Human Capital's Role in Corporate Adoption of Socially Responsible Practices:

The Case of Responsible Artificial Intelligence Principles

Nur Ahmed*
MIT Sloan School of Management

Nan Jia
Marshall School of Business
University of Southern California

* The authors contributed equally and are listed alphabetically by their last names.

This version: January 12, 2024

Research Summary

Artificial intelligence (AI) scientists are strong advocates for responsible AI practices. The competitive market for AI talents compels AI technology firms to commit to responsible AI principles to attract top talent. We hand collected responsible AI principles of over 8,700 firms, and connected the data with firms' AI job postings, academic publications, and patents. We demonstrate that firms with a growing demand for AI scientists, particularly in deep learning, which has an even tighter labor market, are more inclined to adopt responsible AI principles. Moreover, corporate AI scientists' collaborations with academia and their PhD granting institutions' publications on responsible AI further predict firms' commitment to responsible AI principles. Our findings highlight labor market's significant role in shaping corporate social responsibility practices within the AI industry.

Managerial Summary

For AI companies, embracing socially responsible AI practices is not just about ethics; it is a strategic move to attract top talent, because AI scientists are among the most avid advocates for responsible AI practices. Our study of over 8,700 AI companies found a strong link between their commitment to responsible AI principles, hiring patterns, and innovation output. Companies actively seeking AI experts in competitive markets for talents are more likely to commit to responsible AI practices. When their AI scientists collaborate with academia or are trained by universities with stronger responsible AI publications, companies are also more likely to adopt responsible AI principles. This study highlights how the competition for human capital drives companies to prioritize socially responsible practices highly valued by their talents.

It is an increasingly accepted that a firm's employees can wield significant influence over corporate social responsibility (CSR) initiatives. This belief stems from the understanding that higher employee turnover, particularly when fueled by dissatisfaction, can be costly for firms striving to retain talent (Bode, Singh, & Rogan, 2015; Portocarrero & Burbano, 2023). However, labor market research suggests that employees' bargaining power diminishes when firms can easily replace them (Brown, Gianiodis, & Santoro, 2015; Mailath & Postlewaite, 1990). Our research intends to reconcile this tension by focusing on the burgeoning field of responsible Artificial Intelligence (AI). The rapid advancements of AI technologies generate serious concerns over their social consequences on fairness, accountability, transparency, privacy, and security. Consequently, companies at the forefront of AI technology development are increasingly expected to establish "responsible AI principles" that place constraints on their research and development activities towards the broader social good (Jobin, Ienca, & Vayena, 2019). Importantly, it is often the scientists engaged in AI development who are among the most vocal advocates for such principles (Vincent, 2018).

In our study, we argue that an AI technology firm's demand for AI scientists drives the firm to commit to responsible AI principles, because competitive nature of the AI scientist labor market bolsters their bargaining power vis-à-vis firms, such that to attract talent, firms must offer comprehensive packages that include both pecuniary and non-pecuniary benefits (Ahmed, 2022), including a pledge to responsible AI development. Thus, this relationship is particularly strong when the firm actively compete for talent within the most competitive deep learning labor market segment. We further examine the heterogeneity of AI scientists regarding their propensity to advocate for responsible AI principles. Recognizing that the discourse on responsible AI mostly takes place within academic institutions, we explore two pathways linking firms' AI scientists

with universities: a "spillover channel" arising from collaboration with academic researchers, and a "training channel" originating from the PhD programs that educated these AI scientists. We thus hypothesize that firms are more likely to commit to responsible AI principles when their AI scientists collaborate more extensively with university researchers and when these scientists have received their doctoral training from institutions that more actively involved in responsible AI research.

There exists no systematic compilation of AI principles by any industrial association or social activists. Therefore, we hand-coded whether AI-developing firms have released an AI principle and analyzed the content of these principles from 2018 through 2022 using topic modeling (Hannigan et al., 2019).This effort resulted in the collection of 125 AI principles. Before empirically testing our hypotheses, we offer evidence-based descriptions of what responsible AI principles are and firms' need for AI scientists. Our textual analysis of the AI principles reveals that they are highly homogeneous, exhibiting a significant overlap in content. This convergence in what is expected in responsible AI principles is consistent with prior research (Jobin et al., 2019). Such convergence may mitigate concerns that heterogeneity in AI principles may lead to divergent demands on firms. Furthermore, we show that corporations are at the forefront of conducting basic research in AI, which explains their demand for AI scientists, and that there is a shortage of AI scientists in the labor market.

We combine the data on AI principles with various datasets, including the Burning Glass dataset which covers over 200 million job postings in the United States, over 1.7 million patents approved by the USPTO and over 1.1 million publications in top-tier computer science journals and conferences. Our analysis reveals that firms with a 1 unit increase in job postings for AI research positions exhibit a 22% higher odds of embracing responsible AI principles. We find

that this relationships is primarily driven by companies actively seeking talent in deep learning, the most cutting-edge and recent segment of AI technology, which faces the most acute shortage of skilled researchers. To eliminate the possibility that firms adopt AI principles solely because they are engaged in AI research and therefore more inclined to recognize the significance of responsible AI, we make a clear distinction between two groups: AI scientists, who generate AI research and are known to predominantly advocate for responsible AI principles, and AI inventors, who apply this research to develop AI patents,, may also encounter the necessity for responsible AI, but do not actively advocate for responsible AI. Consistent with our argument, a firm's commitment to responsible AI principles only correlated with its AI research but not its AI patents.

Finally, we present results concerning the heterogeneity of AI scientists within firms, akin to a "dose-response" approach aimed at generating evidence that is closer to the theoretical mechanism (Callaway, Goodman-Bacon, & Sant'Anna, 2021). Through ad hoc analysis, it becomes evident that universities are at the forefront of responsible AI research. We then show that a 1 unit increase in collaboration with university researchers is associated with a 28% increase in a firm's odds of committing to responsible AI principles. Furthermore, when AI scientists are graduates of PhD-granting computer science departments ranked top 50 in conducting responsible AI research, firms odd to enhance their adherence to responsible AI principles by 19%. These seemingly disparate findings collectively contribute to the body of evidence supporting the key theoretical argument that the acceptance of the importance of responsible AI by AI scientists is a driving force behind firms' adoption of responsible AI principles.

This paper demonstrates the value of bridging two otherwise separate fields. In the realm of CSR research, our paper provides a robust theoretical foundation and substantial empirical evidence to contextualize a firm's adoption of CSR within the labor market, the source of its human capital. While the CSR literature recognizes that demand by employees can influence a corporation's CSR practices (Bode et al., 2015; Portocarrero & Burbano, 2023), it is noteworthy that not all labor and not at all times can exert this influence. Factors such as tight labor market conditions, a firm's reliance on key employees, and the degree to which these key employees prioritize CSR all play pivotal roles in enhancing a firm's commitment to CSR practices (Brown et al., 2015). To this literature, the study of responsible AI principles shows that firms' adoption of CSR is inherently connected with labor market demand, firms' technological advancements, and features of human capital within firms.

This study generates novelty insights into the rapidly growing field of AI technology development. While firms are actively engaged in AI technology development (Ahmed, Wahed, & Thompson, 2023; Miric, Jia, & Huang, 2023), they also face increasing pressure to address the social consequences of AI, along with the looming threat of regulations (Candelon, di Carlo, De Bondt, & Evgeniou, 2021). Despite the recognition of this issue, there exists limited academic research on when and why firms would voluntarily commit to responsible AI principles. By integrating approaches from the CSR literature, we reveal that AI scientists play a pivotal role in driving a firm's commitment to socially responsible AI. This effect stems from the bargaining power of these talented individuals, which, in turn, is influenced by labor market conditions and firms' innovation activities. Consequently, we can make relatively precise predictions about where, when, and how responsible AI can further advance.

Finally, we underscore the importance of our novel dataset. To our knowledge, it represents the first systematic compilations of firms' responsible AI principles, essential for examining the social responsibilities of AI technologies. Additionally, we demonstrate the practical potential of linking this data with multiple existing large datasets on firms' technology development and labor demand.

RESPONSIBLE AI AND CORPORATE CONCERNS

Rapidly advancing AI technologies are poised to reshape a broad spectrum of socioeconomic activities (Agrawal, Gans, & Goldfarb, 2019). While it is evident that AI technologies can generate substantial benefits, such as increased efficiency (Tong, Jia, Luo, & Fang, 2021), concerns over the potential social harms they may generate have also become more prominent. Issues of fairness and bias arise, questioning the equitable distribution of AI's benefits and burdens (Cowgill, Dell'Acqua, & Matz, 2020). Accountability remains a critical point of contention, as stakeholders demand clarity on who bears responsibility for AI-driven decisions (Collina, Sayyadi, & Provitera, 2023). Transparency in AI processes is another pressing issue, as opaque algorithms challenge the very fabric of open, informed decision-making (Pedreschi et al., 2019). Privacy concerns are amplified as AI's capabilities in data processing outpace current regulatory frameworks, raising alarms about the security of personal information (Brayne, 2017).

AI scientists and technologists champion responsible AI development due to their deep understanding and expertise in the field (Gofman & Jin, 2022). Their position at the forefront of technological innovation places them in a unique position to recognize both the transformative potential and the challenges that come with advanced AI systems. They are often the first to

identify and address the multifaceted risks (Buolamwini & Gebru, 2018). This is evidenced by the fact that most discussions and writings on responsible AI manifest in the form of academic publications (Ahmed, Das, Martin, & Banerjee, 2024). Recent research indicates that frontiers responsible AI research is predominantly undertaken by universities, which excel both in the volume and the quality of their contributions (Ahmed et al., 2024). In contrast, the industry has a limited presence in responsible AI research although they are dominating the frontiers of research in AI technologies (Ahmed et al., 2023).

Activists advocating for responsible AI typically push for several key principles. The first is to guide AI development in a direction that is beneficial, fair for society as a whole and respectful of human rights. For example, concerns about AI tools being unfair to racial minorities or AI tools exhibiting sexism are notable among technology workers. Additionally, they want these tools to be used with certain boundaries in mind, such as no military usage, no usage in capturing illegal immigrants, and no outsourcing decisions to kill humans to AI. Moreover, privacy and security are also key concerns. In particular, the Cambridge Analytica scandal triggered the response against misuse of user data.

However, the advocacy for responsible AI principles often encounters reluctance, if not outright resistance, from the very firms that develop AI technologies (Ali, Christin, Smart, & Katila, 2023). Firms' main concerns are that the adoption of these principles introduces additional constraints that are seen as compromising the commercial objectives.

First, the adoption of responsible AI immediately incurs staffing costs, and maintaining these personnel introduces an additional layer of expenses. In fact, many firms downsized their AI ethics teams during periods of layoffs, indicating that these teams could be costly without significantly contributing to commercial value (Criddle & Murgia, 2023).

Second, the implementation of responsible AI principles can introduce delays in the development process. It may require extensive reworking of AI systems to ensure fairness and privacy, and it can also stall the development of certain profitable technologies that are questioned for their alignment with these principles. This slowdown can result in firms losing their first mover advantage. For instance, consider the case of Google, where internal resistance delayed the launch of their own ChatGPT. The company had to grapple with addressing ethical concerns and ensuring responsible AI practices before introducing the technology to the market, which led to a missed opportunity for an early competitive edge (Metz & Isaac, 2023).

Third, a palpable concern exists that adhering to responsible AI principles might restrict the autonomy of firms, particularly regarding their commercial motives, as these principles can be perceived as external impositions. An illustrative example is Google's decision in October 2018 to withdraw from consideration for a lucrative $10 billion Pentagon contract. The company attributed its withdrawal to a conflict with the ethical principles it had recently introduced for artificial intelligence. This move came in response to employee protests over a previous Pentagon contract known as Maven, which applied machine learning to drone imagery (Wakabayashi & Scott Shane, 2018).

Furthermore, strict adherence to responsible AI might be seen as negatively impacting a firm's competitive edge, particularly if their rivals do not follow suit. In fact, observers tend to hold large technology firms to a different standard than smaller startups, therefore they have to move slower and more cautiously than others (Metz & Isaac, 2023). Similarly, Apple has been cautious in their approach to generative AI, the CEO said "We've been working on generative AI for years and have done a lot of research," … "[a]nd we're going to approach it really thoughtfully and think about it deeply, because we're fully aware of the not-good uses that it can

have, and the issues around bias and hallucination and so forth. You know, we've never felt an urgency to be first, we've always felt an urgency to be best, and that is how we go into this as well." (Phelan, 2023)

Finally, committing to responsible AI principles can give rise to additional concerns, particularly regarding what is often referred to as "ethics washing." An illustrative case involved criticism directed at Amazon for its donation to the National Science Foundation (NSF) in support of AI research. One academic, Nic Weber, Assistant Professor at the University of Washington's iSchool, questioned this action, stating, "Why does Amazon get to prominently feature its logo on a national solicitation (for a relatively modest $7.6 million in basic research) when it profits in the multibillions from AI that is demonstrably unfair and harmful?" This example highlights how public perception and skepticism surrounding a firm's commitment to responsible AI principles can come into play (Lahoti, 2019).

Despite such resistance, it is crucial to engage firms in adhering to responsible AI principles, because firms play a central role in advancing AI research, unlike the diminishing involvement of companies in the corporate science of other fields (Arora, Belenzon, & Patacconi, 2018). We turn to this point next.

HYPOTHESES: LABOR MARKET AND CORPORATE COMMITMENT TO RESPONSIBLE AI

The Corporate Drive Behind AI Research

Over the last thirty years, a strong pattern has formed in scientific research, characterized by a "division of innovative labor" (Arora et al., 2018). Historically significant corporate R&D labs, like Xerox PARC, have seen a reduction in their numbers and research contributions. In

contrast, universities have stepped up as the main contributors to basic research. This shift is largely due to their access to government-funded research grants and a drive to pursue studies that might not yield immediate commercial benefits. Consequently, corporations are increasingly leaning on university research funded and shared publicly, especially in basic research domains (Fleming, Greene, Li, Marx, & Yao, 2019).

However, when it comes to the development of AI technologies, the roles are *reversed*. Firms are now at the forefront of basic AI research and are producing more influential studies in this field (Ahmed et al., 2023). A key factor influencing this trend is a unique aspect of AI research: the necessity for extensive datasets and computational power.

First, universities struggle to collect large-scale databases as effectively as corporations do through their business operations (Shokri & Shmatikov, 2015). In contrast, companies gather substantial user data through their business activities, a practice that has intensified recently (Hartmann & Henkel, 2020). This abundance of data enables corporations to engage in basic AI research to navigate the technical complexities of managing and analyzing large datasets. Google, for instance, has developed numerous methods like MapReduce to address the challenges of processing extensive internet data. Although academic researchers contribute valuable theoretical knowledge to AI, many of the practical breakthroughs and innovations in the AI field have been driven by corporate efforts.

Second, significant computational resources are required to train deep learning models (Hestness et al., 2017). Studies indicate that the advancements in AI and its enhanced effectiveness compared to previous techniques can be largely attributed to the increased availability and application of computing power (Hestness et al., 2017; Thompson, Greenewald, Lee, & Manso, 2020). Industry has a significant advantage in since they own large data centers

which allow them to access compute at a scale that would be prohibitively costly for nonprofits or universities. Furthermore, many AI firms have designed their own computing chips similar to GPUs to gain and maintain competitive advantage. Consequently, research suggests that industry AI models are on average 29 times larger than academic AI models (Ahmed et al., 2023).

Securing AI Scientists in a Competitive Labor Market

To bolster their AI research, companies must attract highly skilled AI scientists, individuals deeply versed in scientific methodology and adept in the scholarly discourse surrounding AI. These experts are typically equipped with PhDs, a qualification that necessitates an average of 5-6 years of rigorous academic training. Academic institutions invest considerable time and effort in identifying prospective students and training them into PhD graduates in AI. However, this training process is slow for the surging demand for such skilled professionals in the labor market. Both corporate entities and society at large have recognized the transformative potential of AI technology and are eager to capitalize on its applications, fueling an intense competition for these valuable experts (Metz, 2017; The Economist, 2016a). Therefore, the job market for AI scientists is remarkably competitive. Due to the shortage in talent supply prompted private firms to hire away faculty members from universities which negatively affected local startup formation (Gofman & Jin, 2022). For example, Carnegie Mellon University, a leading AI research university lost 50 AI scientists including tenure track faculty members to Uber in 2015 (Lowensohn, 2015). As described by Peter Lee, Co-head of Microsoft Research, there is a "bloody war for talent in this [AI] space" (Parloff, 2016).

In a labor market characterized by high demand and a limited supply of workers, the dynamics of value distribution shift in favor of employees, enhancing their bargaining power to obtain what they desire (Brown et al., 2015; Molloy & Barney, 2015). Elite scientists desire

more than monetary compensation. These individuals highly value the freedom to pursue their own research interests—work that not only aligns with their intellectual passions but also makes a significant contribution to their field and is recognized by their peers (Merton, 1973). As a result, non-pecuniary rewards, such as the liberty to investigate topics of personal curiosity and to shape the trajectory of the research community are crucial (Agarwal & Ohyama, 2013; Ahmed, 2022). This autonomy in research direction is not just a peripheral benefit; it is often a decisive factor in attracting and retaining top talent. For instance, Chris Nicholson, CEO of the deep-learning company Skymind, has noted, "if you try to recruit AI researchers by promising lots of money and zero peer recognition, you won't get very far." (Alba, 2017).

Corporations that understand and cater to this intrinsic motivation often find themselves at an advantage. For example, Kim & Mahoney, (2007: 21) describes Merck's success in securing researchers 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 contrast, firms that do not cater to AI talent will have a hard time in recruiting and retaining talent. For instance, Palantir CEO said that "I've had some of my favorite employees leave" due to their objection to the firm's acceptance of the contract with Immigration and Customs Enforcement (ICE) (Allen, 2020).

When considering the various factors at play, we argue that firms heavily involved in AI research have a substantial demand for AI scientists. In a tight labor market, they have less bargaining power vis-à-vis these experts. This dependence makes firms more susceptible to the influence of AI scientists, leading them to be more responsive to the specific needs of these professionals. When AI scientists promote responsible AI practices, firms deeply immersed in AI

research and thus heavily reliant on AI scientists are more inclined to accommodate these preferences, despite the increasing costs and the perceived risks that responsible AI principles may pose to their R&D process. Consequently, these firms tend to adopt and uphold the principles of responsible AI. Therefore, we formulate our first hypothesis (H1) as follows:

Hypothesis 1 (H1): *Firms with a higher demand for AI scientists are more likely to commit to responsible AI principles.*

Exacerbating Factor: Corporate Demand for Deep Learning Scientists

We now turn to a particularly high-demand segment of the AI talent labor market, which consists of deep learning researchers. Deep learning, a subbranch of AI and machine learning, gained significant prominence since the 2010s. It involves the use of deep neural networks, which are composed of multiple layers of interconnected nodes (artificial neurons). These deep neural networks are designed to learn complex patterns and features from large amounts of data (LeCun, Bengio, & Hinton, 2015).

Deep learning models have had a profound impact on the development of AI technology for several key reasons. First, deep learning algorithms excel at processing and learning from massive datasets, enabling them to identify intricate patterns and make advanced predictions. This capability has led to groundbreaking advancements in fields such as image and speech recognition (Russakovsky et al., 2015; The Economist, 2016b). Second, deep learning algorithms possess a self-learning capability that sets them apart from traditional programming. They do not require explicit instructions or feature engineering to learn from data. As a result in in areas where there is ample data available like natural language processing and computer vision they perform well (LeCun et al., 2015). Third, deep learning models are highly versatile and adaptable to a wide range of applications, spanning from healthcare diagnostics to financial modeling. Their scalability enables them to tackle complex real-world problems (The Economist, 2016b).

For these reasons, there is a substantial demand for the advancement of deep learning technologies. However, it's important to note that deep learning represents just one subset of AI/ML technologies, and the number of scientists trained in this field each year remains limited. Within the broader AI field, deep learning constitutes a relatively small subfield. Before 2012, there was a notable stigma associated with neural networks, as this area of research had not yielded the desired results (Hooker, 2020). Only a small group of academics were engaged in research within this specific domain..

Moreover, the tacit nature of underlying deep learning knowledge has contributed to challenges in rapidly developing a vast pool of skilled talent. This difficulty is exacerbated by researchers' limited understanding of the mechanics behind deep learning, a factor contributing to the limited codification of knowledge in this field (Chen, 2019; Martineau, 2019). Consequently, the acquisition of this knowledge requires extensive first-hand experimentation. This has resulted in a slow dissemination of expertise, creating a notable disparity between the demand for and the availability of skilled professionals in this area.

The undersupply of deep learning talent, compared with the market demand for them, resulted in a higher bargaining power for them. This is evident in their salary. The Vox (Bergen & Wagner, 2015) reports from 2015 that “An engineer proficient in deep learning can earn upward of $250,000 a year at places like Google and Facebook, according to several sources; exceptional or more experienced ones can net seven-figure salaries.” This increased payment to deep learning professionals continued even in 2018. The New York Times wrote “… A.I. specialists with little or no industry experience can make between $300,000 and $500,000 a year in salary and stock. Top names can receive compensation packages that extend into the millions.” (Metz, 2018).

Because of the even tighter labor market for deep learning scientists, we argue that firms' demand for deep learning scientists particularly drives their commitment to responsible AI principles.

Hypothesis 2 (H2): *Firms with a higher demand for deep learning scientists are even more likely to commit to responsible AI principles.*

Heterogeneity of Corporate AI Scientists

At the heart of the current analysis is the presupposition that AI scientists are proponents of responsible AI principles. We now examine heterogeneity among AI researchers concerning their engagement with this discourse. Variability in exposure to conversations on responsible AI among scientists can serve as a "dose-response" test (Callaway et al., 2021). This approach allows us to quantify the outcome of committing to AI principles based on corporate scientists' varying level of exposure to the discourse on responsible AI, thus taking us closer to the underlying mechanism.

The responsible AI discourse predominantly occurs within academic institutions (Ahmed et al., 2024). We propose two primary channels that potentially expose corporate AI scientists with these principles. The first is the "spillover channel," where corporate scientists in collaboration with academic researchers may find themselves more exposed to the currents of thought advocating for responsible AI. Such collaborations, which often result in joint academic publications, are common for technology firms (Mindruta, 2013). As corporate scientists work alongside their academic counterparts, they are more likely to engage in the dialogue about responsible AI, increasing the potential to resonate with them. Consequently, firms whose scientists engage more frequently in collaborative efforts with university researchers in AI research and publication are more inclined to adopt responsible AI principles. This relationship is captured by Hypothesis 3 (H3)

Hypothesis 3 (H3): *Firms whose AI scientists collaborate more extensively with academia are more likely to commit to responsible AI principles.*

Corporate scientists may have been influenced by the prevailing academic discourse on responsible AI, primarily through their foundational education in PhD programs. These programs, often housed within various academic computer science departments, serve as an important "training channel," imparting not only technical knowledge but also ethical standards. It is important to note that the level of engagement in responsible AI research is not uniform across these departments; some are known for their prolific contributions to the topic, consistently publishing on responsible AI development. While engagement in conventional AI research is correlated with responsible AI research, there is a marked difference in the level of engagement. For instance, MIT, CMU, GeorgiaTech are the top 3 producers in conventional AI research. However, in responsible AI research CMU, Harvard and UW are the top 3 producers of responsible AI research. In particular, Harvard and UW are number 11th and 15th respectively in conventional AI research. This showcases that being a leader in responsible AI research does not automatically result in a leader in conventional AI research. Scientists who have been trained in such proactive academic environments are likely to have developed a deeper understanding of and commitment to responsible AI. This, in turn, enhances the likelihood that they will champion responsible AI principles in their professional endeavors within the corporate sector. Their academic lineage, therefore, has significant implications for the adoption of responsible AI principles by firms, as captured by Hypothesis 4 (H4):

Hypothesis 4 (H4): *Firms whose AI scientists received their doctoral training from institutions more actively involved in responsible AI research are more likely to commit to responsible AI principles.*

DATA, VARIABLES, AND METHODS

Identify AI Firms

To create a comprehensive list of firms that develop AI technologies, we start with the USPTO patent data. We define a firm with AI research experience if that firm has at least one AI patent with the USPTO under the class *"computer systems based on specific computational models."* This CPC class was selected after extensive consultations with two USPTO examiners. This resulted in a sample of 1826 firms with at least one AI patent between 2000 and 2019.[1]

One limitation of this sample is that there might be many other firms that have published AI principles but are not part of our sample since may not have secured an AI patent by 2019. Therefore, we take a more expansive approach to augment the current sample. We use Burning Glass Technologies' AI job posts data to complement this sample. The presence of AI job postings serves as an indicator of a firm's involvement in AI activities. Specifically, we focused on companies with a minimum of five AI-related job advertisements on BGT's platform in 2019. This criterion yielded a list of 8,734 firms engaging in AI including the AI-patent holding firms. It should be noted, however, that the vast majority of these entities were private firms or startups and their AI research engagements were often minimal or nonexistent (as measured by their AI research or patenting activities).

Collection of Responsible AI Principles

The absence of a centralized repository of organization-level responsible AI principles led us to hand-collect the data. We took an extensive manual process to collect comprehensive firm-level responsible AI principles, relevant links, and associated text data. This extensive data

[1] To validate our result we use (Miric et al., 2023)'s dataset on AI patents. We find that our AI patent holding firms cover more than 90% of AI patents listed by their dataset. This increases confidence that our dataset is capturing the AI research and commercializing organizations.

collection effort was conducted by a team of six research assistants who manually searched the web. First, the research assistants used *Google.com* with the firm name and an extensive list of responsible AI related keywords[2] and if a firm's responsible AI principle was not found they also searched in *Bing.com.* Once a relevant webpage was identified, the research team utilized *archive.org*—a digital repository that regularly archives web pages—to retrieve the corresponding details (e.g., the first time the website was posted online). The full text detailing the responsible AI principle was then carefully transcribed. Additionally, the initial publication year for each set of principles was systematically recorded from the archive.org's first appearance. This website allows us to observe the first time a firm publicly adopts responsible AI principles. Subsequently, a pair of research assistants independently reviewed each firm entry to avoid missing data. Finally, to bolster the reliability of our findings, a co-author performed a random audit of the assembled list, confirming the accuracy of the collated information. This data collection process was undertaken between November 2022 and May 2023.

We restricted our dataset to only include AI principles published in English. During our search process, we excluded any unofficial materials such as firms' engagement on Twitter or other social media posts. Here, we consider only the official website, under the company's main domain (or subdomain), or official blog posts as valid indicators of the adoption of responsible AI principles. This resulted in a dataset of AI ethics principles from 125 firms to our knowledge, of

[2] The keywords include: "AI ethics principles", "Trustworthy AI", "AI ethics guidelines" "Responsible AI", "Accountable AI", "Artificial Intelligence Principles", "Artificial Intelligence Guidelines", "Artificial Intelligence Framework", "Artificial Intelligence Ethics", "Robotics Ethics", "Data Ethics", "Software Ethics", "Artificial Intelligence Code of Conduct", "AI policy", "AI fairness", "Trust in AI" and "Explainable AI".

which 86 of them are public firms. This dataset is the most comprehensive private responsible AI principles dataset to our knowledge.

For additional validation exercise, we consulted two existing databases of AI principles, AIethicist.org and algorithmwatch.org. Both sources include a number of public and private responsible AI principles. However, these sources were not regularly up to date and included data on primarily reputed organizations, therefore increasing concerns about selection bias in the sampling process.

Key Variables

Responsible AI Principles (ResAI) is the dependent variable indicating if a firm has publicly adopted a responsible AI principle or not. This binary variable takes the value 1 if a given firm releases a publicly available responsible AI principle.

Demand for AI Scientists is the first key explanatory variable. To create this variable first, we classify AI job posts with an extensive list of keywords (see Appendix B) based on prior literature (Alekseeva, Azar, Gine, Samila, & Taska, 2019) Subsequently, we search for job postings that specify a requirement for a PhD degree. The total count of job postings with such a requirement are then recorded and log transformed to create the final count. Finally, we take the average of the prior 3 year's of data.

Using the same approach, we calculate the *Demand for Deep Learning Scientists* by identifying job postings related to deep learning, utilizing a specific set of keywords (see Appendix A). We then evaluate whether these posts require a PhD degree for the position. Finally, we take the average for the last 3 year's data and log transform it as before.

We collect all AI firms' patent portfolios from the USPTO, both AI and non-AI patents. Next, we merge this patent 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, we have each firm's AI publications and non-AI publications. To classify AI papers, we use an extensive list of keywords presented in Appendix A1. Finally, we merge this data with the job posting data from Burning Glass Technologies. To merge with burning glass technologies data, we manually searched for company names. To ensure accuracy and completeness, we painstakingly refined these search processes.

Firm-University Collaboration: We take the average of the prior 3 year's of total number of firm papers that had at least one university collaborator and then log transform it as before. We create two more variations of it: *Firm-University Collaboration AI and Firm-University Collaboration DL,* indicating if that is AI collaboration or deep learning paper collaboration respectively.

Firm-University Responsible AI Recruitment: This measures to what extent firms' recruited AI scientists had responsible AI exposure at graduate school. To calculate this, first, we calculate a responsible AI score for the top 50 universities[3]. This score is a ratio of total responsible AI research to total conventional AI research. Then we count the number of AI scientists recruited by each firm and weight the total number with the aforementioned score. This final score helps us to test the training channel for AI scientists.

Number of AI patents. we calculate the average of the total AI patents filed over the past three years. Then we log transform the average count. We collect this AI patent data from USPTO and use a specific CPC class (*computer systems based on specific computational models*) to classify AI patents[4].

[3] The list of universities were obtained from csranking.org, a well-known Computer Science ranking website
[4] We excluded quantum computing related patents from AI patents

Number of non-AI patents. We calculate the average number of AI patents filed over the last three years and apply a logarithmic transformation to this average count. We collect this data from USPTO. If a CPC class was not under the AI patent category, it was regarded as a non-AI patent.

Number of AI papers. To create this, as before we calculate the average number of AI patents filed over the last three years and then apply a logarithmic transformation to this average. We collect this data from Scopus and use an extensive list of words to classify AI papers.

Number of non-AI papers. we tally the total number of publications from the preceding three years that are not relevant to AI, applying a log transformation to this count. As before, this was collected from Scopus and if the paper did not have any AI keywords, we counted this under variable.

Control variables

To accurately isolate the effects of employee pressure on the adoption of responsible AI principles, we include a series of control variables in all of our models. We merge our data with the Compustat data.

Firm size: we count the average total number of employees for the past three years and log transform it. *Firm revenue:* we calculate the average total revenue for the past three years and log transform it. *Firm R&D Expenditure:* we take the average of total R&D spending and log transform the variable to consider the skewness of R&D spending among firms. This variable helps us to control for firm-level time variant factors that could affect their recruiting and spending in CSR.

Method:

We estimated the likelihood that a firm would adopt a public responsible AI principle using the following logistic regression model, captured by Equation (1)

$$Pr\,(ResAI_i = 1) = F\,(\beta_1\,Demand\ for\ AI\ research\ scientist_i + \beta_2\,Demand\ for\ Deep\ Learning\ research\ scientist_i + \beta_3\,Firm\text{-}University\ Collaboration_i + \beta_4\,Industry_j + \beta_5 Controls_i + \varepsilon_i) \quad Eq\ (1)$$

Where i represents firm, j represents industry. ResAI is the dependent variable indicating if a firm has publicly adopted a responsible AI principle or not. We cluster the standard errors to consider heteroskedasticity and correlation in the error term due to repeated measurements for each firm over time. $Controls_i$ is a matrix for firm-level controls like firm-size and firm revenue. Additionally, we include an industry level fixed effects based on the SIC code to control for industry factors that might affect the adoption of responsible AI principles.

RESULTS

Descriptions of Responsible AI Principles

We provide descriptions of the responsible AI principles we collect through the above step to offer a better understanding of our data and the research context. To discern the underlying themes, we use topic modeling, an unsupervised machine learning technique, analysis on the full text of 86 public firms' responsible AI principles. We analyzed the text with different topic numbers, which largely produced similar results. Here, we present the results for two sets of topics (namely 5 topics and 10 topics). In both cases, our analyses suggest that across all these principles core themes of responsible AI are quite similar. Consistent with prior research (Jobin et al., 2019)

we find that, firms have largely converged to a small set of core issues in their responsible AI principles. In particular, we find that *privacy, transparency, fairness* and *bias* are the major themes in these principles. These themes were also found in prior research in public and global nonprofit organizations' responsible AI research (Jobin et al., 2019). This is not surprising that firms were borrowing ideas from public organizations' principles.

INSERT Table 1 HERE

Our topic modeling analysis suggests that firms are not heterogeneous in their approach about AI principles. These principles are also consistent with the demands of AI scientists who care about the technology's downstream cases to be fair, unbiased and more beneficial to the broader society.

The plot of the number of AI firms' adoption of responsible AI principles over time suggests that increasingly firms are publishing more responsible AI principles. However, the total number of responsible AI principle is still quite limited to the active number of AI firms.

INSERT Figure 1 HERE

Descriptions of Labor Market Supply of AI Scientists and Deep Learning AI Scientists

In Fig 2, we present descriptive evidence that there was a significant gap between the demand for deep learning AI scientists and the total supply of AI scientists. This is a very conservative test because not all AI scientists could work on deep learning techniques. This is because deep learning research and commercialization requires extensive training.

INSERT Figure 2 HERE

RESULTS

To test our hypotheses, we use logistic regression and report the results for the independent and control variables in Table 2. Here, Model 1 suggests that the best predictor of the adoption of responsible AI is AI research publications (β=0.265, p=0.001). This indicates that 1 unit increase in AI research increases the odds of responsible AI adoption by almost 30%. On the other hand, the number of AI patents does not have a strong relationship (p=0.700). This confirms our intuition that it is the AI research *not* AI patents which is driving the adoption. In Model 2, we add a key independent variable, the *demand for AI scientists* and we find that the effect size is positive and a strong predictor (β=0.202, p=0.004). This supports our hypothesis 1, which states that firms with a higher demand for AI scientists are more likely to commit to AI principles. We find that 1 unit increase in the demand for AI scientists increases the odds of AI principle adoption by 22%. Similarly, in Model 3, we find that the demand for deep learning scientists is also positive and a strong predictor (β=0.197, p=0.016). This supports our hypothesis 2, which states that firms with a higher demand for deep learning scientists are more likely to adopt AI principles. We find that 1 unit increase in the demand for deep learning scientists leads to 22% additional odds of adopting responsible AI principles.

To get a better understanding of the bargaining power of deep learning scientists, we categorize AI scientists into two groups – deep learning AI scientists and non-deep learning AI scientists. We run a placebo test in Model 4 with the variable demand for *non*-deep learning AI scientists. We find that this variable is not a strong predictor (p=0.576) suggesting that it is primarily the bargaining power of deep learning scientists which led to the adoption of these principles.

INSERT Table 2 HERE

Next, we turn our attention to the next set of hypotheses which also get deeper into the mechanisms of our theory. In Table 3, we present the results of additional models where we consider firms' relationship with universities and the background of the hired AI scientists. Here, we present three different models each of which counts the number of firm-university collaboration. In Model 1, we find that 1 unit increment firm-university collaborative paper increases the odds of AI principle adoption by almost 28% (p=0.000). This confirms our hypothesis 3 is which states that if firms collaborate more with universities they are more likely to adopt responsible AI principles. Similarly, in Models 2 and 3, we find that both variables *Firm-University Research Collaboration AI* and *Firm-University Research Collaboration DL* are positively significant.

Finally, in model 4, we find that *Firm Responsible AI Recruitment*, which explicitly considers the ethics ratio of research of the universities, is also positive and significant (β=0.173, p=0.08). This supports our Hypothesis 4 which states that firms with AI scientists hired from institutions with stronger responsible AI research would have a higher likelihood of adopting responsible AI principles. We find that 1 unit increment in AI scientist recruitment increases the odds of AI principle adoption by almost 19%.

INSER Table 3 HERE

Robustness analysis

We conducted multiple supplemental analyses to corroborate our theoretical argument for the adoption of AI principles. For our measurement instead of 3-year periods, we also took 5-year

periods and the results are consistent with our prior results. This increases confidence that the results are not driven by a small temporal factor.

Additionally, we ran models with the rare event logit model (King & Zeng, 2001) to take into account the fact that our dataset has more 0s or non-adoption than adoption. This method produces relatively impartial and more consistent estimates of logit coefficients and their variance-covariance matrix, adjusting for the limitation of infrequent occurrences. We find that all four of our hypotheses were supported and found similar results.

Limitations

We acknowledge limitations of this study. Including potential unobservable factors that could affect both our independent and dependent variables. For example, one possible alternative explanation is that firms' involvement in AI research may heighten their awareness of the potential social harm these new technologies can cause, leading them to embrace responsible AI principles. However, our qualitative data and additional empirical analyses suggest that it was bargaining power which contributed to the result. In our model with the demand for deep learning scientists, we added AI research as a variable. We find that indeed the most significant factor was the demand for deep learning talent and the AI research variable was statistically insignificant.

One could argue that academic institutions' hiring data is not the best way to measure the impact of training on firms' adoption of AI principles. It is due to the fact that top universities that are highly productive in AI are also likely to be highly productive in responsible AI research. However, our analysis shows that it is not necessarily top institutions that do more research in responsible AI. Indeed, several institutions not typically recognized as top-tier in conventional AI research have outpaced their elite counterparts in responsible AI research. This

observation alleviates the concern about any assumed correlation between conventional AI research expertise and responsible AI research.

DISCUSSION AND CONCLUSION

In the rapidly evolving field of AI technology, there is a growing call from key stakeholders for companies to embrace social responsibility in their AI development efforts. Among the most vocal advocates for responsible AI principles are the AI scientists who conduct research in this domain. However, many companies are hesitant to commit to these principles, fearing that they might impede their business development. Our research reveals that firms with a higher demand for AI scientists are more inclined to adopt responsible AI principles. We propose that this is due to the scarcity of AI scientists in comparison to the surging demand for their expertise in technology-driven companies. Consequently, companies are more willing to accommodate their needs.

Corroborating this theory, we find that a firm's commitment to responsible AI principles is particularly influenced by their need for scientists in deep learning, which is the most sought-after market segment and faces the greatest gap between supply and demand for skilled professionals. Furthermore, to reinforce our argument that AI scientists drive firms to embrace responsible AI principles, we demonstrate that the exposure of corporate AI scientists to academia, where discussions on responsible AI are prevalent, plays a pivotal role in shaping a firm's commitment. This exposure includes collaboration between AI scientists and researchers in universities, as well as training provided by academic institutions actively engaged in responsible AI research.

For the active research on corporate social responsibility, this study demonstrates the significant influence of the labor market on a firm's commitment to responsible practices, as noted by previous research (Bode et al., 2015; Portocarrera and Burbano, 2023). However, our contribution hinges on highlighting how corporate decisions on social responsibility practices are deeply embedded in the context of business operations and market. Critical factors shaping the connections between employees' preference and firms' adoption of socially responsible practices include the importance of these employees to a firm's core business, the ease with which they can be replaced (labor market competitiveness), and the extent to which these employees advocate for specific social responsibility practices. Therefore, a comprehensive understanding of corporate social responsibility, often examined within the nonmarket strategy domain, should be grounded in a deep comprehension of a firm's operational dynamics and the prevailing market conditions. This approach aligns with the call for an "integrated strategy," as advocated by (Baron, 1995).

The emerging field addressing the social responsibility of AI technologies has predominantly focused on external activists' attempts to encourage firms to adopt responsible AI principles, often resulting in resistance from corporations. Our research directs attention inward, within the firms themselves. Specifically, we highlight the role of certain key employees, such as AI scientists directly engaged in technology development, in influencing a firm's decision-making regarding social responsibilities. This presents a promising avenue for expanding responsible AI practices. However, it is worth noting that not all personnel involved in AI technologies seem to possess this influential role. For instance, inventors responsible for developing AI patents do not appear to significantly impact a firm's likelihood of committing to

responsible AI principles. Therefore, our conclusion is not solely based on the personnel involved in AI technology development being the secret source for corporate commitment to responsible AI. Instead, it emphasizes the importance of these individuals themselves actively advocating for responsible AI and recognizing their capacity to wield influence within the organization for such outcomes to materialize.

REFERENCE

Agarwal, R., & Ohyama, A. 2013. Industry or Academia, Basic or Applied? Career Choices and Earnings Trajectories of Scientists. ***Management Science***, 59(4): 950–970.

Agrawal, A., Gans, J., & Goldfarb, A. 2019. ***The economics of artificial intelligence: an agenda***. University of Chicago Press.

Ahmed, N. 2022. Competitive Scientific Labor Market and Firm-level Appropriation Strategy in Artificial Intelligence Research. ***MIT Sloan Working Paper***.

Ahmed, N., Das, A., Martin, K., & Banerjee, K. 2024. The Narrow Breadth and Depth of Corporate Responsible AI. ***MIT Sloan Working Paper***.

Ahmed, N., Wahed, M., & Thompson, N. C. 2023. The growing influence of industry in AI research. ***Science***, 379(6635): 884–886.

Alekseeva, L., Azar, J., Gine, M., Samila, S., & Taska, B. 2019. The Demand for AI Skills in the Labor Market. ***SSRN Working Paper***, 1–38.

Ali, S. J., Christin, A., Smart, A., & Katila, R. 2023. Walking the Walk of AI Ethics: Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs. ***Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency***, 217–226.

Allen, M. 2020. Palantir CEO reflects on work with ICE. ***Axios***.

Arora, A., Belenzon, S., & Patacconi, A. 2018. The decline of science in corporate R&D. ***Strategic Management Journal***, 39(1): 3–32.

Baron, D. P. 1995. The Nonmarket Strategy System. ***Sloan Management Review***, 37(1): 73–85.

Bergen, M., & Wagner, K. 2015. Welcome to the AI Conspiracy: The "Canadian Mafia" Behind Tech's Latest Craze. ***The Vox***. https://www.vox.com/2015/7/15/11614684/ai-conspiracy-the-scientists-behind-deep-learning.

Bode, C., Singh, J., & Rogan, M. 2015. Corporate social initiatives and employee retention. ***Organization Science***, 26(6): 1702–1720.

Brayne, S. 2017. Big Data Surveillance: The Case of Policing. ***American Sociological Review***, 82(5): 977–1008.

Brown, J. A., Gianiodis, P. T., & Santoro, M. D. 2015. Following doctors' orders: Organizational change as a response to human capital bargaining power. ***Organization Science***, 26(5): 1284–1300.

Buolamwini, J., & Gebru, T. 2018. Gender Shades: Intersectional accuracy disparities in

commercial gender classification. ***Conference on fairness, accountability and transparency***, 77–91. PMLR.
Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. C. 2021. Difference-in-differences with a continuous treatment. ***ArXiv Preprint ArXiv:2107.02637***.
Candelon, F., di Carlo, R. C., De Bondt, M., & Evgeniou, T. 2021. AI Regulation Is Coming: How to prepare for the inevitable. ***Harvard Business Review***, 99(5).
Chen, Y. 2019. How evolutionary selection can train more capable self-driving cars. ***Deepmind***.
Collina, L., Sayyadi, M., & Provitera, M. 2023. Critical Issues About AI Accountability Answered. ***California Management Review Insights***.
Cowgill, B., Dell'Acqua, F., & Matz, S. 2020. The managerial effects of algorithmic fairness activism. ***AEA Papers and Proceedings***, 110: 85–90. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.
Criddle, C., & Murgia, M. 2023. Big tech companies cut AI ethics staff, raising safety concerns. ***Financial Times***.
Fleming, L., Greene, H., Li, G., Marx, M., & Yao, D. 2019. Government-funded research increasingly fuels innovation. ***Science***, 364(6446): 1139–1141.
Gofman, M., & Jin, Z. 2022. Artificial Intelligence, Human Capital, and Innovation. ***Journal of Finance***. https://doi.org/10.2139/ssrn.3449440.
Hannigan, T., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V., et al. 2019. Topic Modeling in Management Research: Rendering New Theory From Textual Data. ***Academy of Management Annals***, 13(2): 586–632.
Hartmann, P., & Henkel, J. 2020. The Rise of Corporate Science in AI : Data as a Strategic Resource. ***Academy of Management Discoveries***.
Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., et al. 2017. Deep learning scaling is predictable, empirically. ***ArXiv Preprint ArXiv:1712.00409***.
Hooker, S. 2020. The Hardware Lottery. ***ArXiv Preprint***, 63–88.
Jobin, A., Ienca, M., & Vayena, E. 2019. The global landscape of AI ethics guidelines. ***Nature Machine Intelligence***, 1(9): 389–399.
Kim, J., & 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.
King, G., & Zeng, L. 2001. Logistic regression in rare events data. ***Political Analysis***, 9(2): 137–163.
Lahoti, S. 2019. Amazon joins NSF in funding research exploring fairness in AI amidst public outcry over big tech #ethicswashing. ***Pakthub***. https://hub.packtpub.com/amazon-joins-nsf-funding-fairness-ai-public-outcry-big-tech-ethicswashing/.
LeCun, Y., Bengio, Y., & Hinton, G. 2015. Deep learning. ***Nature***, 521(7553): 436–444.
Lowensohn, J. 2015. Uber Gutted Carnegie Mellon's top robotics lab to build self-driving cars. ***The Verge May***, 19: 2015.
Mailath, G. J., & Postlewaite, A. 1990. Workers versus firms: Bargaining over a firm's value. ***The Review of Economic Studies***, 57(3): 369–380.
Martineau, K. 2019. What a little more computing power can do. ***MIT News***. http://news.mit.edu/2019/what-extra-computing-power-can-do-0916.
Merton, R. 1973. ***The sociology of science: Theoretical and empirical investigations***.
Metz, C. 2017. Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent. ***The New York Times***. https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-

salaries.html.
Metz, C. 2018, April 19. A.I. Researchers Are Making More Than $1 Million, Even at a Nonprofit. ***The New York Times***.
Metz, C., & Isaac, M. 2023. Meta, Long an A.I. Leader, Tries Not to Be Left Out of the Boom. ***The New York Times***. https://www.nytimes.com/2023/02/07/technology/meta-artificial-intelligence-chatgpt.html.
Mindruta, D. 2013. Value creation in university-firm research collaborations: A matching approach. ***Strategic Management Journal***, 34(6): 644–665.
Miric, M., Jia, N., & Huang, K. G. 2023. Using supervised machine learning for large-scale classification in management research: The case for identifying artificial intelligence patents. ***Strategic Management Journal***, 44(2): 491–519.
Molloy, J. C., & Barney, J. B. 2015. Who captures the value created with human capital? A market-based view. ***Academy of Management Perspectives***, 29(3): 309–325.
Parloff, R. 2016, September 28. Why Deep Learning Is Suddenly Changing Your Life. ***Fortune***. https://web.archive.org/web/20220226165354/https://fortune.com/longform/ai-artificial-intelligence-deep-machine-learning/.
Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., et al. 2019. Meaningful explanations of black box AI decision systems. ***Proceedings of the AAAI conference on artificial intelligence***, 33(01): 9780–9784.
Phelan, D. 2023. Apple's Tim Cook Talks About AI, Apps, Vision Pro And iPhone Gaming. ***Forbes***.
Portocarrero, F. F., & Burbano, V. C. 2023. The effects of a short-term corporate social impact activity on employee turnover: Field experimental evidence. ***Management Science***.
Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., et al. 2015. ImageNet Large Scale Visual Recognition Challenge. ***International Journal of Computer Vision***, 115(3): 211–252.
Shokri, R., & Shmatikov, V. 2015. Privacy-preserving deep learning. ***Proceedings of the 22nd ACM SIGSAC conference on computer and communications security***, 1310–1321. ACM.
The Economist. 2016a, April. Million-dollar babies: As Silicon Valley fights for talent, universities struggle to hold on to their stars. ***The Economist***. https://www.economist.com/news/business/21695908-silicon-valley-fights-talent-universities-struggle-hold-their.
The Economist. 2016b, June. From not working to neural networking. ***The Economist***. https://www.economist.com/news/special-report/21700756-artificial-intelligence-boom-based-old-idea-modern-twist-not.
Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. 2020. ***The Computational Limits of Deep Learning***. http://arxiv.org/abs/2007.05558.
Tong, S., Jia, N., Luo, X., & Fang, Z. 2021. The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. ***Strategic Management Journal***, 42(9): 1600–1631.
Vincent, J. 2018. Google promises ethical principles to guide development of military AI. ***The Verge***.
Wakabayashi, D., & Scott Shane. 2018. Google Will Not Renew Pentagon Contract That Upset Employees. ***The New York Times***.

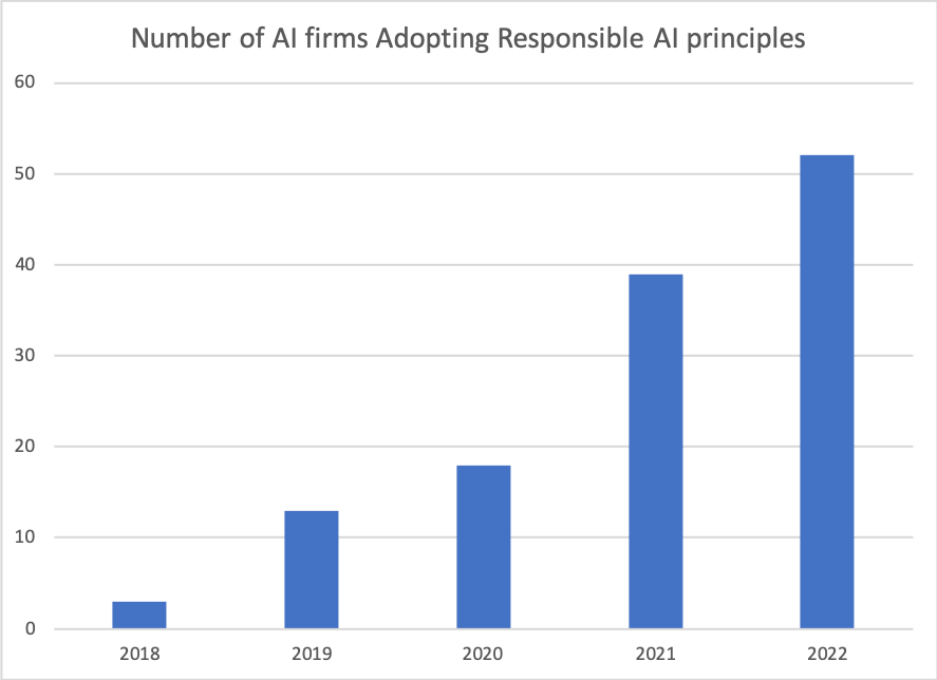

Figure 1: The number of AI firms adopting responsible AI principle is growing over time

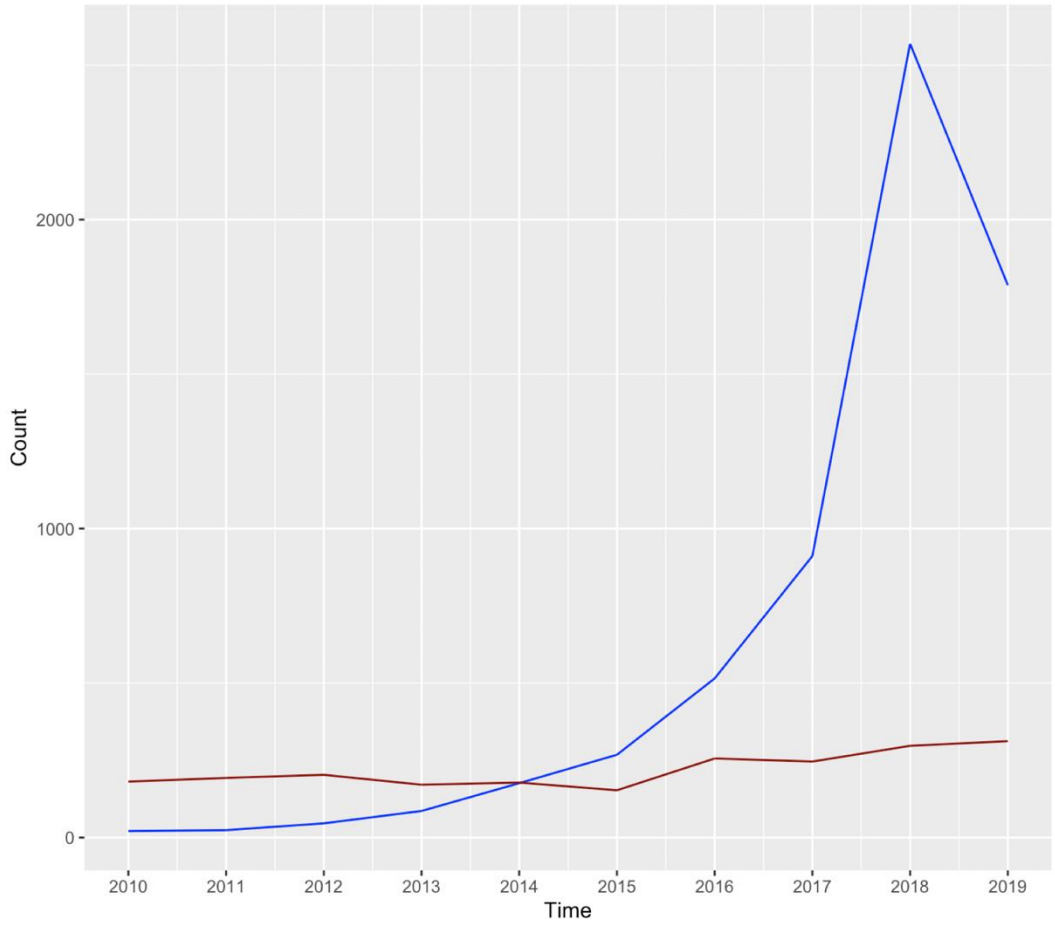

Figure 2: The disparity between the demand for deep learning researchers vs the supply of AI scientists in the US. Here blue line indicates the demand for deep learning talent (Burning Glass Technologies job posts data, the total number of deep learning job posts with PhD requirements) and the red line indicates the supply of AI talent (US data, CRA survey data, annual count of AI specialized PhD)

TABLE 1: Top Keywords from Topic modeling results [Responsible AI principle full text]

5 Topics: top keywords

'Topic 1: ai, technologies, human, principles, utilization, nec, customers, people, development, use',
'Topic 2: ai, group, human, data, use, products, principles, society, services, ethical',
'Topic 3: ai, data, systems, use, shall, governance, system, privacy, policy, development',
'Topic 4: ai, data, systems, use, bias, principles, human, development, privacy, people',
'Topic 5: must, ai, data, systems, rights, transparency, insights, regulations, humans, firms'

10 Topics: top keywords

'Topic 1: ai, technologies, ethical, said, systems, principles, quinn, ethics, human, scientific',
'Topic 2: ai, bmw, group, data, applications, use, principles, human, technologies, intelligence',
'Topic 3: ai, data, relx, system, systems, position, development, legal, information, may',
'Topic 4: ai, data, bias, fairness, systems, use, ml, human, model, ensure',

'Topic 5: ai, data, systems, bias, human, use, development, system, technology, people',
'Topic 6: ai, group, oki, etc, products, business, human, issues, principles, technologies',
'Topic 7: ai, systems, data, use, system, human, people, shall, used, applications',
'Topic 8: ai, lg, sony, products, technologies, services, diversity, customers, work, create',
'Topic 9: ai, data, use, principles, systems, human, privacy, development, society, technology',
'Topic 10: ai, systems, data, ethical, system, use, development, human, bias, ethics'

Table 2: summary statistics

	Obs	Mean	SD	Min	Max
Variable Name					
Demand for AI Scientists	1828	0.648	2.03	0	17.124
Demand for Deep Learning Scientists	1828	0.32	1.31	0	13.59
Demand for Non-Deep Learning AI Scientists	1828	0.525	1.76	0	16.08
Firm-University Research Collaboration	1828	1.411	3.28	0	20.575
Firm Responsible AI Recruitment	1828	0.35	3.26	0	86.01
ResAI	1828	0.0465	0.21	0	1
AI Research stock	1828	1.057	2.73	0	19.746
AI Patents stock	1828	0.741	1.34	0	13.31
Non AI Patents stock	1828	4.295	5.96	0	27.183
Demand for Non-AI Scientist	1828	1.53	3.52	0	24.265
Firm Size	546	3.41	2.16	0.001	10.066
Firm Revenue	546	11.232	4.19	0.001	25.072
R&D Spending	496	7.74	3.65	0	21.66

Table 2. Results of logistic regression with the adoption of responsible AI principles as dependent variable

DV: Adoption of Responsible AI principles	Model 1	Model 2	Model 3	Model 4
Demand for AI Scientists (t-1)		0.202***		
		(0.004)		
Demand for Deep Learning Scientists (t-1)			0.197**	
			(0.016)	
Demand for Non-Deep Learning AI Scientists (t-1)				0.060
				(0.576)
Control variables				
AI Research stock (t-1)	0.265***			
	(0.001)			
AI Patents stock (t-1)	-0.005	-0.003	-0.014	-0.056
	(0.700)	(0.981)	(0.913)	(0.128)
Non AI Patents stock (t-1)	-0.015	0.031	0.019	0.022
	(0.785)	(0.553)	(0709	(0.665)
Demand for Non-AI Scientist (t-1)	0.133**		0.028	0.126*
	(0.022)		(0.708)	(0.083)
Firm Size (t-1)	0.275	0.605***	0.455**	0.526***
	(0.217)	(0.005)	(0.034)	(0.016)
Firm Revenue (t-1)	-0.011	-0.119	-0.067	-0.130
	(0.960)	(0.558)	(0.735)	(0.607)
R&D Spending (t-1)	0.115	0.256	0.262	0.283
	(0.568)	(0.187)	(0.170)	(0.143)
Constant	-20.741	-20.778	-20.912	-20.873
	(1.000)	(1.000)	(1.000)	(1.000)
Industry FE	Yes	Yes	Yes	Yes
Observations	496	496	496	496
Log Likelihood	-83.626	-90.065	-87.245	-90.056
Note: *p<0.1, **p<0.05,***p<0.01; p-values are reported in the parentheses				

Table 3. Results of logistic regression with the adoption of responsible AI principles as dependent variable

DV: Adoption of Responsible AI principles	Model 1	Model 2	Model 3	Model 4
Firm-University Research Collaboration(t-1)	0.247***			
	(0.000)			
Firm-University Research Collaboration AI (t-1)		0.253***		
		(0.007)		
Firm-University Research Collaboration DL (t-1)			0.288**	
			(0.016)	
Firm Responsible AI Recruitment (t-1)				0.173*
				(0.083)
Control variables				
AI Patents stock (t-1)	0.053	-0.058	-0.062	-0.076
	(0.694	(0.680)	(0.659)	(0.559)
Non AI Patents stock (t-1)	-0.051	-0.006	0.005	0.024
	(0.359)	(0.910)	(0.919)	(0.631)
Demand for Non-AI Scientist (t-1)	0.104*	0.135*	0.138**	0.128*
	(0.078)	(0.018)	(0.014)	(0.021)
Firm Size (t-1)	0.275	0.437**	0.461**	0.526***
	(0.110)	(0.047)	(0.036)	(0.013)
Firm Revenue (t-1)	-0.013	-0.051	-0.048	-0.058
	(0.949)	(0.803)	(0.816)	(0.770)
R&D Spending (t-1)	0.115	0.168	0.190	0.226
	(0.568)	(0.394)	(0.331)	(0.234)
Constant	-20.741	-20.741	-20.791	-20.828
	(1.000)	(1.000)	(1.000)	(1.000)
Industry FE	Yes	Yes	Yes	Yes
Observations	496	496	496	496
Log Likelihood	-82.915	-86.273	-87.076	-87.757
Note: *p<0.1, **p<0.05,***p<0.01; p-values are reported in the parentheses				

Appendix A

List of words 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’

Appendix B

This list of keywords is based on (Alekseeva et al., 2019) 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”