2010 firm-university network. Each observation is a firm-university pair. The dependent variable, constructed using the Cognism resume data, is the share of all fresh graduates hired from each university after 2010 in columns 1 and 2 and the share of AI-trained fresh graduates hired from each university after 2010 in columns 3–6. In columns 1 and 3, the independent variable is the share of all fresh graduates hired from each university between 2005 and 2010. We define an individual  $i$  as a fresh graduate from university  $u$  in year  $t$  if individual  $i$  joined a firm in year  $t$  and graduated from university  $u$  in year  $t$  or year  $t - 1$ . In columns 2 and 4, the independent variable is the share of all workers in the firm in 2010 who graduated from each university. In column 5, the independent variable is the share of all STEM fresh graduates hired from each university before 2010. We define STEM workers as employees who have at least one degree with a major in either engineering (e.g., electrical, chemical, mechanical), physical sciences (e.g., math, physics, chemistry, computer science, statistics), or biological sciences (e.g., biology, pharmacology). In column 6, the independent variable is the share of STEM workers in the firm in 2010 who graduated from each university. All columns control for firm fixed effects and university fixed effects. Standard errors are clustered at the university level and reported in parentheses. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	Share of Post-2010 Hires		Share of Post-2010 AI Hires			
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Pre-2010 Hires	0.465*** (0.017)		0.550*** (0.054)			
Share of 2010 Workers		0.147*** (0.006)		0.236*** (0.028)		
Share of Pre-2010 STEM Hires					0.342*** (0.040)	
Share of 2010 STEM Workers						0.197*** (0.021)
Firm FE	Y	Y	Y	Y	Y	Y
University FE	Y	Y	Y	Y	Y	Y
Observations	327,313	327,313	177,097	177,097	177,097	177,097

**Table A.2**

Changes in Hiring from Ex-ante AI-strong Universities during the Pre-period (2005–2010). This table reports the coefficients from regressing the change in the share of fresh graduates from AI-strong universities from 2005 to 2010 on the instrument, which measures ex-ante firm-level exposure to the supply of AI-trained university graduates from AI-strong universities (see the definition of the instrument in Section 5.3 and the details of instrument construction in Appendix A). The independent variable is standardized to mean zero and standard deviation of one. Columns 2–5 control for ex-ante exposure to universities that are strong in CS research and top 10 universities. Columns 3–5 also control for the 2-digit NAICS industry sector fixed effects. Columns 4 and 5 add the baseline controls all measured as of 2010: firm-level characteristics (log sales, cash/assets, firm age, and log number of resumes), log industry wage, and characteristics of the commuting zones where the firms are located (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers). Column 5 additionally controls for state fixed effects. Regressions are weighted by the number of Cognism resumes in 2010. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta$ Share of Fresh Graduates Hired from AI-strong Universities 2005–2010				
	(1)	(2)	(3)	(4)	(5)
Instrument	0.023 (0.088)	0.026 (0.096)	0.033 (0.091)	-0.005 (0.110)	0.060 (0.120)
University Control	N	Y	Y	Y	Y
Industry FE	N	N	Y	Y	Y
Baseline Control	N	N	N	Y	Y
State FE	N	N	N	N	Y
Observations	830	830	829	829	825

instrument leverages the variation in exposure to AI-strong universities across firms. Therefore, it requires that firms do not hire uniformly from the same universities. Empirically, most firms in our data concentrate their hiring in a small number of universities. On average, a firm hires 18% of its fresh graduates from the single main university in its network, 44% from its five main universities, and 59% from its 10 main universities. By contrast, the largest university produces only 1.6% of all fresh graduates, the largest five universities produce 7.1% of all fresh graduates, and the largest 10 universities produce 12.9% of all fresh graduates. Firms also hire disproportionately from universities located in the same state as their headquarters: on average, firms hire 38% of all fresh graduates, 37% of STEM fresh graduates, and 42% of AI fresh graduates from universities located in the same state. Second, in

order for the ex-ante firm-university network to predict ex-post hiring of AI-skilled labor, firm-university networks need to be persistent over time. In column 1 of Table A.1, we regress the share of fresh graduates hired from each university after 2010 on the share of fresh graduates hired from each university before 2010. We find a strong positive relationship, suggesting that firm-university networks are correlated over time. In column 2, we use the share of all workers employed in a firm in 2010 who graduated from each university to predict the share of fresh graduates hired from that university after 2010, again finding a strong positive correlation. The persistence of firm-university hiring networks also manifests in AI hiring. Columns 3 and 4 show that the universities from which a firm hired before 2010 also strongly predict the universities from which the firm hires its AI-skilled workers after 2010. Finally,

**Table A.3**

First Stage of the Instrument. This table reports the first stage of the instrument, where we regress our key independent variable—firm-level changes in the share of AI-skilled workers from 2010 to 2018—on the instrument, which measures ex-ante firm-level exposure to the supply of AI-trained university graduates from AI-strong universities (see the definition of the instrument in Section 5.3 and the details of instrument construction in Appendix A). The dependent variable is the resume-based measure of the growth in the share of AI workers from 2010 to 2018. The dependent variables and the instrument are standardized to mean zero and standard deviation of one. Regressions are weighted by the number of Cognism resumes in 2010. All specifications control for the 2-digit NAICS industry sector fixed effects and ex-ante firm-level exposure to universities that are historically strong in CS research as well as top 10 universities. Columns 2–4 also include the baseline controls measured as of 2010: firm-level characteristics (log sales, cash/assets, firm age, and log number of resumes), log industry wage, and characteristics of the commuting zones where the firms are located (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers). Columns 3 and 4 add controls for firm-level pre-trends: changes in log sales and log employment from 2000 to 2008. Column 4 adds state fixed effects. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. The first-stage F-statistics of the instrument are for all specifications. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta$ Share of AI Workers			
	(1)	(2)	(3)	(4)
Instrument	0.609*** (0.171)	0.422*** (0.096)	0.455*** (0.115)	0.464*** (0.107)
Industry FE	Y	Y	Y	Y
University Control	Y	Y	Y	Y
Baseline Controls	N	Y	Y	Y
Control Pre-trend	N	N	Y	Y
State FE	N	N	N	Y
F Statistic	12.7	19.3	15.7	18.7
Observations	1,001	1,001	777	773

in columns 5 and 6, we show that pre-2010 firm-university hiring networks based only on STEM workers also strongly predict the universities from which firms hire their AI workers after 2010.

Our instrument is defined as follows for each firm  $i$ :

$$IV_i = \sum_u s_{iu}^{2010} AIstrong_u,$$

where  $s_{iu}^{2010}$  is the share of STEM workers in firm  $i$  in 2010 who graduated from university  $u$ , and  $AIstrong_u$  equals one if university  $u$  is identified as an AI-strong university based on pre-2010 publications as described above. We use firm-university hiring networks based on STEM workers in the firm as of 2010, because the instrument based on this measure has a stronger first stage; however, the results are very similar when we construct firm-university hiring networks using all workers in the firm as of 2010. To reduce noise, for each firm's hiring network, we consider the 50 universities from which the firm has the most workers in 2010. To control for the effects of general computer science (and not specifically AI), we construct an analogous measure of firms' exposure to CS-strong universities:  $\sum_u s_{iu}^{2010} CSstrong_u$ , where the weights  $s_{iu}^{2010}$  are firms' 2010 STEM hiring shares, and  $CSstrong_u$  is the average share of (non-AI) CS researchers (the number of CS researchers divided by the number of all other researchers) at university  $u$  between 2005 and 2009. To control for the effects of overall university ranking, we also construct a measure of firms' exposure to top-ranked universities:  $\sum_u s_{iu}^{2010} Top10_u$ , where  $Top10_u$  equals one if university  $u$  is one of the top 10 universities ranked by U.S. News & World Report in 2010.<sup>46</sup>

Before proceeding, we examine an important identification concern regarding our instrument: if firms anticipated the surge in demand for AI, they might have started building their connections to AI-strong universities before 2010, making firm-university hiring networks in 2010 endogenous to firms' ability to hire AI-trained students ex-post. This is unlikely, given the lack of both commercial interest in AI by firms and training of AI-skilled graduates by universities (Fig. A.2) prior to 2010. Indeed, we are able to confirm empirically that firms connected to AI-

strong universities did not increase their share of hired fresh graduates from those universities from 2005 to 2010. Specifically, in Table A.2, we find no significant relationship between the change in the share of fresh graduates from AI-strong universities in the pre-period (from 2005 to 2010) and our instrument.

In Appendix Table A.4, we further show that firms that are more exposed to AI-strong universities are not growing faster before 2010, which supports the exclusion restriction that the exposure to AI-strong universities only affects firm growth through firms' AI investments after 2010.

**First Stage.** Table A.3 presents the first stage of the instrument, where we regress our key independent variable—firm-level changes in the share of AI-skilled workers from 2010 to 2018—on the instrument, which measures ex-ante firm-level exposure to the supply of AI-trained university graduates from AI-strong universities. We control for firm-specific ex-ante exposure to CS-strong universities and top-10 universities and industry fixed effects in all specifications. In column 2, we additionally control for our baseline controls measured as of 2010: (i) firm-level variables (log employment, cash/assets, log sales, R&D/Sales, and log markups), (ii) the characteristics of the commuting zones where the firms are located in 2010 (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers); and (c) the log industry-average wage. The inclusion of these controls helps to address the concern that firms' ex-ante exposure to AI-strong universities might be correlated with other firm characteristics that can drive AI adoption and firm growth. In column 3, we also control for firms' pre-period sales and employment growth between 2000 and 2008 to address unobservable firm characteristics that might simultaneously drive firms' growth trajectories and their hiring of AI workers. In column 4, we further add state fixed effects to control for local labor market characteristics that might drive both universities' ability to produce AI graduates and firm growth. The first stage F-statistics are well above the conventional level of 10 for all specifications.

**Robustness.** We conduct several robustness tests of the IV results. First, we use an alternative measure of universities' research strength

<sup>46</sup> The top 10 universities include: Harvard, Princeton, Yale, University of Pennsylvania, Columbia, MIT, Stanford, Caltech, University of Chicago, Duke.

**Table A.4**

Firm Connections to AI-strong Universities and Firm Growth in the Pre-period. This table reports the coefficients from regressions of firm growth from 2000 to 2008 on the instrument, which measures ex-ante firm-level exposure to the supply of AI-trained university graduates from AI-strong universities (see the definition of the instrument in Section 5.3 and the details of instrument construction in Appendix A). Regressions are weighted by the number of Cognism resumes in 2010. The instrument is standardized to mean zero and standard deviation of one. We consider changes in log sales in columns 1 and 2, log employment in columns 3 and 4, and log market value in columns 5 and 6. All specifications control for the 2-digit NAICS industry sector fixed effects and ex-ante exposure to universities that are strong in computer science research as well as top 10 universities. All columns also control for the baseline controls all measured as of 2010: firm-level characteristics (log sales, cash/assets, firm age, and log number of resumes), log industry wage, and characteristics of the commuting zones where the firms are located (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers). Columns 2, 4, and 6 additionally control for state fixed effects. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta$ Log Sales, 2000–2008		$\Delta$ Log Employment, 2000–2008		$\Delta$ Log Market Value, 2000–2008	
	(1)	(2)	(3)	(4)	(5)	(6)
Instrument	-0.027 (0.034)	-0.032 (0.040)	-0.005 (0.044)	0.006 (0.047)	-0.041 (0.033)	-0.037 (0.040)
Industry FE	Y	Y	Y	Y	Y	Y
University Control	Y	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y
State FE	N	Y	N	Y	N	Y
Observations	821	817	780	776	760	753

**Table A.5**

Growth in University AI Research in the Pre-period. This table reports the coefficients from regressing the change in connected universities' AI research strength from 2005 to 2010 on the change in the share of AI workers from 2010 to 2018. The dependent variable is the weighted average of the change in the log number of AI researchers of connected universities in columns 1 to 4, and the weighted average of the change in the share of AI researchers of connected universities in columns 5 to 8, where the weights are the share of STEM workers in 2010 graduating from each university. Regressions are weighted by the number of Cognism resumes in 2010. All specifications control for the 2-digit NAICS industry sector fixed effects and ex-ante firm-level exposure to universities that are historically strong in CS research as well as top 10 universities. All columns also control for the weighted average of the log number of AI researchers and the share of AI researchers in 2005. Columns 2–4 and 6–8 also include the baseline controls measured as of 2010: firm-level characteristics (log sales, cash/assets, firm age, and log number of resumes), log industry wage, and characteristics of the commuting zones where the firms are located (the share of workers in IT-related occupations, the share of college-educated workers, log average wage, the share of foreign-born workers, the share of routine workers, the share of workers in finance and manufacturing industries, and the share of female workers). Columns 3–4 and 7–8 add controls for firm-level pre-trends: changes in log sales and log employment from 2000 to 2008. Columns 4 and 8 add state fixed effects. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$\Delta$ Log Number of AI Researchers				$\Delta$ Share of AI Researchers			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta$ Share AI Workers	0.012 (0.007)	0.004 (0.007)	0.000 (0.008)	0.005 (0.008)	0.004 (0.003)	0.005 (0.004)	0.008 (0.005)	0.008 (0.005)
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y
University Control	Y	Y	Y	Y	Y	Y	Y	Y
Baseline Controls	N	Y	Y	Y	N	Y	Y	Y
Control Pre-trend	N	N	Y	Y	N	N	Y	Y
State FE	N	N	N	Y	N	N	N	Y
Observations	1,001	1,001	777	773	1,001	1,001	777	773

in CS more generally. The ranking of CS departments is from [csrankings.org](https://csrankings.org) and is based on publications in prestigious CS conferences. The same ranking is also used in Gofman and Jin (2022). To make it consistent with our main measure, we exclude conferences in AI areas (including AI, computer vision, machine learning, and natural language processing) and limit the time period to 2005–2009. The results are robust to this control and are reported in Online Appendix Table A20.

Second, one potential concern is reverse causality: universities might have become stronger in AI research because of their connections to firms who could provide the data and resources necessary to support breakthroughs on the research side. Since AI research started well before the wider commercial application of AI among non-tech firms in the 2010s, the reverse causality is less likely. Moreover, we provide two tests to address this concern. We first regress the change in universities' AI research strength (measured by the log number of AI researchers or the share of AI researchers) from 2005 to 2010 on connected firms'

change in AI investments between 2010 and 2018. Appendix Table A.5 shows that universities connected to firms that later invest more in AI did not increase their AI strength in the pre-period. We also measure universities' AI research strength using an earlier period from 2000 to 2004 instead of 2005–2009 and find similar results (Online Appendix Table A21).

Third, we show that our results are not driven by a single prominent university that is strong in AI and other areas. In Online Appendix Table A22, we drop, in turn, each of the three universities with the largest number of AI faculty—Stanford University, Carnegie Mellon University, and the University of Illinois Urbana-Champaign—from the calculation of the IV, and the results remain unchanged.

Finally, the hiring links could be endogenous and correlated with other omitted variables (e.g., if the hiring links are driven by AI-related or programming-related jobs). Using ex-ante firm-university networks mitigates this concern. Furthermore, in Online Appendix Table A23, we

drop all STEM workers and only use non-STEM workers to construct firm-university networks. The first stage of the IV becomes slightly weaker, but the IV estimates remain similar to the baseline results.

## Appendix B. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jfineco.2023.103745>.

## References

- Abis, S., Veldkamp, L., 2023. The changing economics of knowledge production. *Rev. Financ. Stud.* Forthcoming.
- Abowd, J.M., Stephens, B.E., Vilhuber, L., Andersson, F., McKinney, K.L., Roemer, M., Woodcock, S., 2009. The LEHD infrastructure files and the creation of the quarterly workforce indicators. In: *Producer Dynamics*. University of Chicago Press, pp. 149–234.
- Acemoglu, D., Anderson, G., Beede, D.N., Buffington, C., Childress, E., Dinlersoz, E., Foster, L., Goldschlag, N., Haltiwanger, J.C., Kroff, Z., Restrepo, P., Zolas, N., 2022a. Technology, firms, and workers: Evidence from the 2019 Annual Business Survey. Working paper.
- Acemoglu, D., Autor, D., Hazell, J., Restrepo, P., 2022b. Artificial intelligence and jobs: Evidence from online vacancies. *J. Labor Econ.* 40, S293–S340.
- Acemoglu, D., LeLarge, C., Restrepo, P., 2020. Competing with robots: Firm-level evidence from France. *AEA Pap. Proc.* 110, 383–388.
- Acemoglu, D., Restrepo, P., 2018. The race between man and machine: Implications of technology for growth, factor shares and employment. *Am. Econ. Rev.* 108, 1488–1542.
- Acemoglu, D., Restrepo, P., 2019. 8. Artificial Intelligence, Automation, and Work. University of Chicago Press.
- Acemoglu, D., Restrepo, P., 2020. Robots and jobs: Evidence from U.S. labor markets. *J. Polit. Econ.* 128, 2188–2244.
- Aghion, P., Antonin, C., Bunel, S., Jaravel, X., 2020. What are the labor and product market effects of automation? New evidence from France. Working paper.
- Aghion, P., Bergeaud, A., Boppart, T., Klenow, P.J., Li, H., 2019. A theory of falling growth and rising rents. NBER Working Paper w26448.
- Aghion, P., Howitt, P., 1992. A model of growth through creative destruction. *Econometrica* 60, 323–351.
- Aghion, P., Jones, B.F., Jones, C.I., 2017. Artificial intelligence and economic growth. NBER Working paper w23928.
- Agrawal, A., Gans, J.S., Goldfarb, A., 2019. Artificial intelligence: The ambiguous labor market impact of automating prediction. *J. Econ. Perspect.* 33, 31–50.
- Akcigit, U., Kerr, W.R., 2018. Growth through heterogeneous innovations. *J. Polit. Econ.* 126, 1374–1443.
- Alderucci, D., Branstetter, L., Hovy, E., Runge, A., Zolas, N., 2020. Quantifying the impact of AI on productivity and labor demand: Evidence from U.S. census microdata. Working paper.
- Alekseeva, L., Azar, J., Gine, M., Samila, S., Taska, B., 2021. The demand for ai skills in the labor market. *Labour Econ.* 71, 102002.
- Altonji, J., Elter, T., Taber, C., 2005. Selection on observed and unobserved variables: assessing the effectiveness of catholic schools. *J. Polit. Econ.* 113, 151–184.
- Argente, D., Baslandze, S., Hanley, D., Moreira, S., 2020. Patents to products: Product innovation and firm dynamics. Working paper.
- Argente, D., Fitzgerald, D., Moreira, S., Priolo, A., 2021. How do firms build market share? Working paper.
- Autor, D., Dorn, D., Katz, L.F., Patterson, C., Van Reenen, J., 2020. The fall of the labor share and the rise of superstar firms. *Q. J. Econ.* 135, 645–709.
- Babina, T., 2020. Destructive creation at work: how financial distress spurs entrepreneurship. *Rev. Financ. Stud.* 33, 4061–4101.
- Babina, T., Bernstein, A., Mezzanotti, F., 2023a. Financial disruptions and the organization of innovation: Evidence from the great depression. *Rev. Financ. Stud.* 36, 4271–4317.
- Babina, T., Buchak, G., Gornall, W., 2023b. Customer data access and fintech entry: Early evidence from open banking. Working paper.
- Babina, T., Fedyk, A., He, A.X., Hodson, J., 2022. Firm investments in artificial intelligence technologies and changes in workforce composition. Working paper.
- Babina, T., He, A.X., Howell, S.T., Perlman, E.R., Staudt, J., 2023c. Cutting the innovation engine: how federal funding shocks affect university patenting, entrepreneurship, and publications\*. *Q. J. Econ.* 138, 895–954.
- Babina, T., Howell, S.T., 2023. Entrepreneurial spillovers from corporate R&D. *J. Labor Econ.* Forthcoming.
- Bartel, A., Ichniowski, C., Shaw, K., 2007. How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills. *Q. J. Econ.* 122, 1721–1758.
- Basu, S., Fernald, J.G., Kimball, M.S., 2006. Are technology improvements contractionary? *Am. Econ. Rev.* 96, 1418–1448.
- Basu, S., Fernald, J.G., Shapiro, M.D., 2001. Productivity growth in the 1990s: Technology, utilization, or adjustment? In: *Carnegie-Rochester Conference Series on Public Policy*. Elsevier, pp. 117–165.
- Benmelech, E., Zator, M., 2022. Robots and firm investment. NBER Working paper w29676.
- Bertrand, M., Schoar, A., 2003. Managing with style: the effect of managers on firm policies. *Q. J. Econ.* 118, 1169–1208.
- Bessen, J., 2020. Industry concentration and information technology. *J. Law Econ.* 63, 531–555.
- Bloom, N., Hassan, T.A., Kalyani, A., Lerner, J., Tahoun, A., 2021. The diffusion of disruptive technologies. NBER Working paper w28999.
- Bloom, N., Schankerman, M., Van Reenen, J., 2013. Identifying technology spillovers and product market rivalry. *Econometrica* 81, 1347–1393.
- Braguinsky, S., Ohyama, A., Okazaki, T., Syverson, C., 2021. Product innovation, product diversification, and firm growth: Evidence from Japan's early industrialization. *Am. Econ. Rev.* 111, 3795–3826.
- Brynjolfsson, E., Jin, W., Wang, X., 2023. Information technology, firm size, and industrial concentration. NBER Working Paper w31065.
- Brynjolfsson, E., McAfee, A., Sorell, M., Zhu, F., 2008. Scale without mass: Business process replication and industry dynamics. Harvard Business School Working Paper 07-016.
- Brynjolfsson, E., McElheran, K., 2016. The rapid adoption of data-driven decision-making. *Am. Econ. Rev.* 106, 133–139.
- Brynjolfsson, E., Rock, D., Syverson, C., 2019. Artificial intelligence and the modern productivity paradox. *Econ. Artif. Intell., An Agenda*, 23.
- Brynjolfsson, E., Rock, D., Syverson, C., 2021. The productivity J-curve: How intangibles complement general purpose technologies. *Am. Econ. J. Macroecon.* 13, 333–372.
- Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Henke, N., Trench, M., 2017. Artificial Intelligence the Next Digital Frontier? McKinsey and Company Global Institute.
- Bustamante, M.C., Cujean, J., Fresard, L., 2020. Knowledge cycles and corporate investment. Working paper.
- Caliendo, L., Mion, G., Oromolla, L.D., Rossi-Hansberg, E., 2020. Productivity and organization in portuguese firms. *J. Polit. Econ.* 128, 4211–4257.
- Cao, S., Jiang, W., Wang, J.L., Yang, B., 2021. From Man vs. Machine to Man+ Machine: the Art and AI of Stock Analyses. NBER Working Paper w28800.
- Cardona, M., Kretschmer, T., Strobel, T., 2013. ICT and productivity: Conclusions from the empirical literature. *Inf. Econ. Policy* 25, 109–125.
- Chen, M.A., Wu, Q., Yang, B., 2019. How valuable is fintech innovation? *Rev. Financ. Stud.* 32, 2062–2106.
- Cockburn, I.M., Henderson, R., Stern, S., 2018. The impact of artificial intelligence on innovation. NBER Working Paper w24449.
- Corrado, C., Haskel, J., Jona-Lasinio, C., Iommi, M., 2016. Intangible investment in the E.U. and U.S. before and since the Great Recession and its contribution to productivity growth. EIB Working Papers No. 2016/08.
- CorrelationOne, 2019. *Future of Data Talent, 2019 Annual Report*.
- Covarrubias, M., Guriérrez, G., Philippon, T., 2019. From good to bad concentration? U.S. industries over the past 30 years. NBER Macroecon. Annu. 34.
- Crouzet, N., Eberly, J.C., 2019. Understanding weak capital investment: the role of market concentration and intangibles. NBER Working Paper w25869.
- Curtis, E.M., Garrett, D.G., Ohrn, E.C., Roberts, K.A., Suárez Serrato, J.C., 2021. Capital Investment and Labor Demand. NBER Working Paper w29485.
- D'Acunzio, F., Prabhala, N., Rossi, A.G., 2019. The promises and pitfalls of robo-advising. *Rev. Financ. Stud.* 32, 1983–2020.
- Davis, S.J., Haltiwanger, J., Jarmin, R., Miranda, J., Foote, C., Nagypal, E., 2006. Volatility and dispersion in business growth rates: Publicly traded versus privately held firms. NBER Macroecon. Annu. 21, 107–179.
- De Loecker, J., Eeckhout, J., Unger, G., 2020. The rise of market power and the macroeconomic implications. *Q. J. Econ.* 135, 561–644. Forthcoming.
- Dedrick, J., Gurbaxani, V., Kraemer, K.L., 2003. Information technology and economic performance: A critical review of the empirical evidence. *ACM Comput. Surv.* 35, 1–28.
- Eisfeldt, A.L., Kim, E., Papanikolaou, D., 2020. Intangible value. NBER Working Paper w28056.
- Eisfeldt, A.L., Papanikolaou, D., 2013. Organization capital and the cross-section of expected returns. *J. Finance* 68, 1365–1406.
- Farboodi, M., Mihet, R., Philippon, T., Veldkamp, L., 2019. Big data and firm dynamics. *AEA Pap. Proc.* 109, 38–42.
- Farboodi, M., Veldkamp, L., 2021. A growth model of the data economy. NBER Working Paper w28427.
- Farboodi, M., Veldkamp, L., 2022. A model of the data economy. NBER Working Paper 28427.
- Fedyk, A., 2016. How to tell if machine learning can solve your business problem. *Harv. Bus. Rev.*
- Fedyk, A., Hodson, J., 2023. Trading on talent: Human capital and firm performance. *Rev. Finance* 149, 92–113.
- Felten, E., Raj, M., Seamans, R., 2019. The variable impact of artificial intelligence on labor: The role of complementary skills and technologies. Working paper.
- Fizsbein, M., Lafortune, J., Lewis, E., Tessada, J., 2020. Electrifying? How new technologies impact productivity and jobs. Working paper.
- Forman, C., 2005. The corporate digital divide: Determinants of Internet adoption. *Manag. Sci.* 51, 641–654.
- Foster, L., Haltiwanger, J., Syverson, C., 2008. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *Am. Econ. Rev.* 98, 394–425.

- Foster, L., Haltiwanger, J., Syverson, C., 2016. The slow growth of new plants: learning about demand? *Economica* 83, 91–129.
- Furman, J., Seamans, R., 2019. AI and the economy. *Innov. Policy Econ.* 19, 161–191.
- Furman, J.L., MacGarvie, M.J., 2007. Academic science and the birth of industrial research laboratories in the U.S. pharmaceutical industry. *J. Econ. Behav. Organ.* 63, 756–776.
- Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., Walther, A., 2020. Predictably unequal? The effects of machine learning on credit markets. In: *The Effects of Machine Learning on Credit Markets*.
- Ganglmair, B., Robinson, W.K., Seeligson, M., 2021. The rise of process claims: Evidence from a century of U.S. patents. Working paper. University of Mannheim.
- Garcia-Marin, A., Voigtlandern, N., 2019. Exporting and plant-level efficiency gains: It's in the measure. *J. Polit. Econ.* 127, 1777–1825.
- Gofman, M., Jin, Z., 2022. Artificial intelligence, education, and entrepreneurship. *J. Finance*. Forthcoming.
- Goldfarb, A., Taska, B., Teodoridis, F., 2023. Could machine learning be a general-purpose technology? A comparison of emerging technologies using data from online job postings. *Res. Policy* 52, 104653.
- Grennan, J., Michaely, R., 2019. Artificial intelligence and the future of work: Evidence from analysts. Working Paper.
- Grullon, G., Larkin, Y., Michaely, R., 2019. Are us industries becoming more concentrated? *Rev. Finance* 23, 697–743.
- Gutierrez, G., Philippon, T., 2017. Declining competition and investment in the U.S. NBER Working Paper w23583.
- Gutierrez, G., Philippon, T., 2019. Fading stars. *AEA Pap. Proc.* 109, 312–316.
- Haltiwanger, J., Hyatt, H.R., McEntarfer, E., Sousa, L., Tibbets, S., 2014. Firm age and size in the longitudinal employer-household dynamics data. Studies Paper No. CES-WP-14-16. US Census Bureau Center for Economic Studies.
- Hauptmann, A.G., Hodson, J., Li, J., Sebe, N., Rettinger, A., 2015. Cross-lingual cross-media content linking: Annotations and joint representations. *Dagstuhl Rep.* 5.
- Hershbein, B., Kahn, L.B., 2018. Do recessions accelerate routine-biased technological change? Evidence from vacancy postings. *Am. Econ. Rev.* 108, 1737–1772.
- Hirvonen, J., Stenhammar, A., Tuhkuri, J., 2022. New evidence on the effect of technology on employment and skill demand. Working paper.
- Hoberg, G., Phillips, G., Prabhala, N., 2014. Product market threats, payouts, and financial flexibility. *J. Finance* 69, 293–324.
- Hoberg, G., Phillips, G., 2016. Text-based network industries and endogenous product differentiation. *J. Polit. Econ.* 124, 1423–1465.
- Hobijn, B., Jovanovic, B., 2001. The information-technology revolution and the stock market: evidence. *Am. Econ. Rev.* 91.
- Hodson, J., 2016. How to make your company machine learning ready. *Harv. Bus. Rev.*
- Hottman, C.J., Redding, S.J., Weinstein, D.E., 2016. Quantifying the sources of firm heterogeneity. *Q. J. Econ.* 131, 1291–1364.
- Hsu, P.H., Li, D., Li, Q., Teoh, S.H., Tseng, K., 2021. Valuation of new trademarks. *Manag. Sci.* 68, 257–279.
- Humlum, A., 2019. Robot adoption and labor market dynamics. Working Paper.
- Janssen, M., Nguyen, H., Shams, A., 2020. Human vs. machine: Underwriting decisions in finance. Fisher College of Business Working Paper. 019.
- Jones, C.I., Tonetti, C., 2020. Nonrivalry and the economics of data. *Am. Econ. Rev.* 110, 2819–2858.
- Klette, T.J., Kortum, S., 2004a. Innovating firms and aggregate innovation. *J. Polit. Econ.* 112, 986–1018.
- Kline, P., Petkova, N., Williams, H., Zidar, O., 2019. Who profits from patents? Rent-sharing at innovative firms. *Q. J. Econ.* 134, 1343–1404.
- Kogan, L., Papanikolaou, D., Schmidt, L., Seegmiller, B., 2019. Technological change and occupations over the long run. Working paper.
- Kogan, L., Papanikolaou, D., Seru, A., Stoffman, N., 2017. Technological innovation, resource allocation and growth. *Q. J. Econ.* 132, 665–712.
- McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E., 1955. A proposal for the dartmouth summer research project on artificial intelligence. Reprinted online at <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>.
- Mihet, R., Philippon, T., 2019. The economics of big data and artificial intelligence. Disruptive innovation in business and finance in the digital world. *Int. Fin. Rev.* 20, 29–43.
- Organisation for Economic Co-operation and Development, 2019. Artificial Intelligence in Society. OECD Publishing, Paris.
- Peters, R., Taylor, L.A., 2017. Intangible capital and the investment-q relation. *J. Financ. Econ.* 123, 251–272.
- Rock, D., 2019. Engineering value: The returns to technological talent and investments in artificial intelligence. Working paper.
- Romer, P.M., 1990. Endogenous technological change. *J. Polit. Econ.* 98, S71–S102.
- Seamans, R., Raj, M., 2018. AI, labor, productivity and the need for firm-level data. NBER Working paper w24239.
- Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.J.P., Wang, K., 2015. An overview of Microsoft Academic Service (MAS) and applications. In: *International World Wide Web Conferences*.
- Stock, J.H., Watson, M.W., 2015. Introduction to Econometrics, 3rd updated edition. Age, vol. X3, p. 3.
- Tambe, P., Hitt, L., Rock, D., Brynjolfsson, E., 2020. Digital capital and superstar firms. NBER Working Paper w28285.
- Tanaka, M., Bloom, N.B., David, J.M., Koga, M., 2020. Firm performance and macro forecast accuracy. *J. Monet. Econ.* 114, 26–41. forthcoming.
- Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z., 2008. Arnetminer: Extraction and Mining of Academic Social Networks. Association for Computing Machinery, New York, NY, USA, pp. 990–998.
- Webb, M., 2020. The impact of artificial intelligence on the labor market. Working Paper.