

# Scientific Labor Market and Firm-level Appropriation Strategy in Artificial Intelligence Research

Nur Ahmed<sup>1</sup>

## Abstract<sup>2</sup>

This study examines the tension over appropriation strategy between firms and scientists, a key human capital. Whereas scientists prefer to publish, firms tend to minimize outgoing knowledge to maintain competitive advantage. This study investigates how a tight labor market, which affords scientists higher bargaining power, can influence firm publications. Using a novel dataset of 200 million job posts and 1.1 million publications from the US Artificial Intelligence (AI) industry, I show that recruitment efforts increase the number of AI publications, but primarily in the same fields of heightened demand. For identification strategy, I exploit the variation in AI exposure at the firm level, which directly influences firm-level demand for AI talents but not AI publications. A machine learning-based approach demonstrates that to balance the trade-off between knowledge leakage and recruiting, firms publish papers that are less commercially valuable. Further mechanism tests on the use of AI research in patents and the science intensiveness of AI patents bolster our theoretical explanation. Findings underscore the importance of human capital in firms' appropriation strategies.

**Keywords:** Corporate Science, Appropriation Strategy, Corporate R&D, Corporate R&D Disclosure, Human Capital, Scientists, Patent-Paper Pairs

---

<sup>1</sup>Email: nurahmed@mit.edu, Postdoc, MIT Sloan & CSAIL

<sup>2</sup>I would like to thank Matt Marx, Riitta Katila, Waverly Ding, Mahka Moeen, JP Vergne, Neil C. Thompson, Abhishek Nagaraj, Markus Simeth, Nan Jia, Sharon Belenzon, Lia Sheer, Milan Miric, Romel Mostafa, Subrina Shen, Leo Schmallenbach, Mayur P. Joshi, Andrew Sarta, Sukwoong Choi, and seminar participants at AOM, SMS, Bocconi BAIC, Industry Studies Association, NBER Seminar Series, MIT Initiative on the Digital Economy, AI & Strategy Group, The Max Planck Institute for Innovation, Northeastern E&I, Boston University TPRI, Ivey Research Series for their feedback and suggestions. Amit Das and Kawshik Banerjee have provided excellent research assistance. All errors are my own. I am grateful to Ivey's Fintech Research Lab for a grant to purchase the Burning Glass Technologies data.

# 1 Introduction

There is a “bloody war for talent in this [AI] space”

Peter Lee, Co-head of Microsoft Research ([Parloff, 2016](#))

“[...] if you try to recruit AI researchers by promising lots of money and zero peer recognition, you won’t get very far.”

Chris Nicholson, CEO, deep-learning startup SkyMind ([Alba, 2017](#)).

It is well-acknowledged that, due to appropriability concerns ([Teece, 1986, 2018](#)), firms tend to maximize incoming knowledge, minimize outgoing knowledge ([Alexy et al., 2013](#)), and disclose internal research strategically ([Gans et al., 2017](#); [Polidoro Jr and Theeke, 2012](#)) to maintain a competitive advantage. While patents are a well-established appropriation strategy ([Somaya, 2012](#)), publications offer fewer specific rights and may not effectively capture the value of R&D investments and increase the likelihood of expropriation ([Ding, 2011](#)). Furthermore, research suggests that to publish strategically, firms need “scientific disclosure capabilities,” or distinct competencies, workforce, and routines to reveal internal research selectively ([Simeth and Lhuillery, 2015](#)).

Different streams of literature have suggested inconsistent and varied answers to motivations behind a firm’s decision to publish papers ([Hicks, 1995](#); [Rosenberg, 1990](#); [Ding, 2011](#); [Simeth and Cincera, 2016](#)). One view is that publishing internal research allows firms to recruit talent ([Rotolo et al., 2022](#); [Hicks, 1995](#)). The key idea is that scientists have a “taste for science,” or a preference to publish papers that allow them to gain peer recognition ([Partha and David, 1994](#); [Merton, 1973](#)). This preference is acquired during educational training and is considered a strong norm in academic science ([Ding, 2011](#); [Merton, 1973](#)). This relationship is consistent with the evidence that scientists are willing to be paid less in exchange for the freedom to publish ([Sauermann and Roach, 2014](#); [Stern, 2004](#)). Thus, scholars posit that to recruit skilled scientists firms will need to publish more papers.

However, to date, there has been no evidence that firms cater to scientists’ non-pecuniary interests by using publications to recruit talent. We have evidence from the scientists’ side that they are willing

to let go of certain remuneration to publish papers (Stern, 2004). Surprisingly, we do not have any evidence on the *firms'* side of the equation.<sup>3</sup> This is important because publications could enable firms to economize labor costs by paying lower wages, specifically in industries, where spillovers have a limited impact on firm performance. However, recent research (Gans et al., 2017) notes this puzzle that firms do not show any such tendencies to lower scientists' salaries by publishing more papers, even in industries with lower expropriation risks.

Furthermore, research argues that there is an assortative matching between firms' and scientists' preferences (Agarwal and Ohyama, 2013; Roach and Sauermann, 2010). This line of research suggests that scientists with a higher preference for scientific publications tend to join academia and scientists with a lower preference for publishing tend to join the industry (Agarwal and Ohyama, 2013). This assortative matching is also consistent with research suggesting that not all scientists have a strong preference for publications (Sauermann and Roach, 2014). Thus, inasmuch as assortative matching exists between scientists' and firms' preferences, then firms should be able to recruit scientists with lower preferences for publications even without publishing papers. Furthermore, it is still unclear whether firms consistently resort to publications or only in specific cases (or in specific industries) to recruit scientists.<sup>4</sup> This lack of clarity leads to my research question: *To attract scientists, under which conditions do research-active firms increase publications?*<sup>5</sup>

Additionally, it is unclear how firms balance the trade-off between recruiting scientists and strategically releasing internal knowledge. For instance, firms could disclose papers that they have already filed patents for in order to reduce expropriation. Put simply, firms could use patent-paper pairs (Ducor, 2000; Murray, 2002)—the simultaneous disclosure of the same invention in both patents and papers—to balance the trade-off. Furthermore, it is unclear whether firms prioritize increasing their general research reputation or specialized reputation (e.g., reputation in that specific

---

<sup>3</sup>One exception is (Ding, 2011)'s work, which documents that even when there are expropriation risks, startups founded by scientists with PhDs have a higher probability to publish scientific papers.

<sup>4</sup>It is important to note that scientists have different levels of preference for publications in different industries. For instance, in engineering fields like Computer Science, scientists have *lower* preferences for publications (Sauermann and Roach, 2014).

<sup>5</sup>In this study, I limit my attention to firms' *attempts at hiring* scientists rather than actual recruitment.

sub-field) in order to attract scientists.

Drawing on the innovation literature, I argue that organizations' shift toward increased academic publishing happens, in part, due to scientists' increased bargaining power (Brown et al., 2015; Coff, 1999; Gans et al., 2017). Satisfying the interests and demands of scientists, a key stakeholder, is important for firms because they provide critical resources (e.g., access to human capital, which often relies on tacit knowledge). Prior research suggests that scientists tend to have higher bargaining power in knowledge-intensive firms (Kehoe and Tzabbar, 2015). I argue that competition for their unique knowledge and skills allows AI scientists higher bargaining power with firms. Consequently, scientists' increased bargaining power influences the extent to which a firm reveals internal research to recruit skilled scientists.

To answer my research question, I use Artificial Intelligence (AI), where both publishing and patenting are prevalent, as a context. Notably, recent empirical evidence (Hartmann and Henkel, 2020; Ahmed et al., 2023) suggests that in Artificial Intelligence (AI), a large number of firms—including secretive firms like Apple and Amazon—have increased their research presence (see Figure A4). This contrasts with the well-documented secular decline of corporate participation in basic science (Arora et al., 2018; Mowery, 2009). Recent evidence highlights the importance of scientists for this particular technology, specifically for deep learning, which has experienced a tight labor market in recent years (Gofman and Jin, 2022). As a result, the demand for AI talent has led to higher bargaining power for scientists and a subsequent wage premium (Alekseeva et al., 2021).

Using a novel dataset of 1828 firms that have at least one AI patent I test my propositions. I complement my quantitative data with rich qualitative interviews, archival data, and field observations. I use 200 million firm-level job posting data from Burning Glass Technologies as a proxy for firm-level labor demand for that specific talent. Employing a two-way fixed effects model, I find that firms reveal more internal research in publications when they need to hire AI talent. The estimates suggest that a one standard deviation increase in job posts would increase the number of publications by 1.22 papers each year by each firm, which is a substantial change for most research-active firms. This observed increase in publications is over and beyond the increased salaries for AI scientists and

controlling for firm-level AI patents.

Empirically, I employ a variety of robustness analyses, including a difference-in-differences (Gardner, 2022; Athey et al., 2021) and an instrumental variable design to establish that scientists' bargaining power led to increased AI publications. I leverage a recently developed AI exposure measure (Felten et al., 2021) to provide more insight into the underlying mechanisms. The identifying assumption is that AI job posts are greater when firm-level AI exposure is greater for reasons unrelated to firms' decisions to publish. My difference-in-differences study shows that firms with higher exposure to AI start to increase publications only from 2012 when AI scientific labor market became a key issue. I complement this with an instrumental variable strategy to show that firms' numbers of AI job posts rise in response to AI exposure (Acemoglu et al., 2020), which, in turn, increases the number of AI publications.

Additionally, using a novel machine learning technique, I show that firms balance the trade-off between expropriation and recruiting by disclosing papers that are *unrelated* to their patents. In other words, firms are disclosing less commercially valuable papers to protect internal knowledge. This appropriation strategy allows firms to protect commercially valuable knowledge and simultaneously recruit scientific talent.

I investigate possible mechanisms to explain the increased publications in AI. First, I find that the effect is more pronounced when firms need to hire AI talent with higher salaries. To further isolate the mechanism, I show that startups are likely to be more concerned with expropriation (Ding, 2011; Katila et al., 2008), and thus do not increase publications. Additionally, I use a placebo test to demonstrate which shows that job posts with PhD requirements in non-AI areas do not increase non-AI publications. This is because non-AI fields did not experience a tight labor market. These tests give credence to the idea that it is the bargaining power that leads to increased publications in AI.

Second, examining the intensity of AI patents that rely on science (Marx and Fuegi, 2020), I rule out the possibility that in recent years AI has become more science-intensive. I find that the

percentage of commercial AI patents that rely on science has largely stayed the same. Additionally, I find similar results when I consider the use of AI science in firms' AI patents. Finally, exploiting the granular nature of job posts and publications, this study sheds light on the “black box” of a firm's signaling mechanism to recruit scientists. More specifically, I highlight that firms primarily increase publications in the areas (e.g., deep learning) where they are posting jobs instead of increasing general research reputation (e.g., general AI or other non-AI research fields).

This study contributes to the innovation strategy literature in three important ways. First, I resolve a long-standing puzzle in the innovation literature on whether firm-level publications matter for recruiting scientists. This theory explains under what conditions firms use the publication as an instrument to hire human capital. In contrast to previous literature, this study suggests firms tend to change their appropriation strategy when scientists have higher bargaining power. By focusing on human capital for appropriation strategy (Kang and Lee, 2022), this paper brings scientists' bargaining power to the forefront of the disclosure literature. To the best of my knowledge, this is also the *first empirical study* to link firm-level publications with firm-level human capital needs.

Second, this study elucidates how firms balance the trade-off between recruiting scientists and selectively revealing internal knowledge. More specifically, I explore an important facet of appropriation strategy—patent-paper pairs (Ducor, 2000; Murray, 2002)—which received limited empirical attention in the strategy literature. This study's result paves the way for future research regarding when and why firms tend to disclose in patent-paper pairs.

Finally, the extant strategy literature has mostly focused on factors such as the stock market (Simeth and Cincera, 2016; Pellens and Della Malva, 2018) or competition (Pacheco-de Almeida and Zemsky, 2012; Polidoro Jr and Theeke, 2012) that affect a firm's decision to publish. Contrary to this, this study provides evidence that shows the salience of human capital and the external labor market (Starr et al., 2018a) in appropriation strategy. More specifically, I highlight an important trade-off in knowledge-intensive industries: the benefits of hiring a talent versus the risks associated with R&D disclosure. Taken together, this study emphasizes the need for innovation scholarship to envision appropriation strategy in relation to scientists, an important conduit of knowledge.

## 2 Literature Review

### 2.1 The Downside of Publishing

Disclosing internal R&D information is a critical decision for a firm's appropriability strategy (Teece, 1986). Survey-based evidence suggests that firms prefer lead time and secrecy over patents or publications (Cohen et al., 2000; Hall et al., 2014; Richard et al., 1987). Despite its limitations, the benefits of patenting are well-acknowledged in the literature due to relatively direct and clear intellectual ownership (Bhaskarabhatla and Hegde, 2014; Somaya, 2012). However, disclosing R&D in peer-reviewed articles is a more contentious area in the literature, with inconsistent arguments and varied empirical results.

There are multiple reasons why knowledge-intensive firms might want to avoid publishing. First, publishing internal research can heighten the risks of expropriation (Ding, 2011). Getting accepted at top academic outlets requires a certain novel contribution, and the research has to go through a rigorous academic review process (Merton, 1973). In other words, firms will have to reveal certain "interesting" and "novel" work to get accepted into academic outlets. Further, firms will also need to provide significant details of an invention for academic reviews. Thus, participating in a peer-reviewed process increases the risk that competitors will scoop the published discovery or will find a workaround to file the same or a similar invention before the focal firm. For instance, Cistron Biotechnology scientists submitted a paper to the journal *Nature* in 1984 for peer review describing the gene sequence of *interleukin 1*. Interleukin is a naturally occurring protein in the immune system with multiple therapeutic applications. One of the reviewers, Steven Gillis, was the head of the research department at a rival company named Immunex Corporation (Ding, 2011). Later, Cistron claimed that Immunex Corporation inappropriately used the chemical sequence for a gene critical in making interleukin. In the end, after a three-year protracted litigation, Cistron won a \$21 million settlement in 1996. This anecdote depicts the challenges associated with appropriation from innovation while trying to engage with the broader academic community.

This is consistent with prior literature, which acknowledges the key role knowledge spillover plays in a firm's decision to invest in research and disclose them publicly (Arora et al., 2021; Simeth and Raffo, 2013). In general, knowledge spillover reduces a firm's rent from innovation as rivals can capture some of the benefits from the innovator's investment. Therefore, knowledge spillover reduces the cost of rival firms' entry at the expense of the focal firm. This inability to capture profit due to knowledge spillover motivates firms to reduce publications.

A second reason knowledge-intensive firms might want to avoid publications is that publishing papers in reviewed outlets requires distinct firm-level capabilities and is expensive (Kleid, 2002; Simeth and Lhuillery, 2015). Research (Simeth and Lhuillery, 2015) suggests that publication is *not* a by-product of internal research activities; rather, firms need to have a specific R&D staff composition to enable disclosure. For instance, firms require competencies to codify knowledge in a standard manner and then selectively reveal a specific part of their internal research. Specifically, for publishing, firms have an internal review process to check for intellectual property and to avoid unwanted spillovers (Kleid, 2002). In general, if a paper has an idea worth patenting, lawyers tend to encourage the scientist to patent the idea.<sup>6</sup> Therefore, this review process often requires both scientists and lawyers or IP experts. Jeff Dean, Google's Senior Vice President of Engineering, describes their internal process: "[Our internal review process is] more than just a single approver or immediate research peers; it's a process where we engage a wide range of researchers, social scientists, ethicists, policy & privacy advisors, and human rights specialists from across Research and Google overall.[. . .] While more than 1,000 projects each year turn into published papers, there are also many that don't end up in a publication."<sup>7</sup> In other words, codification to internal review is expensive and time-consuming for firms. Overall, publishing in peer-reviewed outlets requires time and resources for scientists and lawyers, which can divert resources from other profit-oriented

---

<sup>6</sup>One illustrative quote is from a leading scientist at Genentech: "What usually happens at Genentech is the scientist wants to publish his paper, or he wants to go to a meeting and talk about his project. So before he does that, he's got to go to the patent attorneys and see if there's anything patentable there. If there is, then they've got to write up a patent [application] for it before he sends in his paper. So there is a trade-off between those two things, and actually this [policy] caused all of Genentech's major discoveries to be published and have patent applications filed very quickly after the work was done" (Kleid, 2002, p. 110).

<sup>7</sup>This is notable that Google, a well-known open innovation advocate, does not publish all of its research projects. Source: <https://tinyurl.com/googAI2>



activities.

Third, publications could reduce friction for employees to move from one firm to another. More specifically, highly skilled employees such as scientists are likely to take advantage of the investments in their human capital made by their firms (Starr et al., 2018b). A publication based on the focal firm's investment would allow a scientist to be more visible within the scientific community. Therefore, by publishing additional papers, firms could increase the likelihood that scientists would leave the firm. Accordingly, firms are mindful of the risks associated with the departure of their human capital and respond strategically (Agarwal et al., 2009; Kang and Lee, 2022). For instance, prior research (Kang and Lee, 2022) documents that firms apply for more patents under a heightened threat of employee departure to protect knowledge. Consequently, organizations are likely to be more strategic in publications to ensure that scientists do not benefit at the expense of firms. They may even choose to reduce the number of publications.

Finally, publications may impede the patenting process for specific technologies by creating a prior art for patents (Rubin, 2011). A published paper as a prior art has an expiring deadline of one year for the USPTO. In other words, firms need to file a patent within a year after publishing a paper to file with the USPTO. This time constraint puts firms at a strategic disadvantage in patenting because they may be forced to apply for a patent before the firms are ready to disclose that invention (Kim et al., 2016). When innovation is considered a race between competing firms and technologies, the timing of a certain disclosure is particularly important. Viewed from this perspective, the publication process adds another bottleneck to a firm's appropriability strategy.

## **2.2 Publications as an Appropriation Strategy**

Given the high risks associated with publishing, firms tend to consider both the benefits and costs associated with publications (Toh and Miller, 2017; Gans et al., 2017; Polidoro Jr and Theeke, 2012). First, research suggests that participating in open science is conducive to accessing external knowledge, specifically from academia. In other words, publishing is a "ticket of admission" to

communicate with a broader research community (Rosenberg, 1990). Publishing in top academic venues provides firms with credibility to be in an exchange relationship with academics (Hicks, 1995). Put simply, a firm's research reputation allows participation in knowledge exchange with academic collaborators, and, in turn, increases organizational absorptive capacity (Cockburn and Henderson, 1998; Zucker et al., 2002; Cohen and Levinthal, 1990).

Second, the literature also posits that publications act as an instrument to recruit scientists (Hicks, 1995; Rotolo et al., 2022). Hicks (1995, p. 413, emphasis added) notes that “[a]pparently, good salaries, excellent equipment and freedom from teaching, administrative duties and all the hassle of university research are not enough to entice top scientists into corporate work; the opportunity to publish proves a further necessary inducement. Enhanced competitiveness in recruiting might be the *only* benefit companies derive from publishing by their scientists.”<sup>8</sup> Publications are important to scientists because they have a particular preference for publications (a subject upon which I will elaborate in the following section of this paper). However, until now, there has not been any empirical evidence that firms use publications to recruit talent.

Third, firm publications work as a signal to communicate with external stakeholders such as the stock market (Simeth and Cincera, 2016; Pellens and Della Malva, 2018; Arora et al., 2018). However, literature has documented varied results and inconsistent responses by markets to firm publications. For example, Simeth & Cincera (2016) find evidence that publishing in high-impact factor journals positively correlates with a firm's market value. The authors argue that the underlying mechanism is that publishing works as a signal of a firm's research quality to the market. However, they find that the market does not value publishing for firms in the Information Communication Technology (ICT) industry. To explain the industry heterogeneity, Pellens & Malva (2018) propose that the market reacts positively when a firm's business model is congruent with its disclosure strategy. The authors find that upstream firms are rewarded, and downstream firms are penalized for publications. The reason is upstream firms tend to be more knowledge-intensive, and research

---

<sup>8</sup>In this study, the author's own interviews and archival data suggest that leading academic scientists were demanding *both* higher salaries and the option to publish papers in academic outlets.

quality from their publications helps the market to discern their underlying quality. In summary, not all firms benefit from the market by publishing research articles, and appropriation strategy is an important concern. However, more recently, Arora et al. (2018) documented that the stock market is valuing publications less over time.

### **2.3 Scientists' Preference for Publications and the Tension between Scientists and Firms**

Research suggests that scientists have a “taste for science” or a strong preference to publish papers in academic outlets (Partha and David, 1994; Merton, 1973). Scientists are part of a professional group with their own norms and values; in particular, scientists who gain PhDs are also socialized in a way that encourages publishing. This professionalization leaves an imprint on scientists even if they leave academia. Scientists tend to publish to maintain relationships with other scientists and to gain higher status within their communities (Merton, 1973; Gittelman and Kogut, 2003). For instance, empirical evidence suggests that startups founded by scientists have a higher propensity to publish (Ding, 2011). This relationship demonstrates that even after leaving academia scientists tend to follow scientific norms. Furthermore, prior literature posits that science is manufactured in laboratories where scientists compete for recognition and status within their respective communities for publishing, at times, at the expense of the firm (Gittelman and Kogut, 2003; Latour and Woolgar, 1979; Merton, 1973; Starr et al., 2018b). In sum, the empirical literature has documented that the benefits of publications accrue to individual scientists; however, the costs are borne primarily by firms.

Overall, there is a tension between scientists' propensity to publish and firms' appropriability strategy (Gans et al., 2017). Science as an institution is driven by a priority-based reward system where scientific prestige is awarded to the first person for discovery, and the second discoverers get little or nothing (Partha and David, 1994; Merton, 1973). This system motivates scientists to focus strategically on a research topic and expedite publication. On the other hand, firms tend to

consider the strategic benefits of keeping a discovery secret. Chief Scientific Officer of Millennium Pharmaceuticals, Robert Tepper, summarized this tension as follows: “There are times when we’ve held back a publication when the scientist would have preferred to publish. We realize that that’s a delicate balance, not only for our scientists but also for [the company’s] well-being. (Quoted in [Polidoro Jr and Theeke, 2012](#), p. 1138).” This discrepancy in priorities creates tension between scientists and firms. This particular tension often comes down to the bargaining power of each side.

## **2.4 Scientific Labor Market and Increased Publications**

Research acknowledges the importance of power that skilled individuals, such as scientists, hold over an organization ([Coff, 1999](#); [Gittelman and Kogut, 2003](#); [Kehoe and Tzabbar, 2015](#); [Kim and Mahoney, 2007](#)). More specifically, the bargaining power of skilled employees plays a vital role in appropriating value. For instance, tacit knowledge increases an employee’s influence within a knowledge-intensive organization ([Coff, 1999](#)). The reason for this relationship is that oftentimes organizations depend on their employees’ tacit knowledge to create value. In particular, when knowledge is characterized by both scarcity and tacitness, human capital becomes the central resource for the formation and transformation of firms’ competitive advantage ([Zucker et al., 2002](#)). Consequently, scientists can have potentially high bargaining power at a knowledge-intensive firm ([Kehoe and Tzabbar, 2015](#)).

A prime example of the significant bargaining power of scientists can be observed at the well-known pharmaceutical company Merck. The firm pursued a drug discovery program and later distributed it for free partly to ensure that the morale of the research scientists was not damaged ([Kim and Mahoney, 2007](#)). Research highlights Merck’s unique capability to attract and retain highly qualified scientists them. Kim and Mahoney (2007, p. 21) summarized it as follows: “Research scientists come to Merck because Merck offers them the resources and the freedom to pursue research projects that are not necessarily the most economically profitable ones, and these scientists are essential for the competitive advantage of Merck in developing pharmaceutical drugs.” In sum, scientists have significant power in knowledge-intensive firms, which, in turn, often make changes

to accommodate their demands, even if it is expensive for firms.

Prior research (Brown et al., 2015; Molloy and Barney, 2015) acknowledges that a tight labor market allows a firm's employees to capture higher value due to increased bargaining power. In particular, when scientists realize that their skill is non-substitutable, and thus, they are not easily replaceable, and they have alternative options (e.g., other firms or universities to choose from), it makes them less reliant on focal firms. Further, the additional costs of replacing a noted scientist increase scientists' bargaining power to demand changes. Overall, the external labor market shortage gives scientists more power over firms, and thus they can demand more from their employers. For instance, Brown et al. (2015) documented how the bargaining power of physicians triggered organizational changes at a hospital and resulted in a new organizational form. Following this logic, I argue that scientists can use their bargaining power to trigger certain organizational changes like firm-level appropriation strategy.

Recent theoretical literature has modeled the tension between scientists and firms as an ongoing negotiation on R&D disclosure strategy (Gans et al., 2017). More specifically, Gans et al. (2017) posit that while scientists try to maximize their pecuniary and non-pecuniary benefits, firms tend to focus only on their bottom line. Thus, when scientists have higher negotiating power, firms will change their disclosure strategy and publish more papers in academic outlets. Anecdotally, the importance of allowing publishing and the reputation of an organization's research ability are highlighted by the quote of Yann LeCun, Facebook's chief AI scientist and the winner of 2019's Turing Award:

[Y]ou can't attract the best scientist unless you allow them to publish . . . publishing innovative research contributes to establishing the company as a leader and innovator. This helps recruiting the best people. In the tech industry the ability to attract the best talent is everything.

Based on the above discussion, I hypothesize:

*Hypothesis 1: In a knowledge-intensive industry, when there is a tight labor market for*

*scientists, a research-active firm's increased demand for scientists will be associated with increased publications in that research area.*

### **3 Empirical Context**

I test my propositions with data from 2000 to 2019 in Artificial Intelligence, a knowledge-intensive industry. Previous research ([Gofman and Jin, 2022](#)) notes the importance of scientists in this sector.

#### **3.1 Tight Scientific Labor Market in AI**

The AI industry experienced a sudden increase in interest and investment due to a technological breakthrough known as the “deep learning revolution” ([The Economist, 2016](#)). The increased interest resulted in a significant demand for AI labor. The increased demand for AI talent has created a competitive labor market([Gofman and Jin, 2022](#)). Co-head of Microsoft Research Peter Lee says it is a “bloody war for talent in this [AI] space,” which led to top talent demanding salaries “along the lines of NFL football players” ([Parloff, 2016](#)). Furthermore, recent research from job posting data suggests that it is not only AI scientists but also engineers with AI knowledge who are experiencing a wage premium ([Alekseeva et al., 2021](#)). The intense competition for talent has also forced firms to hire talent away from universities ([Gofman and Jin, 2022](#); [Ahmed et al., 2023](#)). In sum, it is well-documented that AI as a field has gone through a competitive labor market for scientists.

Figure 1 demonstrates how the demand for AI talent surpassed the supply of AI talent between 2010 and 2019. The Figure shows that AI job posts with PhD requirements surpassed the total number of students who completed PhD degrees. This is in contrast to non-AI research fields (see Figure A1 in the Appendix), where the supply of PhDs is much higher than the demand for PhDs for the firms in our sample, which is consistent with prior observations ([Charette, 2013](#); [Cyranski et al., 2011](#)).

INSERT FIGURE 1 HERE

## 3.2 Qualitative Data

To better understand the emerging AI field, I collect rich qualitative and quantitative data. I also participate in a few important AI conferences including the Conference on Neural Information Processing Systems (NeurIPS), the leading AI conference. Additionally, I interview 24 computer scientists and research managers affiliated with firms and universities (see Table A1). These interviews (presented in Table 1) follow a semi-structured format and last between 28 minutes and 75 minutes. In my field observations at AI conferences, I noticed that firms had large displays and distributed colorful pamphlets to showcase their latest peer-reviewed publications. They would also invite their top AI scientists to present research and announce their vacancies.

A key objective of this qualitative research is to delve deeper into how different actors view AI research and why they engage with it. I ask respondents about what led to the sudden rise of AI. Unequivocally, more than 90% of them mention ImageNet moment (see more in the identification strategy). ImageNet moment demonstrated that deep learning can be made to work, as the reason for increased interest in AI. Some respondents liken the ImageNet shock to a “breakthrough moment” within AI research. Further, I also inquire about why there is an increased presence of companies at AI conferences. More than 80% of the respondents suggest that firms are publishing to recruit talent and the shortage of skilled AI talent is the key driver of increased firm publications.

INSERT TABLE 1 HERE

## 3.3 Case Studies: Apple and Amazon’s Change in Appropriation Strategy

Since the widespread adoption of AI across the technology industry, industry observers noticed that Apple was struggling with this technology. For instance, in 2015, *Bloomberg* wrote an article arguing that “Apple is ramping up AI efforts, but the company’s reticence to publish its research is limiting its effectiveness and hiring” (Clark, 2015). This report highlights that Apple was struggling to recruit AI talent who have a strong preference for publications. Later on, in 2016, Apple’s

decision to open up about AI publications received extensive media coverage (Metz, 2016; Statt, 2016; Tilley, 2016). For instance, *Wired* magazine wrote an article titled “Artificial Intelligence Just Broke Steve Jobs’ Wall of Secrecy,” citing the increased openness as a strategy to recruit talent (Metz, 2016). Others have echoed similar points (Dvoskin, 2017; Alba, 2017; Mickle, 2017). The CEO and co-founder of deep-learning startup Skymind, Chris Nicholson, summarized it as follows: “[T]hose people like to publish. Publishing is like breathing to them: You do it or you die [. . .] So if you try to recruit AI researchers by promising lots of money and zero peer recognition, you won’t get very far. There are some people who will never join Apple for that reason” (Alba, 2017).

Similarly, Amazon, another relatively secretive firm, struggled to recruit AI talent due to a lack of openness (Clark, 2015; Levy, 2018). For example, *Wired* magazine argued that the company’s culture of secrecy hurt its efforts to recruit top AI talent. Later when Amazon started to publish AI papers that facilitated their talent recruitment. Spyros Matsoukas, Senior Principal Scientist at Amazon, stated, “[publishing has] helped quite a bit with recruiting top talent as well as providing visibility of what type of research is happening at Amazon” (Levy, 2018).

Overall, these cases illuminate how, despite formidable resources, large firms such as Apple and Amazon struggled with recruiting AI scientists. Notwithstanding higher salaries, research environment, and access to data and computing power, scientists shied away from joining these two organizations due to their propensity to maintain secrecy. However, the tight labor market afforded scientists additional bargaining power, which, in turn, changed these firms’ appropriation strategy, at least in AI.

## 4 Quantitative Data

For my quantitative analysis, I rely on multiple data sources to construct the dataset. I collect (a) patent data from the USPTO, (b) firm-level data from Compustat, (c) publications data from Scopus, and (d) job posting data from Burning Glass Technologies (BGT). BGT collects data from 40000 American job sites and is widely used for labor market-related literature (Acemoglu et al., 2020;



[Alekseeva et al., 2021](#)). In total, I have 200 million job posts, a comprehensive dataset of the whole U.S. job market data.<sup>9</sup>

To construct my sample, I start with the USPTO data. I define a firm with AI research if that firm has at least one AI patent under the CPC class “computer systems based on specific computational models.” This CPC class was selected after extensive consultations with two USPTO examiners. Thus, I collect data on all the firms with at least one AI patent. After that, I collect all of their patent portfolios, both AI and non-AI patents. Next, I merge this data with their publications data from Scopus. All the different variations of a firm name are used to look for Scopus publications. This data collection entailed a combination of manual verification and Python web scraping given the nontrivial number of variations of these firm names. As before, for publications, I have each firm’s AI publications and non-AI publications. Finally, I merge this data with the job posting data from BGT. To ensure accuracy and completeness, I painstakingly refined these different data sets and corroborated them manually.

The empirical results focus on firms that had some operation in the US, given that all the job posts are also from the US. The sample consists of a panel dataset of 18280 firm-year observations from 2010 to 2019.<sup>10</sup> This data set improves and extends the prior studies (e.g., ([Arora et al., 2018](#))), which primarily focused on US public firms. Notably, this sample includes both public firms and startups.

---

<sup>9</sup>One potential concern is whether the BGT data accurately reflects firm-level hiring attempts since BGT data allows us to observe only job posts but not hiring. However, using both BGT’s job posting data and Cognism Ltd’s hiring data, Babina, Fedyk, He, and Hodson (2023) find that these two are highly correlated. In other words, BGT’s job posting data is a reasonable representation of firm-level hiring. Furthermore, even if there is a gap between job posts and actual hiring, it would not affect my theory, since it predicts that increased demand is enough for triggering a change in publications.

<sup>10</sup>Here, I am counting DeepMind, Waymo and Google separately. The results still hold, when I combine them at the parent level. Overall, from 2000 to 2019, the total number of observations comes to 35920. However, this particular data does not have AI Job post data for earlier years since BGT data starts in 2010. Therefore, except for difference-in-differences, in every other analysis, I limit my attention to 2010 to 2019 data.

## 4.1 Empirical Approach

**Dependent Variable:** The propositions relate to scientists' bargaining power and firm-level publications. To calculate that,  $\ln(\text{theNumberOfAIpublications}_{it} + 1)$  is used as the number of publications. As a publication, I include both journal articles and conference proceedings because, depending on the fields, firms might prefer one over the other (Gofman and Jin, 2022; Pellens and Della Malva, 2018; Simeth and Cincera, 2016). After that, I classify AI papers based on an extensive list of keywords (see Appendix A1) which were prepared with my consultation with computer scientists.<sup>11</sup>

**Independent Variables:** To test the influence of bargaining power I use firm-level job posting as an independent variable. To calculate AI job posts I take  $\ln(\text{NumberOfAIjobposts}_{it} + 1)$ , which takes into account the skewed nature of job posts by firms. To create  $\text{NumberOfAIjobposts}_{it}$  I build on the keywords list by Alekseeva et al. (2021). From the BGT dataset, I remove part-time or intern job posts, as full-time employees possess greater bargaining power with firms compared to part-timers or interns.

*AI PhD Job Posts* is another independent variable. To create this variable first, I classify AI job posts as before and then look for educational requirements with a PhD degree. If an AI job post has a PhD degree requirement, I count that under this variable. In the same way, I calculate *DL PhD Job Posts* by classifying deep learning (DL) job posts and examining if the job post had any PhD requirements. To calculate *HigherSalary AI Job Posts*, first, I classify AI PhD job posts as before and then look for higher salary information. If an AI PhD job post has a salary of more than 100000 USD in base salary, I count that under this variable. All of these variables have been log-transformed as before.

**Control Variables:** I control for several potential factors that might affect the number of publications by firms. I account for firms' scientific capabilities and also their propensity to publish by including annual *non-AI publications*, which is the total annual count of non-AI publications at

---

<sup>11</sup>For additional robustness analysis, I have also used citation-weighted paper counts, which produce similar results (available upon request).

the firm level.

*AI Patents*: I also include the annual number of AI patents filed by firms to control AI research capability. Additionally, I control for firms' *non-AI patents* or all the non-AI patents granted in a given year. To calculate that, I subtract AI patents from all the patents for a given year. Literature suggests that having complementary resources allows firms to disclose more (Toh and Miller, 2017). Non-AI patents are a proxy for a firm's complementary capabilities in non-AI areas. *AI pub Trends* is a variable that measures the annual non-industry AI publications. To calculate that, I count all the annual AI papers in Scopus and subtract annual industry AI papers from them. This measure helps us to control for the annual publication trend in AI. *AI PhD Supply* is the annual count of AI PhDs in US universities, which comes from the CRA.

Following previous literature, I include a firm's *R&D investment*, which is a logged measure of total R&D expense (in millions of dollars). I also include a firm's annual *sales*. These two variables control for two important firm-level factors that can influence firms' increased presence in AI. Additionally, I include *Tobin's Q* to take into account firms' market value. The summary statistics are presented in Table 2.

INSERT TABLE 2 HERE

## 4.2 Model Estimation

Following previous literature, I use a two-way fixed effect (TWFE) model. The regression equation is given as

$$\begin{aligned} \ln(AIPapers_{it} + 1) &= \alpha + FirmID_i + V_t + \beta * \\ \ln(AIPhDJobPosts_{it} + 1) &+ \beta_2 * X_{it} + \epsilon_{it} \quad (1) \end{aligned}$$

where  $i$  represents the firm, and  $t$  represents time. The dependent variable: log of the total number of annual AI publications at the firm-level.  $FirmID_i$  is the firm fixed effect, which allows

me to control for within-firm drivers of changes related to the number of publications.  $V_t$  is the time-fixed effect to control for potential market changes that affect a firm's number of publications. The main explanatory variable is the number of AI job posts by firms in that given year: *AI PhD Job Posts<sub>it</sub>*. Standard errors are clustered at the firm level since treatment (i.e., labor demand) occurs at this level.

Table 3 presents the estimates of the influence of labor demand on firm-level publications. Model 1 tests the hypothesis, which predicts that increased AI PhD job posts will be associated with increased publications from firms. Consistent with this hypothesis, the coefficient on the number of AI job posts is positive and significant ( $\beta = 0.121$ ,  $p = .000$ ). The result from Model 1 suggests that firms with higher AI PhD job posts have a positive and significant relationship with AI publications. These relationships still hold after controlling for firms' AI patents, non-AI patents, and other firm-level factors that could affect the relationship. Model 1 suggests that a 1% increase in job posts results in a .121% increase in publications. In other words, one standard deviation increase in AI PhD job posts would increase the number of publications by 1.22 papers each year per firm, which is a substantial number of publications when aggregated over a period of time for hundreds of firms.<sup>12</sup>

INSERT TABLE 3 HERE

To gain deeper insights into potential heterogeneity in firm-level response to increased bargaining power we analyze both public firms and research firms. Following (Arora et al., 2018) research firms are defined as those with at least one publication. Model 2 presents the findings for research firms. The effect size ( $\beta = 0.081$ ,  $p = .000$ ) is smaller than Model 1 ( $\beta = 0.121$ ). Model 3 reports the results ( $\beta = 0.105$ ,  $p = .000$ ) for public firms that also include additional financial controls, the estimates are closer to Model 1.

The labor market demand was the highest in the deep learning domain (Gofman and Jin, 2022). Accordingly, Deep Learning PhDs are expected to hold greater leverage, resulting in more AI papers.

---

<sup>12</sup>Here,  $\exp(0.121 * p) \approx 1.22$ , where,  $p = \log(1 + 0.607/0.154) \approx 1.61$

Models 4-6 investigate this hypothesis. I find that all these models produce similar effect sizes. Likewise, Models 8-10 examine the influence of AI PhD job posts offering *higher salaries* on the substantial growth in the number of AI papers. These outcomes are consistent with my qualitative data, which reveal that AI scientists sought *both* increased salaries and the capacity to publish. Although not statistically significant, the effect sizes from these Models (8-10), appear marginally larger in comparison (relative to Models 1-3). Therefore, these results provide suggestive evidence that bargaining power was a major mechanism.

### 4.3 Estimates from Difference-in-Differences

The TWFE models have some strong assumptions ([Gardner, 2022](#); [Liu et al., 2022](#)), including homogeneous treatment effect. Furthermore, a key concern is that publications are not random events; rather firms are very strategic about their disclosure ([Gans et al., 2017](#)). To address these concerns, this study leverages an exogenous shock to the AI field to demonstrate the impact of increased AI labor demand on firm-level publications.

### Identification Strategy

Since 2010, the ImageNet contest has been evaluated large-scale AI models for object detection and image classification. In 2012, a team of three scientists from the University of Toronto at the ImageNet contest demonstrated that “deep learning” based AI models work better with substantial computing power. Specifically, they demonstrated that Graphics Processor Units (GPUs) could be *repurposed* to train deep learning to produce significantly superior results. Their model outperformed the second-best model by an astonishing 70% that year, catching both scientists, including the ImageNet contest organizer (“unexpected outcome”), the winning team and other scientists, off guard by the unanticipated success of deep learning.

Consequently, media outlets ([Gershgorn, 2018](#); [The Economist, 2016](#)) and computer scientists ([Alom et al., 2018](#); [Russakovsky et al., 2015](#)) have acknowledged ImageNet 2012 as a watershed

moment in AI's trajectory. As articulated by *The Economist*: “The rehabilitation of ‘AI’, and the current excitement about the field, can be traced back to 2012 and an online contest called the ImageNet Challenge” (2016). (For additional qualitative evidence see Appendix A3).

The ImageNet moment demonstrated that GPUs could be repurposed to train deep learning models. However, that does not guarantee that these models will be economically valuable for firms. My interviews and archival data suggest that making AI required some experimentation (i.e., trial and error) and customization at the firm level. Most of the knowledge was tacit at that stage and resided within AI scientists (Zucker et al., 2002). Therefore, firms needed to recruit AI scientists to acquire this technology. Consequently, this unanticipated shock resulted in a ten-fold increase in the demand for AI talent (Alekseeva et al., 2021), and industry started to poach AI scientists from universities (Gofman and Jin, 2022; Ahmed et al., 2023). However, this is not correlated with the ex-ante increase in firm-level AI publications. Therefore, in the next section, I use ImageNet 2012 as an exogenous shock.

## **The Difference-in-Difference (DiD) Results based on AI Exposure**

*Classifying the Treated and Control Firms:* This study utilizes the “AI Occupational Exposure” measure (Felten et al., 2021), which estimates the likelihood of occupations being impacted by AI advances. This allows for the calculation of each firm’s AI exposure based on its occupational structure. Acemoglu et al. (2020) expanded this measure to a firm-level metric that varies over time based on job posts. Following their methodology, I compute each firm’s annual AI exposure as the sum of all job posts weighted by their corresponding occupational AI exposure (see Appendix A2). Since firms post different job posts in different years, we get a time-variant measure for each firm.

Next, firms are classified into two groups based on their *median* AI exposure. Those with exposure scores above the median industry level are deemed “treated” units, while those below are considered “control” units.<sup>13</sup> There are three reasons why it is not obvious why firms with higher AI

---

<sup>13</sup>For robustness tests, I took classified the top 40% in terms of firm-level AI exposure as treated” units. Conversely, firms constituting the bottom 40% in mean AI exposure scores are categorized as “control” units. The results are similar

exposure will start publishing more. Notably, higher AI exposure does not automatically lead to increased engagement in upstream research and peer-reviewed publications.<sup>14</sup> Based on the median AI exposure, the top five firms that are exposed to AI are *Wells Fargo, JP Morgan, Deloitte, Booz Allen, and Accenture*. Thus, I contend that firms with higher exposure will have a higher need to hire AI scientists to reduce technological uncertainty either by developing the technology or customizing existing solutions. Accordingly, to recruit AI scientists firms will have to publish more AI papers. Therefore, we are likely to observe that firms that are more exposed to AI will have a higher number of publications since the ImageNet moment compared to firms that have a lower level of exposure.

Second, firms may be less inclined to disclose information to maintain a competitive advantage in a competitive market (Kao, 2022). During such an environment, rivals are likely to be more opportunistic and firms have more to lose from their disclosure. Consequently, firms are likely to publish more papers post-ImageNet only if the benefits of publishing (e.g., recruiting key talents) outweigh the risks (e.g., expropriation) of disclosing internal knowledge. Finally, the high wage premium for AI talent (Alekseeva et al., 2021) during this competitive stage implies that recruiting talent to publish papers would less likely be a preferred strategy.

**DiD Specification:** Next, by comparing the outcomes of these two groups before and after the ImageNet shock, we can estimate the causal effect of ImageNet moment-induced labor demand on firms' AI publications. I use the following specification:

$$\text{Log}(AIPapers_{it} + 1) = \beta(\text{HighAIExposure}_i * \text{Post}_t) + \text{Firm}_i + \gamma_t + X_{it} + \epsilon_{it} \quad (1)$$

Here,  $\text{Log}(AIPapers_{it} + 1)$  is the log count of AI publications for a firm  $i$  in year  $t$ .  $\text{HighAIExposure}_i$  is a dummy variable indicating whether the firm had a higher level of AI exposure or not.  $\text{Post}_t$  is another binary variable that denotes whether the publications happened before or after the ImageNet 2012 moment. The quantity of interest is  $\beta$ , the interaction term. I also include firm and year fixed effects. Furthermore, I add other control variables as before, which are

to before.

<sup>14</sup>It is important to note that publishing scientific articles require "scientific disclosure capabilities" (Simeth and Lhuillery, 2015) and downstream firms could be penalized by firms for publishing papers (Pellens and Della Malva, 2018).

indicated by  $X_{it}$ . Here, we use the full sample size of 35920 observations from 2000 to 2019.

Table 4, Model 1 presents the main results of our difference-in-differences estimation on AI publications. After the ImageNet moment, firms in the treated group (i.e., higher AI Exposure) increased their publications by 0.134% ( $p = .000$ ) compared to the control group (i.e., firms with lower AI exposure). In Table 4 Model 2, we report the same analysis for research active firms (firms with at least one publication) and find an 0.167% ( $p = .000$ ) increase in publications after the ImageNet moment. Similarly, in Model 3 we present the result for public firms, and where we find similar results ( $\beta = 0.123$ ,  $p = .000$ ). In Models 4 to 6, we include additional controls and report similar results. In all of these models we find that relative to lower-exposed AI firms, firms with higher exposure to AI, have published more AI papers since 2012. These results are closer to our TWFE results since we included control variables in the last three models.

INSERT TABLE 4 HERE

The dynamic treatment effect is presented in Figure 2, which shows that a parallel trend persists until 2012 and the treated group increased their AI publications by 0.06% - 0.23% only after the ImageNet, compared to the control group. This incremental growth in publications aligns with the expected duration from project initiation to the eventual publication of research outcomes. We confirm the validity of the parallel trend for the two groups and find a diverging trend only after the ImageNet moment in 2012.

It is possible that firms that are more capable in AI research are likely to publish more in general. If such a case, we should observe a pre-trend before the ImageNet shock. Examining the pre-trend in Figure 2 we can rule out this possibility. This figure suggests that both groups — which are all AI firms— did publish AI papers at a similar rate. However, the difference between these two groups increased only after ImageNet 2012, suggesting the need for AI talent might have increased the number of publications.<sup>15</sup>

INSERT FIGURE 2 HERE

---

<sup>15</sup>An Instrumental variable strategy (Appendix A8) provides additional insights into this mechanism.



**Additional robustness analyses:** Following the recent literature ([Gardner, 2022](#); [Athey et al., 2021](#)), I use two different difference-in-differences methods. I use the two-stage DiD ([Gardner, 2022](#)) and matrix completion method by ([Athey et al., 2021](#)) to address potential limitations of existing frameworks. The results are presented in the Appendix in Table A6. These supplementary findings are consistent with the core argument of this study, and the results are similar in significance and estimates.

## 5 Mechanism Tests

### Do Firms Increase Publications Closely Related to Patents?

Another appropriation strategy firms can follow is the patent-paper pair, which is the simultaneous disclosure of the same knowledge in both scientific outlets and commercial patents ([Ducor, 2000](#); [Gans et al., 2017](#); [Murray, 2002](#)).<sup>16</sup> In particular, it has been well-acknowledged as a preferred disclosure strategy for firms. For instance, in industries like biotechnology and nanotechnology, patent-paper pairs' significant presence is well-documented ([Fehder et al., 2014](#)). This disclosure is a safer strategy for firms since they can attract scientists by publishing and mitigate risks of expropriation by owning the relevant intellectual property rights at the same time.

To examine this particular mechanism, I calculate the total number of patent-paper pairs by using natural language processing techniques (more details in Appendix A4). Results from our patent-paper pair analysis are presented in Table 5. The dependent variable is the log-transformed count of the total number of patent-paper pairs in that year for a specific firm. Similar to before we test based on the AI job posts with PhD requirements (Model 1,  $\beta = 0.004$ ,  $p = .532$ ), and deep learning job posts with PhD requirements (Model 3,  $\beta = -0.015$ ,  $p = .306$ ). I also use lagged variables for each of these models (Models 2, and 4 respectively). I do not find any evidence that scientific labor demand had any impact on firms' decision to increase patent-paper pairs.

---

<sup>16</sup>Patent-paper pair is possible for research projects that happen only in “Pasteur’s quadrant,” where research projects can be both commercially and scientifically important. This was named after Pasteur’s simultaneous advances in vaccination (a product with commercial value) and microbiology (knowledge with high scientific value).

The absence of increased patent-paper pairs provides additional insight into how firms balance the trade-off between recruiting scientists and disclosing internal knowledge. It is plausible that firms, under pressure from scientists, revealed internal knowledge that has limited commercial value, leading to papers that are not related to patents. This way firms are protecting commercially valuable knowledge and simultaneously recruiting scientific talents. My qualitative data further corroborate this result. In particular, my interview data suggests that managers are less likely to be concerned with the publication of ideas that did not have an immediate commercial impact.

INSERT TABLE 5 HERE

### **Bargaining Power & Additional Falsification tests**

One could argue that an increase in PhD job posts might naturally result in a rise in publications, as firms employ scientists to conduct research. However, in Table A4's (in the Appendix) Models 1 ( $\beta = -0.026$ ,  $p = .274$ ) and 2 ( $\beta = -0.009$ ,  $p = .767$ ), we examine whether AI and DL PhD job posts increase publications, respectively, for startups. However, the results suggest that knowledge leakage might be a bigger concern for startups, which is why we do not observe this relationship. This aligns with prior literature emphasizing that expropriation is a more pressing concern for startups (Katila et al., 2008; Ding, 2011). This further increases confidence in our theory that firms consider the risks of expropriation and releases knowledge when scientists have increased bargaining power.

On the other hand, one might suggest that non-AI job posts could also lead to increased non-AI publications, indicating a linear relationship between job posts and publications. It is plausible to consider research job posts, in general, to be associated with publications, and to view PhD job posts as proxies for research positions. Yet, Model 3 in Table A4 implies ( $\beta = 0.001$ ,  $p = .955$ ) that non-AI job posts, even those requiring a PhD, do not influence non-AI publications. I contend that the nonsignificant results are partly due to the lack of a discernible tight labor market in non-AI fields, as illustrated in Figure A1. This is another falsification test, demonstrating that the models are not merely detecting a “recruitment of talent for publication” trend. Instead, the relationship between AI job posts and AI publications arises from the tight labor market within the AI sector.

Overall, this analysis supports the argument that the bargaining power of AI scientists is the primary factor driving firms to publish more papers in AI.

### **General Research Reputation vs Specialized Research Reputation:**

Next, I turn my attention to whether firms are concerned about increasing “general” research reputation (publications in general) or “specialized” reputation (publications in the area where the demand is higher) during a tight labor market. The result ( $\beta = 0.001$ ,  $p = .958$ ) from Table A4 Model 4 suggests that AI job posts with PhD requirements do not increase publications in non-AI areas.<sup>17</sup> This result suggests that firms were not concerned with general research reputation, but only increased publications in AI, specifically in areas where human capital had higher bargaining power.

Additionally, I did a few more granular analyses to examine conference and journal preferences and sub-field preferences in AI publications. Table A4 (in the Appendix) Models 5 ( $\beta = 0.118$ ,  $p = .000$ ) and 6 ( $\beta = 0.051$ ,  $p = .007$ ) suggest that AI PhD job posts increase publications in both AI conferences and in AI journals. It should be noted that journal publications require a longer time to go through the peer-reviewing process. Models 7 ( $\beta = 0.296$ ,  $p = .000$ ) and 8 ( $\beta = 0.104$ ,  $p = .000$ ) suggest that deep learning job posts with PhD requirements increased deep learning publications in both journals and conferences. The effect sizes for deep learning job posts are higher than AI job posts. This is expected since the competitive labor market stems primarily from the shortage of talent in deep learning (Gofman and Jin, 2022). Overall, I document that AI job posts increase the number of AI publications, however, primarily, in the same fields of increased demand.

## **5.1 Alternative Explanations**

**Science-intensiveness of AI research:** It is plausible that innovation in AI requires more science than before. In other words, to invent in AI, it is plausible that firms need more investment and engagement in science to identify more commercial opportunities. I use the reliance on science data (Marx and Fuegi, 2020), to calculate the science intensiveness of AI Patents (see details in Appendix

---

<sup>17</sup>Additional analysis with other independent variables produces similar results: tight labor market in AI does not seem to affect publications in non-AI areas.

A7). The estimates of the intensity of science in AI patents are presented in Figure A3, which shows that the ratio of science-intensive patents has slightly decreased from almost 89% to 72% from 2000 to 2019. This suggests that over the last decade, AI patents or commercial AI have not suddenly become more dependent on science (at least based on AI patents). If the commercialization of AI were more dependent on additional research, we would have observed more AI patents that were reliant on science.<sup>18</sup>

**Science as an Input to Invention:** Recent literature highlights the importance of internal use of research in the invention as measured by citations made by firms' patents to their own publications (Arora et al., 2021). In particular, this stream of literature finds that firms tend to publish more papers if they are more likely to use (i.e., cite) them in their patents. This line of argument aligns with the absorptive capacity argument, where firms need to develop internal research capabilities to recognize and use external research in their products and services (Cohen and Levinthal, 1990). I examine the possibility that firms may have increased AI publications since they are using more internal research in their research.

I use the reliance on science (Marx and Fuegi, 2020) data's front-page non-patent literature (NPL) references. The results are presented in the Appendix in Table A3. Models 1-3 document that the primary hypothesis still holds even after controlling for the internal use of AI papers, which essentially controls for absorptive capacity in AI research. In all three models, AI job posts have a positive relationship with AI papers.

**Reverse Causality:** The relationship between job posts and publications may suggest that companies are hiring scientists to produce papers. However, this is unlikely to be the case because literature has extensively documented that firms tend to *minimize* outgoing knowledge (Alexy et al., 2013; Gans et al., 2017; Teece, 1986). More specifically, recent literature finds that firms publish fewer papers if knowledge spillover happens (Arora et al., 2021; Simeth and Raffo, 2013). Additionally, in a tight labor market, the opportunity cost of using human resources to publish a

---

<sup>18</sup>One potential reason for a slight decrease in science intensity could be that there are more AI patents now than in the past.

paper is higher than during an ordinary period, making it even less likely that companies would prioritize publishing over more profit-oriented activities. Finally, firms' tendency to publish to attract scientists is substantiated by both interview data and archival evidence, such as case studies from large firms like Apple and Amazon.

To provide empirical evidence against reverse causality, I use an instrumental variable strategy (see Appendix A5). Finally, my analysis with lagged variables also provides additional suggestive evidence against reverse causality (see Appendix A8).

**Signaling to External Market:** Firms may also publish more in AI to signal their capabilities to the stock market. Prior research posits that under certain conditions and in specific industries, the stock market rewards firms for publications (Simeth and Cincera, 2016; Pellens and Della Malva, 2018). To consider the signaling to the stock market argument, Table 3 Models 3, 6, and 9 have the key control variable: Tobin's Q. The results imply that even after controlling for market reactions, firms are publishing more to recruit AI talent.

**The Impact of Firm-University Collaboration:** Research suggests that firms collaborating with universities are also more likely to publish (Simeth and Raffo, 2013). To consider this factor, I include firm-university publication counts for both AI and non-AI. Additionally, I also focus on firm-only AI publications, where I exclude if firms had any outside collaborations (e.g., other firms or universities). Results are presented in Table A3. In these models, I introduce a new control variable: the annual count of firm-university collaborations. Models 4-6 are consistent with prior hypotheses. These results suggest that even after controlling for firms' networks with external actors, firms have increased publications in AI. However, it is interesting to note that the effect size is smaller than prior results. In other words, firms tend to publish less on their own, which is consistent with previous literature on appropriation (Alexy et al., 2013).

## 5.2 Limitations

This study uses a large dataset consisting of multiple sources complemented by rich qualitative data. However, there are still a few limitations of this study. First, I am only examining peer-reviewed published research. It is possible that this study is missing “working papers” or pre-prints that are not peer-reviewed. If that is the case, then I am *underestimating* the effect of the AI labor market shortage. In other words, the hypothesis would still hold that firms are publishing more to recruit talent.

Second, it is also possible that not all AI publications are actual AI scholarship, rather some are reclassified AI-adjacent research efforts. One way firms could do that is by using more AI keywords in their abstracts and titles. However, this still does not invalidate the key argument. More specifically, it increases confidence in the result that to attract talent firms need to change their appropriation strategies.

## 6 Discussion and Conclusion

I contribute to the innovation literature by resolving a long-standing puzzle: To attract scientists, under which conditions do research-active firms increase publications? Using rich qualitative and quantitative data from AI research, I argue and then demonstrate that when scientists’ bargaining power is formidable, firms use publications to recruit scientists. To the best of my knowledge, this is the *first empirical evidence* that links scientists and firm-level publications. Furthermore, I show that firms balance the trade-off between expropriation and recruitment by releasing less commercially valuable ideas. I also examine and rule out a number of other specific mechanisms that could motivate firms to increase publications.

**Why Disclose Research in the First Place?:** This study also clarifies several empirical irregularities in the empirical literature. For instance, this paper elucidates *why* firm-university collaborations resulted in higher innovation for firms. Prior studies ([Cockburn and Henderson, 1998](#);

Zucker et al., 2002) document that firms that co-authored scientific articles with academics tend to produce more innovation. However, these studies do not explain *why firms publish papers in the first place*. The key distinction is that firms can always work with scientists without disclosing or publishing those results. For example, prior studies document that industry sponsorship can reduce public disclosure of academic research (Blumenthal et al., 1996; Czarnitzki et al., 2015). Thus, it is not clear why firms tend to publish those results that could spill to their competitors. This study partly answers this important but hitherto unanswered question. Based on this study, it is plausible that those star scientists had enough bargaining power to nudge firms toward more publications. Consequently, the ability of firms to innovate likely stems not from co-authorship, but rather from the recruitment of top-tier scientists.

**Secrecy and Appropriation Strategy:** Whereas extant strategy literature has primarily focused on competition (Polidoro Jr and Theeke, 2012; Polidoro Jr and Toh, 2011) or factors such as the stock market (Pellens and Della Malva, 2018; Simeth and Cincera, 2016) to explain a firm's decision to publish scientific papers, this study highlights an underexplored factor: scientists' role in the appropriation strategy. Thus, this study underscores the importance of tension: a firm's disclosure strategy aims to balance the trade-off between the benefits of acquiring talent versus the risks associated with the disclosure of underlying knowledge of an invention. Specifically, in knowledge-intensive industries, where firms rely on human capital to create and capture value from innovation (Kang and Lee, 2022), it is important to examine how firms balance this tension. Future research could build on this and examine the specific strategies (e.g., delaying publication) firms take to minimize knowledge spillover. Furthermore, prior research on corporate science relied on an implicit assumption that published papers are primarily determined by firms' engagement in basic science. My result suggests that appropriation strategies affect the number of firm-level publications. This result is consistent with the recent research (Bhaskarabhatla and Hegde, 2014) which also demonstrated that management practices can incentivize certain kinds of disclosure (e.g., patents) over others (e.g., publications). Thus, this study opens the door for potential alternative explanations that lead to industry's declining presence in basic research in recent decades.

**STEM Labor Market and the Secular Decline of Corporate Science:** This work complements previous work (Arora et al., 2018; Mowery, 2009) on the decline of corporate science which prior studies documented. Based on this study, it is plausible that scientists lack bargaining power in most research areas due to the increased availability of trained scientists. Thus, the need for firm-level publications is less salient in most fields to attract talent. Instead, in many non-AI fields, talent is already competing to gain employment at such firms. This lack of bargaining power is evident in most fields since wages have not grown significantly in those fields (Cyranoski et al., 2011). Contrary to popular belief that there is a STEM talent shortage, recent evidence from the U.S. suggests that rather than a shortage in certain areas such as in life sciences, there is even an *oversupply* of PhD graduates (Cyranoski et al., 2011; Xue and Larson, 2015). These observations are consistent with the trend in publications in pharmaceutical and biotechnology firms, which have decreased publications over recent years (Arora et al., 2018). This lends credence to my theory that the scientific labor market plays an important role in firm-level publications.

A potential follow-up study could focus on the implications of firm-level increased publications on worker mobility and its subsequent effects on expropriation (Starr et al., 2018a). As the authors of research publications may venture out to establish their own startups or join rival firms, it is crucial to examine the strategic implications of such talent movements. Ultimately, a deeper understanding of the interplay between human capital and firm-level publications can contribute to our understanding of maintaining competitive advantage.

**Patent-Paper Pairs as Appropriation Strategy:** This study contributes to our understanding of patent-paper pairs as an appropriation strategy (Ducor, 2000; Murray, 2002). This is an important but overlooked topic in the literature, perhaps due to a lack of data. The results from this study suggest that it is plausible that strategically disclosing an idea in both patents and publications requires a lot of effort and coordination (Gans et al., 2017; Simeth and Lhuillery, 2015). Therefore, the number of patent-paper pairs might be less common than previously assumed. An interesting future research avenue would be to investigate under which conditions firms release their internal research with patent-paper pairs. Studies could also examine which processes and capabilities are



needed to pursue this appropriation strategy.

## References

- Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2020). AI and Jobs: Evidence From Online Vacancies. Technical report, National Bureau of Economic Research.
- Agarwal, R., Ganco, M., and Ziedonis, R. H. (2009). Reputations for toughness in patent enforcement: Implications for knowledge spillovers via inventor mobility. *Strategic Management Journal*, 30(13):1349–1374.
- Agarwal, R. and Ohyama, A. (2013). Industry or Academia, Basic or Applied? career Choices and Earnings Trajectories of Scientists. *Management Science*, 59(4):950–970.
- Ahmed, N., Wahed, M., and Thompson, N. C. (2023). The growing influence of industry in ai research. *Science*, 379(6635):884–886.
- Alba, D. (2017). A Year After Pledging Openness, Apple Still Falls Behind on AI. *BuzzFeed News*.
- Alekseeva, L., Azar, J., Gine, M., Samila, S., and Taska, B. (2021). The Demand for AI Skills in the Labor Market. *Labour economics*, 71:102002.
- Alexy, O., George, G., and Salter, A. J. (2013). Cui Bono? the Selective Revealing of Knowledge and Its Implications for Innovative Activity. *Academy of management review*, 38(2):270–291.
- Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Van Esesn, B. C., Awwal, A. A. S., and Asari, V. K. (2018). The History Began from Alexnet: A Comprehensive Survey on Deep Learning Approaches. *arXiv preprint arXiv:1803.01164*.
- Arora, A., Belenzon, S., and Pataconi, A. (2018). The Decline of Science in Corporate R&D. *Strategic Management Journal*, 39(1):3–32.
- Arora, A., Belenzon, S., and Sheer, L. (2021). Knowledge Spillovers and Corporate Investment in Scientific Research. *American Economic Review*, 111(3):871–98.
- Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*.
- Bengio, Y. (2020). Time to Rethink the Publication Process in Machine Learning. *Retrieved January, 21:2021*.
- Bhaskarabhatla, A. and Hegde, D. (2014). An Organizational Perspective on Patenting and Open Innovation. *Organization Science*, 25(6):1744–1763.
- Blumenthal, D., Causino, N., Campbell, E., and Louis, K. S. (1996). Relationships Between Academic Institutions and Industry in the Life Sciences—An Industry Survey. *New England Journal of Medicine*, 334(6):368–374.
- Brown, J. A., Gianiodis, P. T., and Santoro, M. D. (2015). Following Doctors’ Orders: Organizational Change as a Response to Human Capital Bargaining Power. *Organization Science*.
- Charette, R. N. (2013). The STEM Crisis Is a Myth. *IEEE Spectrum*, 50(9):44–59.

- Clark, J. (2015). Apple's secrecy hurts its AI software development. *Bloomberg.com*.
- Cockburn, I. M. and Henderson, R. M. (1998). Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery. *The Journal of Industrial Economics*.
- Coff, R. W. (1999). When Competitive Advantage Doesn't Lead to Performance: The Resource-based View and Stakeholder Bargaining Power. *Organization Science*, 10(2):119–133.
- Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, pages 128–152.
- Cohen, W. M., Nelson, R., and Walsh, J. P. (2000). Protecting Their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (or not).
- Cyranoski, D., Gilbert, N., Ledford, H., Nayar, A., and Yahia, M. (2011). The PhD Factory: The World is Producing More PhDs Than Ever Before. Is It Time to Stop. *Nature*, 472(7343):276–279.
- Czarnitzki, D., Grimpe, C., and Toole, A. A. (2015). Delay and Secrecy: Does Industry Sponsorship Jeopardize Disclosure of Academic Research? *Industrial and Corporate change*, 24(1):251–279.
- Ding, W. W. (2011). The Impact of Founders' Professional-education Background on the Adoption of Open Science by for-profit Biotechnology Firms. *Management Science*, 57(2):257–273.
- Ducor, P. (2000). Coauthorship and Coinventorship. *Science*, 289(5481):873–875.
- Dwoskin, E. (2017). Why Apple is Struggling to Become an Artificial Intelligence Powerhouse. *The Washington Post*.
- Fedyk, A., He, A., and Hodson, J. (2023). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*.
- Fehder, D. C., Murray, F., and Stern, S. (2014). Intellectual Property Rights and the Evolution of Scientific Journals as Knowledge Platforms. *International Journal of Industrial Organization*, 36:83–94.
- Felten, E., Raj, M., and Seamans, R. (2021). Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses. *Strategic Management Journal*.
- Gans, J. S., Murray, F. E., and Stern, S. (2017). Contracting Over the Disclosure of Scientific Knowledge: Intellectual Property and Academic Publication. *Research Policy*, 46(4):820–835.
- Gardner, J. (2022). Two-stage differences in differences. *arXiv preprint arXiv:2207.05943*.
- Gershgorn, D. (2018). The Inside Story of How AI Got Good Enough to Dominate Silicon Valley. *Quartz*. Retrieved, 5.
- Gittelman, M. and Kogut, B. (2003). Does Good Science Lead to Valuable Knowledge? biotechnology Firms and The Evolutionary Logic of Citation Patterns. *Management Science*.
- Gofman, M. and Jin, Z. (2022). Artificial Intelligence, Education, and Entrepreneurship. *Journal of Finance, Forthcoming*.

- Hall, B., Helmers, C., Rogers, M., and Sena, V. (2014). The Choice between Formal and Informal Intellectual Property: A Review. *Journal of Economic Literature*, 52(2):375–423.
- Hartmann, P. and Henkel, J. (2020). The Rise of Corporate Science in AI: Data as a Strategic Resource. *Academy of Management Discoveries*, 6(3):359–381.
- Hicks, D. (1995). Published Papers, Tacit Competencies and Corporate Management of the Public/Private Character of Knowledge. *Industrial and Corporate Change*, 4(2):401–424.
- Kang, H. and Lee, W. (2022). How Innovating Firms Manage Knowledge leakage: A Natural Experiment on the Threat of Worker Departure. *Strategic Management Journal*, 43(10):1961–1982.
- Kao, J. L. (2022). Information disclosure in the presence of competition: Evidence from the pharmaceutical industry. Available at SSRN 4081398.
- Katila, R., Rosenberger, J. D., and Eisenhardt, K. M. (2008). Swimming with sharks: Technology ventures, defense mechanisms and corporate relationships. *Administrative science quarterly*, 53(2):295–332.
- Kehoe, R. R. and Tzabbar, D. (2015). Lighting the Way or Stealing the Shine? an Examination of the Duality in Star Scientists' Effects on Firm Innovative Performance. *Strategic Management Journal*, 36(5):709–727.
- Kim, B., Kim, E., Miller, D. J., and Mahoney, J. T. (2016). The Impact of the Timing of Patents on Innovation Performance. *Research Policy*, 45(4):914–928.
- Kim, J. and Mahoney, J. T. (2007). Appropriating Economic Rents From Resources: An Integrative Property Rights and Resource-based Approach. *International Journal of Learning and Intellectual Capital*, 4(1-2):11–28.
- Kleid, D. (2002). Scientist and Patent Agent at Genentech. Oral History Conducted in 2001 and 2002 by Sally Smith Hughes. Regional Oral History Office, Bancroft Library. *University of California, Berkeley*, 7.
- Latour, B. and Woolgar, S. (1979). *Laboratory Life*. Beverly Hills. Calif.: Sage.
- Levy, S. (2018). Inside Amazon's Artificial Intelligence Flywheel. *Wired*, February, 1.
- Liu, L., Wang, Y., and Xu, Y. (2022). A Practical Guide to Counterfactual Estimators for Causal Inference with Time-series Cross-sectional Data. *American Journal of Political Science*.
- Marx, M. and Fuegi, A. (2020). Reliance on Science: Worldwide Front-page Patent Citations to Scientific Articles. *Strategic Management Journal*, 41(9):1572–1594.
- Merton, R. K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago press.
- Metz, C. (2016). Artificial Intelligence Just Broke Steve Jobs' Wall of Secrecy. *Wired*, 6(December).

- Mickle, T. (2017). Secretive Apple Tries to Open up on Artificial Intelligence. *The Wall Street Journal*.
- Molloy, J. C. and Barney, J. B. (2015). Who Captures the Value Created with Human Capital? a Market-based View. *Academy of Management Perspectives*, 29(3):309–325.
- Mowery, D. C. (2009). Plus ca Change: Industrial R&D in the “Third Industrial Revolution”. *Industrial and Corporate Change*, 18(1):1–50.
- Murray, F. (2002). Innovation As Co-evolution of Scientific and Technological Networks: Exploring Tissue Engineering. *Research policy*, 31(8-9):1389–1403.
- Pacheco-de Almeida, G. and Zemsky, P. B. (2012). Some Like It Free: Innovators’ Strategic Use of Disclosure to Slow Down Competition. *Strategic Management Journal*, 33(7):773–793.
- Parloff, R. (2016). Why Deep Learning is Suddenly Changing Your Life. *Fortune*. New York: Time Inc.
- Partha, D. and David, P. A. (1994). Toward a New Economics of Science. *Research policy*, 23(5):487–521.
- Pellens, M. and Della Malva, A. (2018). Corporate Science, Firm Value, and Vertical Specialization: Evidence From the Semiconductor Industry. *Industrial and Corporate Change*, 27(3):489–505.
- Polidoro Jr, F. and Theeke, M. (2012). Getting Competition Down to a Science: The Effects of Technological Competition on Firms’ Scientific Publications. *Organization Science*.
- Polidoro Jr, F. and Toh, P. K. (2011). Letting Rivals Come Close or Warding Them Off? the Effects of Substitution Threat on Imitation Deterrence. *Academy of Management Journal*, 54(2):369–392.
- Richard, C. L., Alvin, K. K., Richard, R. N., Sidney, G. W., Richard, G., and Zvi, G. (1987). Appropriating the Returns from Industrial Research and Development. *Brookings Papers on Economic Activity*, 1987(3):783.
- Roach, M. and Sauermann, H. (2010). A Taste for Science? PhD Scientists’ Academic Orientation and Self-selection Into Research Careers in Industry. *Research policy*, 39(3):422–434.
- Rosenberg, N. (1990). Why do firms do basic research (with their own money)? *Research Policy*, pages 165–174.
- Rotolo, D., Camerani, R., Grassano, N., and Martin, B. R. (2022). Why Do Firms Publish? a Systematic Literature Review and a Conceptual Framework. *Research Policy*, 51(10):104606.
- Rubin, S. (2011). Do NOT Publish that Article (if you Care About Patents Right in the United States. *IEEE-USA Today’s Engineer*, 291671v1.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252.

- Sauermann, H. and Roach, M. (2014). Not All Scientists Pay to be Scientists: PhDs' Preferences for Publishing in Industrial Employment. *Research Policy*, 43(1):32–47.
- Simeth, M. and Cincera, M. (2016). Corporate Science, Innovation, and Firm Value. *Management Science*, 62(7):1970–1981.
- Simeth, M. and Lhuillery, S. (2015). How Do Firms Develop Capabilities for Scientific Disclosure? *Research policy*, 44(7):1283–1295.
- Simeth, M. and Raffo, J. D. (2013). What Makes Companies Pursue an Open Science Strategy? *Research Policy*, 42(9):1531–1543.
- Somaya, D. (2012). Patent Strategy and Management: An Integrative Review and Research Agenda. *Journal of Management*, 38(4):1084–1114.
- Starr, E., Ganco, M., and Campbell, B. A. (2018a). Strategic human capital management in the context of cross-industry and within-industry mobility frictions. *Strategic Management Journal*, 39(8):2226–2254.
- Starr, E., Prescott, J., and Bishara, N. (2018b). Noncompetes in the us labor force (university of michigan law & economic research paper 18-013). *SSRN*.
- Statt, N. (2016). Apple Will Break With Tradition and Start Publishing AI Research. *The Verge*.
- Stern, S. (2004). Do Scientists Pay to be Scientists? *Management science*, 50(6):835–853.
- Teece, D. J. (1986). Profiting From Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy. *Research policy*, 15(6):285–305.
- Teece, D. J. (2018). Profiting from innovation in the digital economy: Enabling technologies, standards, and licensing models in the wireless world. *Research policy*, 47(8):1367–1387.
- The Economist (2016). From Not Working to Neural Networking. *The Economist*.
- Tilley, A. (2016). Apple Publishes Its First Artificial Intelligence Paper. *Forbes*.
- Toh, P. K. and Miller, C. D. (2017). Pawn to Save a Chariot, or Drawbridge into the Fort? Firms' Disclosure During Standard Setting and Complementary Technologies within Ecosystems. *Strategic Management Journal*, 38(11):2213–2236.
- Xue, Y. and Larson, R. C. (2015). STEM Crisis or STEM Surplus? Yes and Yes. *Monthly labor review*, 2015.
- Zucker, L. G., Darby, M. R., and Armstrong, J. S. (2002). Commercializing knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology. *Management Science*.

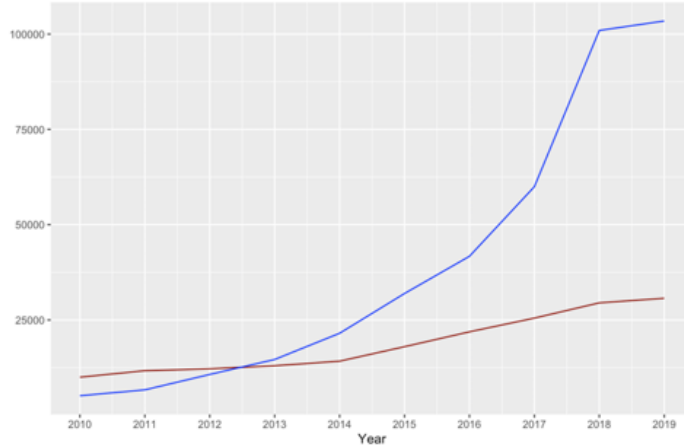


Figure 1: The Supply and Demand of AI Talent

Note: This Figure depicts AI job posts with PhD requirements (blue line) and the total number of computer science PhD graduates per year (red line). Please note that the supply side includes *all* computer science PhDs from the US. In other words, some graduates may not have the necessary training for AI jobs. Therefore, this conservative comparison demonstrates the AI industry’s voracious appetite for AI talents. Job post data is from Burning Glass Technologies and supply-side data is from Computing Research Association (CRA).

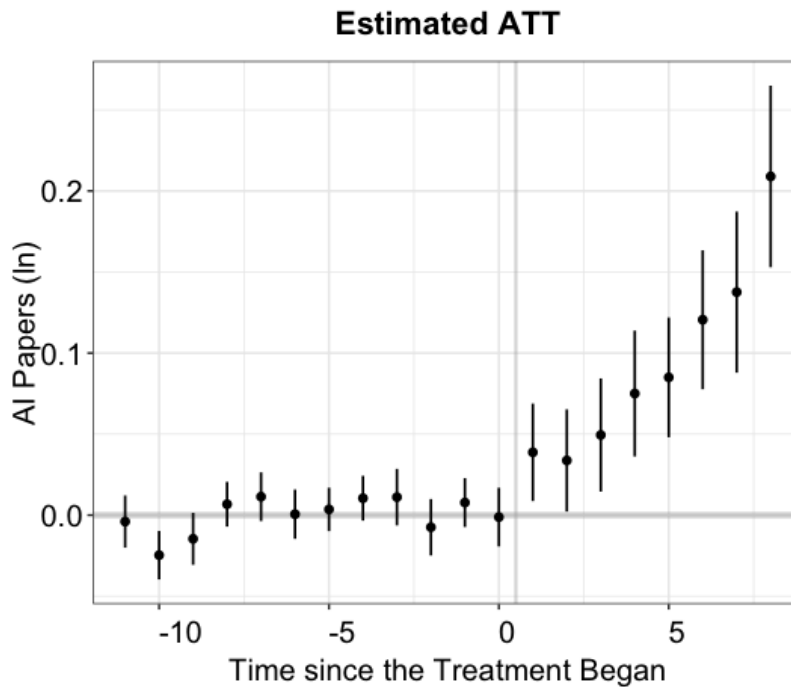


Figure 2: The dynamic treatment effects estimates from the difference-in-differences from the ImageNet moment. Firms with higher AI exposure were classified as treated, and firms with lower AI exposure were classified as control units. 2012 is the intervention year because of the ImageNet moment. Vertical lines represent a 95% confidence interval. Standard errors are clustered at the firm level.

Table 1: Summary of Select Quotes

<b>Publications as an instrument to recruit talent</b>
<p>“[Salary] is important, particularly when there is a competitive situation with Microsoft, DeepMind, Google etc. But the other fundamentals have to be right. If they’re not right, people are just not even considering coming to work for you.” – Yann LeCun, Facebook Chief AI scientist.</p>
<p>“Ability to attract talent is almost existential for AI companies. They would do almost anything to hire talented scientists. Starting a new research center, publishing, higher salary. . . . anything” – Interview #16 (industry senior research scientist)</p>
<p>“I was recruited to lead the research lab only last year [2017]. Before that this company had no research presence. Now the company encourages us to publish in top AI conferences. In many ways, this has been a huge change for us. [. . . ] People won’t come to work for you if you don’t publish.” – Interview #21 (industry manager)</p>
<p>“[Apple and other companies]. . . are now allowing publishing to please the AI scientists. They are trying to deal with the scarcity of [AI] scientists.” – Interview #10 (industry scientist)</p>
<p>“I have been struggling to recruit talent because our team is not allowed to publish. Secrecy is more important for us [our division]. This makes things difficult for us. On the other hand, other teams in the same company can recruit from my alma mater because [graduate] students can work and publish easily. I am a little frustrated with that.” - Interview #22 (industry research manager)</p>



Table 2: Summary Statistics

Statistics	Mean	St. Dev.	Min	Max
AI Papers (ln)	0.324	0.881	0	7.026
Non-AI Papers (ln)	0.737	1.562	0	8.298
AI Job Posts (ln)	0.585	1.382	0	9.776
AI PhD job posts (ln)	0.154	0.607	0	7.238
DL PhD job posts (ln)	0.058	0.356	0	6.349
HigherSalary AI job posts (ln)	0.055	0.347	0	6.052
Non-AI Patent (ln)	1.309	1.989	0	9.128
AI Patent (ln)	0.182	0.481	0	6.265
Patent-Paper Pair (ln)	0.069	0.343	0	4.419
Firm-University Collaboration (ln)	0.458	1.098	0	7.180
Tobins' Q	2.484	2.038	0.654	24.0
Sales (ln)	8.486	2.379	0	13.164
R&D Spending (ln)	5.723	2.0	0	10.489
AI Pub Trends (ln)	5.009	0.13	4.81	5.245

Table 3: Estimates from the Fixed Effects Model

Sample Model	AI Papers (ln)								
	All (1)	Research-firms (2)	Public (3)	All (4)	Research-firms (5)	Public (6)	All (7)	Research-firms (8)	Public (9)
AI PhD Job posts	0.121 (0.016) [0.000]	0.081 (0.019) [0.000]	0.105 (0.021) [0.000]						
DL PhD Job posts				0.150 (0.021) [0.000]	0.084 (0.024) [0.001]	0.115 (0.024) [0.000]			
HigherSalary AI Job posts							0.133 (0.017) [0.000]	0.086 (0.025) [0.001]	0.107 (0.018) [0.000]
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Financial Controls	No	No	Yes	No	No	Yes	No	Yes	Yes
R <sup>2</sup>	0.279	0.301	0.319	0.279	0.300	0.317	0.278	0.332	0.310
Observations	18280	5330	5309	18280	5330	5309	18280	5330	5309

Note: Standard errors are clustered at the firm-level, which are shown in parentheses. p-values are shown in brackets. Controls: AI patents, non-AI patents, non-AI publications, AI pub Trends, AI PhD Supply. Financial controls: Tobin's Q, R&D Spending, Sales.

Table 4: Estimates from Difference-in-Differences

Sample Model	AI Papers (ln)					
	All (1)	Research-firms (2)	Public (3)	All (4)	Research-firms (5)	Public (6)
HighAIExposure*Post	0.134 (0.008) [0.000]	0.167 (0.027) [0.000]	0.123 (0.021) [0.000]	0.064 (0.014) [0.000]	0.060 (0.021) [0.005]	0.041 (0.017) [0.017]
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	Yes	Yes	Yes
Financial Controls	No	No	No	No	No	Yes
R <sup>2</sup>	0.008	0.004	0.003	0.388	0.387	0.392
Observations	35920	10660	11770	35920	10660	10608

Note: Data: 2000-2019; standard errors are clustered at the firm-level, which are shown in parentheses. p-values are shown in brackets. Controls: AI patents, non-AI patents, non-AI publications, AI pub Trends, AI PhD Supply. Financial controls: Tobin's Q, R&D Spending, Sales.

Table 5: Scientific Labor Demand and Patent-Paper Pairs

Model	Patent Paper Pairs (ln)			
	(1)	(2)	(3)	(4)
AI PhD Job posts	0.004 (0.006) [0.532]			
AI PhD Job posts (t-1)		0.003 (0.008) [0.708]		
DL Job posts			-0.015 (0.014) [0.306]	
DL Job posts (t-1)				-0.014 (0.016) [0.396]
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.010	0.010	0.012	0.011
Observations	18280	18216	18280	18216

Note: Standard errors are clustered at the firm-level, which are shown in parentheses. p-values are shown in brackets. Controls: AI patents, non-AI patents, non-AI publications, AI pub Trends, AI PhD Supply.

## Appendix A1

Table A1: Summary of the Interviews Conducted and Respondent Characteristics

<b>Respondent type</b>	<b>Number of interviews</b>	<b>Mean years of research experience</b>	<b>Mean interview length (in minutes)</b>
Senior graduate students	3	4.67	75
Tenure track professors	6	7.5	60
Tenured professors	5	12.6	51
Industry Scientists	7	7.1	44
Industry research managers	3	14	28

### List of Keywords Used to Classify AI papers

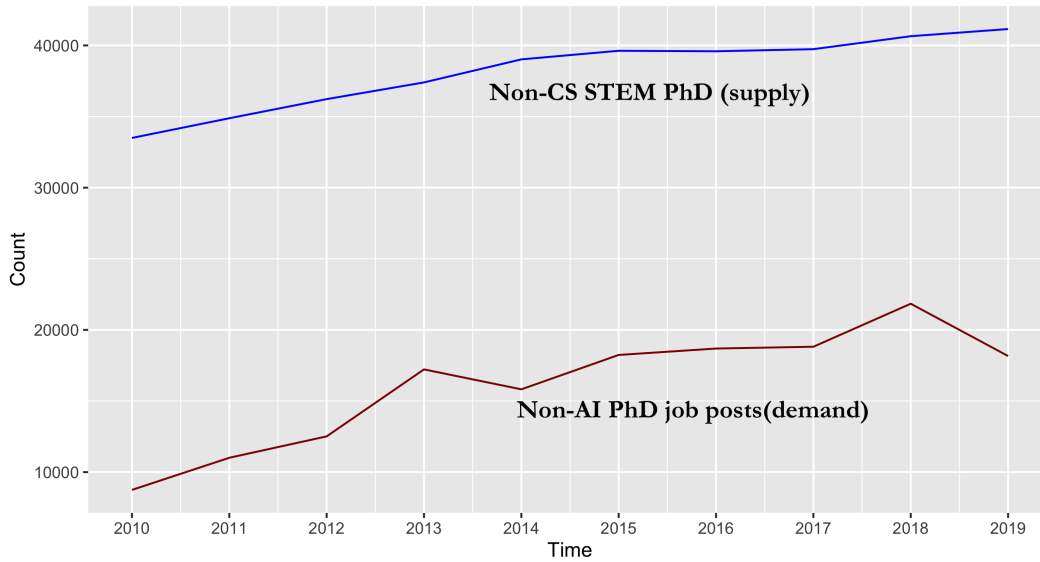
‘Artificial Intelligence’, ‘Experts System’, ‘Automatic Speech Recognition’, ‘Caffe Deep Learning Framework’, ‘Chatbot’, ‘Computational Linguistics’, ‘Computer Vision’, ‘Data Mining’, ‘Decision Trees’, ‘Deep Learning’, ‘Deeplearning4j’, ‘Distinguo’, ‘Deep Representation Learning’, ‘Google Cloud Machine Learning’, ‘Convolutional neural networks’, ‘Pattern Recognition’, ‘Support Vector Machine’, ‘Robotics’, ‘Knowledge Discovery and Data Mining’, ‘Advances in Neural Information Processing Systems’, ‘Gradient boosting’, ‘H2O (software)’, ‘IBM Watson’, ‘Image Processing’, ‘Image Recognition’, ‘ImageNet’, ‘Resnet’, ‘Keras’, ‘Knowledge based systems’, ‘Latent Dirichlet Allocation’, ‘Latent Semantic Analysis’, ‘Lexalytics’, ‘Lexical Acquisition’, ‘Lexical Semantics’, ‘Libsvm’, ‘LSTM’, ‘Machine Learning’, ‘Machine Translation’, ‘Machine Vision’, ‘Madlib’, ‘Mahout’, ‘Microsoft Cognitive Toolkit’, ‘MLPACK’, ‘Mlpy’, ‘Modular Audio Recognition Framework’, ‘MXNet’, ‘Natural Language Processing’, ‘Natural Language Toolkit (NLTK)’, ‘natural language understanding’, ‘ND4J (software)’, ‘Natural Language Learning’, ‘Nearest Neighbor Algorithm’, ‘Data clustering’, ‘Neural Networks’, ‘Object Recognition’, ‘Object Tracking’, ‘OpenCV’, ‘OpenNLP’, ‘Pattern Recognition’, ‘Pybrain’, ‘Random Forests’, ‘Recommender Systems’, ‘Language Model’, ‘Semantic Driven Subtractive Clustering Method’, ‘Semi-Supervised Learning’, ‘Sentiment Analysis’, ‘Opinion Mining’, ‘Sentiment Classification’, ‘Speech Recognition’, ‘Supervised Learning’, ‘Support Vector Machines’, ‘TensorFlow’, ‘Text Mining’, ‘Text to Speech’, ‘Tokenization’, ‘Topic model’,

‘Unsupervised Learning’, ‘Virtual Agents’, ‘Vowpal’, ‘Wabbit’, ‘Word2Vec’, ‘Word Embedding’, ‘Xgboost’, ‘AI ChatBot’, ‘Conversational agent’, ‘Robotic’, ‘Learning Representations’, ‘Boltzmann Machine’, ‘Apertium’, ‘Hidden Markov Model’, ‘sequence model’, ‘Supervised Learning’, ‘generative adversarial network’, ‘Reinforcement Learning’

## **List of Keywords Used to Classify AI Job Posts**

This list of keywords is based on ([Alekseeva et al., 2021](#)) with a few minor additions:

“AI ChatBot”, “AI KIBIT”, “ANTLR”, “Apertium”, “Artificial Intelligence”, “Automatic Speech Recognition (ASR)”, “Caffe Deep Learning Framework”, “Chatbot”, “Computational Linguistics”, “Computer Vision”, “Decision Trees”, “Deep Learning”, “Deeplearning4j”, “Distinguo”, “Google Cloud Machine Learning Platform”, “Gradient boosting”, “H2O (software)”, “IBM Watson”, “Image Processing”, “Image Recognition”, “IPSoft Amelia”, “Ithink”, “Keras”, “Latent Dirichlet Allocation”, “Latent Semantic Analysis”, “Lexalytics”, “Lexical Acquisition”, “Lexical Semantics”, “Libsvm”, “Machine Learning”, “Machine Translation (MT)”, “Machine Vision”, “Madlib”, “Mahout”, “Microsoft Cognitive Toolkit”, “MLPACK”, “Mlpy”, “Modular Audio Recognition Framework (MARF)”, “Moses”, “MXNet”, “Natural Language Processing”, “Natural Language Toolkit (NLTK)”, “ND4J (software)”, “Nearest Neighbor Algorithm”, “Neural Networks”, “Object Recognition”, “Object Tracking”, “OpenCV”, “OpenNLP”, “Pattern Recognition”, “Pybrain”, “Random Forests”, “Recommender Systems”, “Semantic Driven Subtractive Clustering Method (SDSCM)”, “Semi-Supervised Learning”, “Sentiment Analysis / Opinion Mining”, “Sentiment Classification”, “Speech Recognition”, “Supervised Learning (Machine Learning)”, “Support Vector Machines (SVM)”, “TensorFlow”, “Text Mining”, “Text to Speech (TTS)”, “Tokenization”, “Torch (Machine Learning)”, “Unsupervised Learning”, “Virtual Agents”, “Vowpal”, “Wabbit”, “Word2Vec”, “Xgboost”, “Weka”, “Konstanz Information Miner (KNIME)”, “Scikit-learn”, “Theano”, “Data Science”



**Figure A1:** Demand (red line) and Supply (blue line) for non-AI PhDs. This graph demonstrates that for non-AI fields, there is no discernible labor shortage and hence no visible bargaining power for average scientists. The demand data is from Burning Glass Technologies job posts data. If job posts had a PhD requirement and were not classified as AI, they were counted under this. To calculate supply, I took NSF STEM PhD data and subtracted annual American Computer Science PhDs, which was obtained from the CRA.

## Appendix A2: Firm-level AI Exposure

Following (Acemoglu et al., 2020), I calculate the AI exposure for each year as the sum of all job posts weighted by their corresponding occupational AI exposure (Felten et al., 2021). This value is then normalized by the AI industry's total number of AI job posts for each year. Since firms post different job posts in different years, that would change their total AI exposure score each year. A higher total AI exposure would indicate that the firm is more exposed to AI. For firm  $i$  in year  $t$  for a job post with the occupation  $j$ , this is how I calculate this measure:

$$\text{AI exposure}_{it} = \frac{\sum(\text{AIJobposts}_{ijt} \times \text{AIOE}_j)}{\sum(\text{AIJobposts}_t)}$$

One might assume that firms with higher AI exposure would publish more papers. Interestingly, based on the mean AI exposure, the top five firms that are exposed to AI are *Wells Fargo, JP Morgan, Deloitte, Booz Allen, and Accenture*. On the other hand, firms with the lowest average AI exposure are *Lowe's, Koch Industries, GE, Walmart, and Mcdonald's*. This observation implies that AI exposure does not necessarily lead to increased engagement in upstream research and subsequent peer-reviewed publications. Qualitative data suggests that due to technological uncertainty around deep learning, many of these firms needed internal capabilities either to develop or to customize this technology. I contend that firms with higher exposure will need to hire AI scientists to reduce technological uncertainty either by developing the technology or customizing existing solutions. Accordingly, firms will have to publish more AI papers to recruit AI scientists. Therefore, we are likely to observe that firms that are more exposed to AI will have a higher number of publications since the ImageNet shock compared to firms with a lower exposure level.

## Appendix A3: Qualitative evidence for ImageNet Moment

Here additional supporting evidence showcases that ImageNet 2012 was surprising to both AI scientists and outsiders:

The organizer of the ImageNet contest termed deep learning’s success an “unexpected outcome”.<sup>19</sup>

Other observers have used the term “ImageNet moment” to describe breakthroughs in different fields of AI such as AI’s use in voice recognition or Biology. “Why is the ”ImageNet moment” in the voice field not coming late?”.<sup>20</sup> For Biology see: “DeepMind’s AlphaFold & the Protein Folding Problem”.<sup>21</sup> Similarly, the unexpected popularity of ChatGPT was also mentioned as an ”ImageNet moment”.<sup>22</sup>

The reaction to ImageNet’s success surprised even the winner of that competition “[Alex] Krizhevsky, [...], chuckles when recalling the weeks after the 2012 ImageNet results came out. “It became kind of surreal,” he says. “We started getting acquisition offers very quickly. Lots of emails” (Gershgorn, 2018).

My qualitative data also confirmed that ImageNet 2012 was the pivotal moment that convinced computer scientists that deep learning could be a major way to make progress in AI research.

## Appendix A4: Patent-Paper Pair

**Patent-Paper Pair Count (In):** To create this variable at first, I matched a firm’s publications’ author names with their own patent holders’ names. I only match papers with patents in the same year or within 1 year of the patent’s application year, since firms need to file a patent within 12 months of publishing a paper. To match names, I use fuzzy string matching. More specifically, I use the Levenshtein Distance to calculate the differences between sequences and patterns of the names. I use 80% overlap in the author names. This results in 52423 potential paper-patent pairs. After that, I calculate cosine similarity for both abstracts and titles of those potential paper-patent pairs. To create the final paper-patent pair, I use the similarity threshold 50% for abstracts and 30% of titles.<sup>23</sup>

---

<sup>19</sup>Fei Fei Li’s presentation on ImageNet: <https://tinyurl.com/imagenet12> (Page: 55)

<sup>20</sup><https://www.programmersought.com/>

<sup>21</sup><https://www.kdnuggets.com/>

<sup>22</sup>Benedict Evans, venture capitalist

<sup>23</sup>I have used different variations as well mostly focusing on conservative estimates (higher thresholds for the similarity score), but the results are still similar. Manual validation by two research assistants suggests that when abstract similarity

I also look at higher matches in the title. If there is a 100% similarity in the titles, then I label it as a Paper-Patent Pair. If there is a 75% similarity in abstracts, I label it as a pair (the process is described in Figure A2). Similarly, if the titles are 30% or more similar and the abstracts are 50% or more similar, I consider that as a Patent-Paper Pair. In total, I find 11413 Paper-Patent Pairs. After that, I count the total number of patent-paper pairs at the firm level for each year and log transform it to consider the skewness of the data.

---

is lower than 50%, the chances of being a patent-paper pair go down significantly.



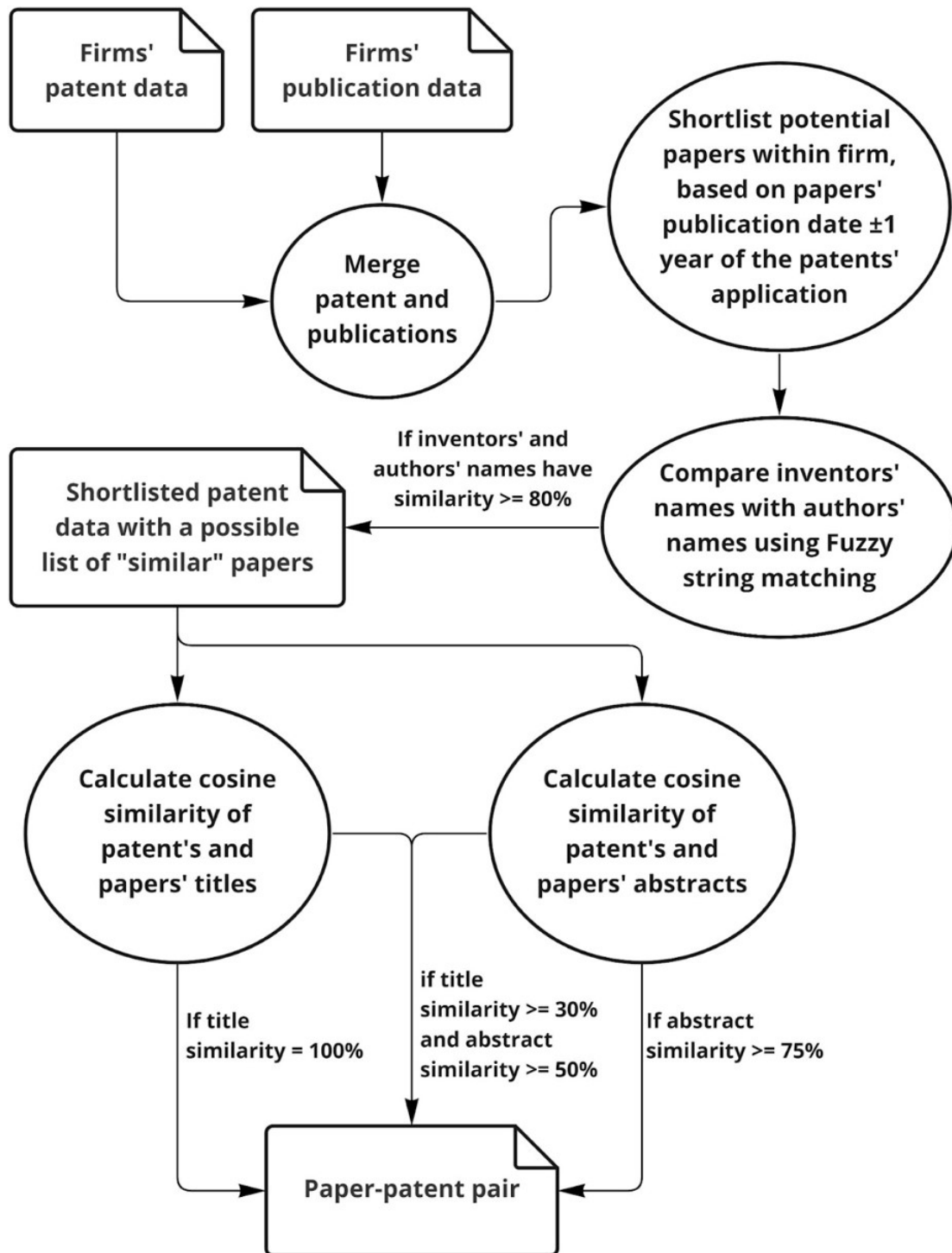


Figure A2: Patent-paper pair process flowchart

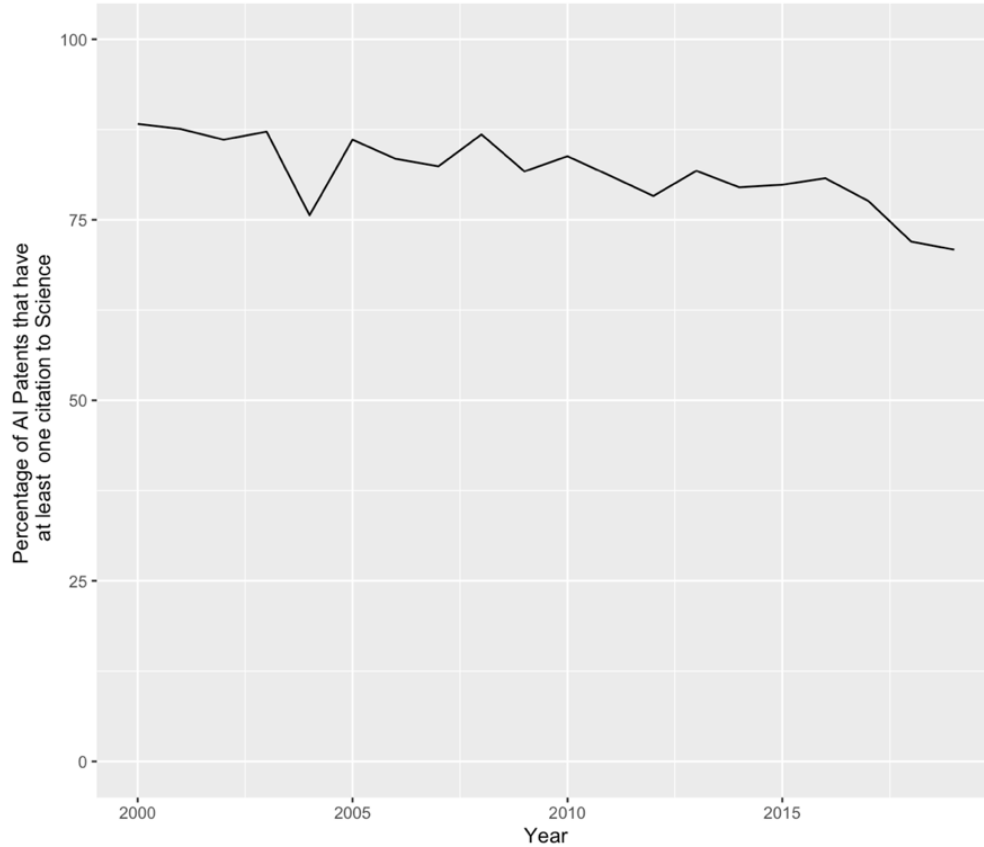


Figure A3: This figure presents the trends in the AI patents that cite science (either industry or academia), measured as the ratio between the total number of AI patents that have at least one citation to science divided by the total number of AI patents in each year.

## Appendix A5: Instrumental Variable Strategy

To address further issues around potential omitted variable bias and reverse causality, I use an instrumental variable strategy to demonstrate that it is the demand for human capital in AI that led to increased publications. My IV strategy is to use the variation in the exposure to AI as a source of exogenous variation in labor demand (i.e., job posts). Firm-level AI exposure directly affects the AI labor demand but not AI publications. Therefore, they offer a source of variation in publications that is independent of the potential confounding factors. Consequently, firm-level AI exposure affects AI publications only via the increased labor demand.

The core idea behind this instrument is that a firm that is more exposed to AI technologies will

have more demand for AI technologies. In contrast, it is less likely to be the case that a firm’s AI exposure will directly affect the firm’s AI publications. However, AI exposure could affect a firm’s publications by increasing its AI labor demand (e.g., job posts).

I find that firms’ AI exposure created more AI labor demands with the PhD requirement (Table A2 Model 1, with an F-stat = 18.28). These first-stage results are consistent with (Acemoglu et al., 2020), who used the same measurement to find that AI job posts and AI exposure are positively correlated. In the second stage, I find that the predicted number of AI PhD job posts increases AI publications (Model 2). Similarly, the results for deep learning job posts are reproduced in Models 3 and 4 (F-stat = 12.83).

Table A2: Estimates from the Instrumental Variables Strategy

Model	AI Papers (ln)			
	(1)	(2)	(3)	(4)
	First Stage (AI PhD Job posts)	Second Stage	First Stage (DL PhD Job posts)	Second Stage
AI PhD Job posts		0.47 (0.20)		
DL PhD Job posts				0.59 (0.22)
AI Exposure	0.095 (0.021)		0.076 (0.025)	
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.008	0.010	0.010	0.014
Observations	5309	5309	5309	5309
F-stat	18.28		12.83	

Note: Standard errors are clustered at the firm-level, which are shown in parentheses. Controls: AI patents, non-AI patents, non-AI publications, AI pub Trends, AI PhD Supply, Tobin’s Q, R&D Spending, Sales.

In both cases (Models 2 and 4), the estimated coefficients are positive ( $\beta_1=0.47$  and  $\beta_2=0.59$ , respectively), highlighting the importance of labor demand in AI publications. However, the estimates in both cases are slightly larger than the prior estimates. There is a plausible explanation for this: the original fixed effects estimates might have a downward bias due to measurement error. The IV estimator alleviates the noise in the measurement error by being more precise—focusing on

firms that were more exposed to AI.

Taken together, results from my IV estimates suggest that, indeed, it is likely that the demand for increased AI talent is driving the increased number of publications.

### Examining the use of AI Science in AI Patents and Firm-University Collaboration

Table A3: Internal Use of AI Science in AI Patents and Firm-University Collaboration

Model	AI Papers (ln)			Firm-only AI Papers (ln)		
	(1)	(2)	(3)	(4)	(5)	(6)
AI Job posts	0.067 (0.008)			0.031 (0.011)		
AI PhD Job posts		0.118 (0.018)			0.050 (0.020)	
DL PhD Job posts			0.149 (0.024)			0.105 (0.028)
Internal Use of AI Papers (ln) <sub>(t-1)</sub>	-0.017 (0.085)	-0.005 (0.083)	0.063 (0.084)			
Firm-University Collaboration (ln)				0.218 (0.040)	0.221 (0.040)	0.222 (0.038)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.277	0.276	0.276	0.201	0.202	0.211
Observations	18280	18280	18280	18280	18280	18280

Note: Standard errors are clustered at the firm-level, which are shown in parentheses. Controls: AI patents, non-AI patents, non-AI publications, AI pub Trends, AI PhD Supply

Table A4: Placebo Test and Granular Analysis

Sample	Startup			All				
	AI Papers (1)	AI Papers (2)	Non-AI Papers (3)	Non-AI Papers (4)	AI Conference (5)	AI Journal (6)	DL Conference (7)	DL Journal (8)
AI PhD Job Posts	-0.026 (0.024)			0.001 (0.012)	0.118 (0.020)	0.051 (0.019)		
DL PhD Job Posts		-0.009 (0.030)					0.296 (0.030)	0.104 (0.027)
Non-AI PhD Job Posts			0.001 (0.009)					
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Financial Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.222	0.222	0.078	0.079	0.289	0.188	0.240	0.138
Observations	1130	1130	5308	5309	5309	5309	5309	5309

Note: Standard errors are clustered at the firm-level, which are shown in parentheses. Controls: AI patents, non-AI patents, non-AI publications, AI pub Trends, AI PhD Financial Controls: Supply, Tobin's Q, R&D Spending, Sales.

## **Appendix A6: Use of Science in AI Patents**

In this section, we describe how we take into account the use of AI science in firms' AI Patents. First, I combine all the corporate AI patents with the reliance on science data's front-page non-patent literature (NPL) references (Marx and Fuegi, 2020). Next, I match all the NPL titles with all the papers published by corporate papers, which resulted in 5342 papers. From this dataset, I take the subset of AI papers that have been cited in AI patents and compute the annual count for the firms that were citing them. I use this annual count of AI papers at the firm-level as internal use of AI papers. Following prior literature (Arora et al., 2021), I use a lagged variable. However, other usage of this control also provides similar estimates.

## **Appendix A7: Science Intensiveness of AI Patents**

To calculate the intensity of science in AI patents, I combine all the AI patents from 2000 to 2019 with the reliance on the science dataset (Marx and Fuegi, 2020). After that, for each year, I calculate the intensity of science for patent innovation as the ratio of the total number of AI patents receiving at least one citation from scientific literature within a given year to the overall number of AI patents filed within the same year.

I calculate the intensity of science for patent innovation as follows:

$$\text{intensity of science for AI patents} = \frac{\text{total number of AI patents that have at least one citation to science in a year}}{\text{total number of AI patents in that year}}$$

## **A Appendix A8: Analysis with Lagged variable**

To provide additional evidence against the reverse causality argument, I use lagged variables. Qualitative data suggests that in AI, hired talent would require one to two years to start publishing

in peer-reviewed outlets.<sup>24</sup> Therefore, one-year and two-year lags should provide ample evidence of whether firms were hiring scientists to publish AI papers. In Table A5, Models 2 and 3, I use one-year and two-year lagged variables respectively. If hiring would lead to more publications then Models 2 and 3 would have bigger coefficients than Model 1. Consistent with the theory, I find that firms Model 1 is bigger than Models 2 and 3. I do the same analysis with deep learning PhD job posts in Models 4-6, which again produce similar results. Taken together, these tests provide suggestive evidence that firms *simultaneously* publish more papers as they try to recruit AI talent.

---

<sup>24</sup>In AI, the leading conference deadlines are spread out throughout the year, which facilitates a faster turnout time. Interestingly, the pace of fast publications has been criticized by leading AI researchers (Bengio, 2020).

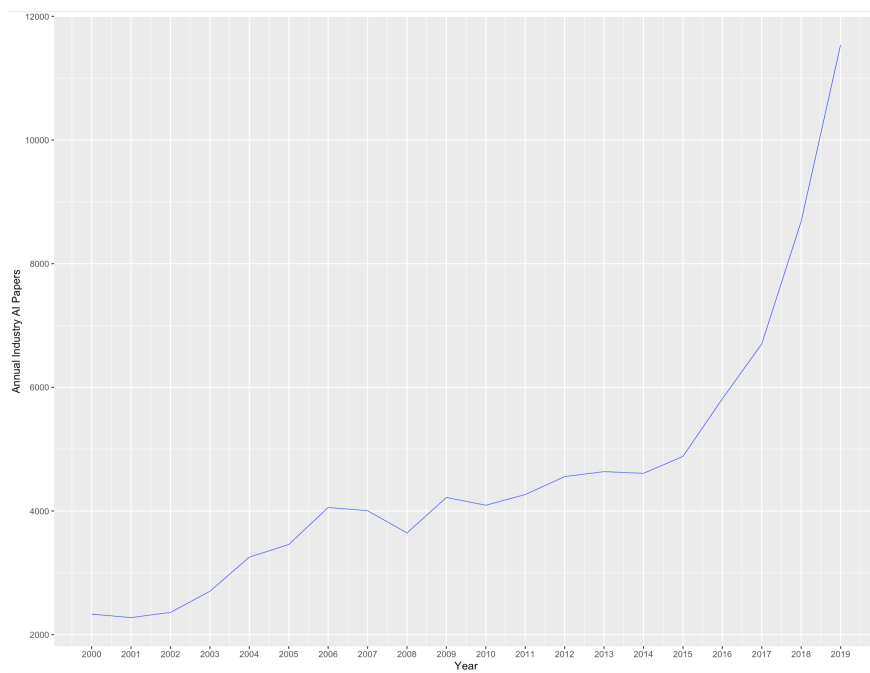


Figure A4: The rise of AI research. In the Y-axis, we present the annual count of all the AI publications by all the firms in my sample from 2000-2019.



Table A5: Estimates from the Fixed Effects Model

Model	AI Papers (ln)					
	(1)	(2)	(3)	(4)	(5)	(6)
AI PhD Job posts	0.105 (0.021)					
AI PhD Job posts (t-1)		0.053 (0.010)				
AI PhD Job posts (t-2)			0.037 (0.010)			
DL PhD Job posts				0.115 (0.024)		
DL PhD Job posts (t-1)					0.112 (0.030)	
DL PhD Job posts (t-2)						0.099 (0.044)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Financial Controls	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R <sup>2</sup>	0.319	0.317	0.312	0.317	0.315	0.312
Observations	5309	5308	5307	5309	5308	5307

Note: Standard errors are clustered at the firm-level, which are shown in parentheses. Controls: AI patents, non-AI patents, non-AI publications, AI pub Trends, AI PhD Supply. Financial controls: Tobin's Q, R&D Spending, Sales.

Table A6: Estimates from Difference-in-Differences

Method	AI Papers (ln)					
	Gardner			Athey		
	All	Research-firms	Public	All	Research-firms	Public
Sample Model	(1)	(2)	(3)	(4)	(5)	(6)
HighAIExposure*Post	0.082 (0.017)	0.073 (0.023)	0.054 (0.024)	0.081 (0.016)	0.073 (0.024)	0.054 (0.025)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Financial Controls	No	No	Yes	No	No	Yes
Observations	35920	10660	10608	35920	10660	10608

Note: Data is from 2000-2019; Bootstrapped Standard errors are clustered at the firm-level, which are shown in parentheses. Controls: AI patents, non-AI patents, non-AI publications, AI pub Trends, AI PhD Supply. Financial controls: Tobin's Q, R&D Spending, Sales.