# Algorithmic Bilinguals

Prasanna (Sonny) Tambe*
Wharton School, U. Pennsylvania

February 13, 2024

Draft copy. Results are preliminary and subject to change.

**Abstract**

Using workforce data from US firms, this study tests the hypothesis that generating value from algorithms requires employing domain experts who can effectively interact with data and algorithms. This decentralization of technical expertise stands in contrast to earlier generations of business technologies for which the complementary skills were primarily embodied in IT specialists and it is due to the task complementarities that arise when integrating decision-making algorithms into production. Using two different workforce data sets, I show that 1) employers have been shifting hiring towards requiring greater expertise with algorithms from domain experts, 2) technical human capital in frontier firms has become more dispersed across occupations, and 3) the market assigns higher value to firms' algorithmic investments when they have also made these workforce changes, indicating the presence of valuable intangible assets that can yield future productivity benefits. Finally, I show that the recent advance of no-code and natural language tools that make it easier to perform technical work accelerates these changes. Implications for training, education, and algorithmic decision-making are discussed.

**Keywords**: human capital, algorithms, IT intangibles, future of work, IT complements, digital literacy, general purpose technology, GPT

## 1 Introduction

The impact of algorithmic decision-making is having on organizations is a topic of growing interest (Rock, 2019; Wu et al., 2019; Agrawal et al., 2018; Zolas et al., 2021). Much research in this area focuses on the labor reallocation effects of AI and automation technologies (Acemoğlu and Restrepo, 2016; Autor and Salomons, 2018; Brynjolfsson et al., 2018; Eloundou et al., 2023), with some of this work demonstrating that these technologies are not simply labor displacing (Agrawal et al., 2019; Gregory et al., 2022). Rather, new technologies are also likely to generate new jobs and new

*Correspondence: tambe@wharton.upenn.edu. Comments are very much appreciated. I am grateful for helpful conference discussions by Christopher Stanton and Anna Salomons and for valuable conversations with Matt Beane and Daniel Rock as well as participants at the INFORMS Conference on Information Systems and Technology, the Harvard Business School Digital Initiative Discussion and Symposium, the NBER Summer Institute on Digital Economics and Artificial Intelligence, and the Temple AIML conference. I am also grateful to Zhiwei (Berry) Wang for valuable research assistance, to Ben Zweig, Patrick Julius, and Lisa Simon for providing access to and assistance with the Revelio data, to Dan Restuccia and Bledi Taska for providing access to the Lightcast (Burning Glass) data, and to the Wharton Mack Institute for providing financial assistance.

types of jobs (Bessen, 2019; Autor et al., 2022), and scholars have explored how humans can be most effective when working alongside algorithms (Faraj et al., 2018; Cowgill, 2018; Beane, 2019; Agrawal et al., 2019; Lebovitz et al., 2022; Babina et al., 2022).

This paper develops new theory and evidence arguing that a complement to the effective organizational use of algorithms is the bundling of algorithmic literacy into jobs that require significant levels of domain expertise. The argument focuses on two types of skills. *Algorithmic literacy* is that related to the use of algorithmic tools and technologies, such as data science, machine learning, and AI tools, which enable firms to convert data into strategic decisions in the pursuit of business goals.[1] Prior work suggests that both domain expertise and "interactional" technical expertise may be important when placing predictive technologies into production, particularly in sensitive contexts like law or medicine, where the payoff function for a decision is difficult to define or where the tolerance for machine-based prediction error is low (Kleinberg et al., 2018; Choudhury et al., 2020). *Domain expertise* is the knowledge related to a specialized field such as nursing, sales, marketing, or accounting. Domain expertise has become a central construct in the literature on data science tool development and use (White et al., 2009; Mao et al., 2019; Sambasivan and Veeraraghavan, 2022; Bayer et al., 2022; Kumar and Sharma, 2022). The need to bundle these skills together is the central focus of the paper, and it requires "algorithmic bilinguals" to fill these positions (Collins, 2004).

This migration of interactional technical expertise into domain expert jobs stands in contrast to a design in which technical expertise is centralized in specialized information technology (IT) workers and it relates to the observation that much of the adaptation in work to technological change occurs at the sub-occupational level (Spitz-Oener, 2006). To explain these changes, this paper develops new hypotheses that build on the literature related to job design (Smith, 1776; Becker and Murphy, 1992; Dessein and Santos, 2006; Teodoridis, 2017; Lindbeck and Snower, 2000; Postrel, 2002). It considers how algorithms differ from prior business information technologies and generates hypotheses related to i) how firms make workforce adjustments when using algorithms for decision-making and ii) how these adjustments may impact organizational performance.

These hypotheses are then tested using databases on corporate hiring and employment. The first database captures a "near-universe" of job listings issued by US firms and has been used in prior work on the changing skill requirements of jobs (Deming and Kahn, 2018; Acemoglu et al., 2022) as well as to track the spread of new technologies (Goldfarb et al., 2023). The second data source is a fourteen-year panel of how workers with different technology skills move across occupations in different firms over time.[2] These databases are combined with administrative data on the knowledge content of occupations from the Bureau of Labor Statistics O*NET database and with employers' financial data from the Compustat-Capital IQ database.

Empirical analysis of these data sources produces three principal findings. First, using the job

[1] I use the term "literacy". Whether jobs require literacy or expertise with the technology is an important question but it is beyond the scope of the data sources used in this paper so it is left to future work.

[2] We provide details on this data source in a later section as well as an Appendix that conducts comparisons with data collected by administrative agencies.

listings data, I demonstrate that algorithmic skills gradually spread across listings from 2010 to 2016 in an occupational pattern that more closely resembles office software (e.g. word processing tools) than more centralized technology skills like database administration. These technological markers became notably more dispersed across occupations than skills related to other business technologies and by 2016, only one-third of these technologies were embedded in IT listings. Indicators of algorithmic literacy were particularly likely to migrate into occupations requiring domain expertise.

These findings illustrate changes in employer preferences, but not whether the market can meet these preferences, so I next show that these trends in the job listings data are consistent with changes in the corporate employment data over an overlapping but longer fourteen year sample period (2008-2022). I find that in public firms, markers of algorithmic literacy increasingly spread to business, management, and financial occupations which is consistent with the job listing data and is in contrast to other business technologies which have not experienced similar changes during the same time period. For example, the fraction of workers in business occupations with skills related to mobile technologies remained flat in this period and this fraction fell for skills related to cloud and network based technologies.

Third, I show that these workforce adjustments are generating productive, intangible assets for public firms. The spread of algorithmic literacy into business occupations is most clear for high market value firms. Financial markets assign higher value to public firms with investments in algorithms when they make complementary workforce adjustments which suggests that firms derive the largest benefits from investing in algorithmic tools when their business professionals are engaged in integrating these technologies into production. Similar correlations are not present for investment in other information technologies or with other types of employee expertise, suggesting that the returns to bundling skills in this way are specific to algorithmic tools. Fourth, to move beyond primarily descriptive evidence, I leverage discrete technological advancements in software and tools that lower the cost of acquiring algorithmic skills to show that these changes increase the propensity of employers to make these workforce changes.

This study contributes to two academic literatures. First, with its focus on employers, it contributes to a literature identifying organizational complements to information technology (Bresnahan et al., 2002; Black and Lynch, 2001; Caroli and Van Reenen, 2001; Bartel et al., 2007; Bloom et al., 2012). These analyses have principally been rooted in a perspective based in IT as a technology that can automate "routine" tasks, but the application of algorithmic technologies to contexts where decision rules are not easily mapped to software has reopened the discussion on how IT might affect firms' labor force needs (Brynjolfsson et al., 2018). In doing so, this paper contributes to an emerging literature that examines management practices that complement investments in predictive algorithms (Brynjolfsson et al., 2021; Zolas et al., 2021; Dixon et al., 2021; Xue et al., 2022).

Second, it contributes to a literature on how the widespread use of algorithms might shape the future of work, which is an increasingly important area of research as new technologies subsume many of the tasks done by humans while simultaneously generating new areas of demand for human labor (Agrawal et al., 2019). Most prior work focuses on the IT workforce (Ang et al., 2002;

Levina and Xin, 2007; Mithas and Krishnan, 2008; Wiesche et al., 2019; Tambe et al., 2020), but there has been limited work on the implications of technical skills for broader workforce outcomes (Atasoy et al. (2016); Deming and Noray (2020) are exceptions). The absence of work in this area is important given the growing demand from students and workers from all disciplines for "coding" and other technical skills and the growing emphasis on how humans and AI will interact in the workforce. Contemporaneously, the costs of doing technical work continue to fall, as new advances from tech companies (e.g. no-code tools for data manipulation or natural language tools like Code Interpreter/ChatGPT from OpenAI) make it cheaper and faster to do data science with little or no coding requirement. These findings, therefore, contribute to our understanding of how the connection between humans and algorithms will shape the demand for skills as employers continue to embrace algorithmic decision-making.

# 2 Theory and Hypothesis Development

## 2.1 Algorithmic technologies and domain expertise

For decades, the diffusion of computing technologies has driven a relative increase in the demand for college-educated workers (Berman et al., 1994; Bresnahan et al., 2002; Bartel et al., 2007). Computers complement educated workers because by automating routine tasks, they raise the productivity of front-line workers who can make decisions about non-routine problems (Autor et al., 2003). For example, when decision authority is decentralized, hotel desk agents can rapidly adapt to changing customer preferences and factory floor workers can fix manufacturing problems as they arise. This literature has principally focused on changes to education levels or to the mix of occupations that firms employ, but employers can also adjust the mix of skills *within* occupations in response to technological change (Spitz-Oener, 2006). Lindbeck and Snower (2000) argue that task complementarities in knowledge-rich jobs have shifted work away from specialization towards "holistic" work in which workers handle a diverse array of tasks. Multi-task work raises productivity in the presence of informational complementarities among tasks because productivity in one task can be interdependent with levels of activity in other tasks. In such contexts, bundling tasks to avoid the need for continuous coordination among workers can yield productivity benefits (Postrel, 2002). An advantage of hiring educated workers is that they can easily adapt to a multi-task job design.[3]

AI and data science algorithms challenge our understanding of technology and job design because these technologies can make decisions even when the relationship between inputs and outputs may not be viewed as routine (Brynjolfsson et al., 2018; Agrawal et al., 2018). When decision rules cannot be clearly articulated, there may be additional costs to separating technical and domain expertise, when building model as well as when interpreting their output. For instance, in complex

[3] Relative to changes in occupational demand, this intra-occupational change has been empirically less widely documented because administrative data agencies do not capture it as well. To fill this gap, scholars often turn towards alternative data sources. An example is (Spitz-Oener, 2006), who uses German data to show that within-occupational change was happening particularly quickly in occupations that were being computerized. In that sample, within-occupational change accounted for 36% of educational upgrading.

domains, like medicine or engineering, making sense of data can require medical expertise needed to create meaningful healthcare indicators from medical test results. This is not a new challenge. Organizations that rely on data-mining processes, have routinely been faced with the challenge of how to inject domain expertise into the data modeling process. For example, one standardized process widely used to balance data modeling decisions with business objectives is "CRISP-DM" (Cross Industry Standard Process for Data Mining) (Wirth and Hipp, 2000; Chapman et al., 2000).[4] The CRISP-DM model separates the data mining process into six discrete steps: 1) Business Understanding, 2) Data Understanding, 3) Data Preparation, 4) Modeling, 5) Evaluation, and 6) Deployment. Within this model, domain expertise is viewed as separate from technical expertise and conceptualized as drawn from other experts within the organization or from outside clients. However, coordination between workers with different expertise is costly, and studies of CRISP-DM identify coordination costs across stakeholders as a weakness of this paradigm (Saltz, 2021).

These challenges are not limited to CRISP-DM. A recent literature extends these challenges to different stages of the data science process (Mao et al., 2019; Choudhury et al., 2020; Park et al., 2021). The inherently iterative nature of data exploration, experimentation, and learning required when doing data science often favors generalists, who have a diversity of relevant skills, rather than specialists (Colson, 2019). A prominent example of this tension is with "data scientists" themselves, who by definition of the job title, combine technical and statistical skills with domain expertise (Davenport and Patil, 2012; Provost and Fawcett, 2013). The importance of domain expertise for data science has been discussed online[5], in industry panels[6], and increasingly, in the business press (Oostendorp, 2019). Even beyond data scientists, "unicorns", who couple domain expertise with technical skills, are becoming important to many algorithmic decision-making contexts (Jha and Topol, 2016).[7] For instance, users of machine learning tools in high-stakes contexts must evaluate the tradeoffs required when choosing which data to include in a model, how to construct model features, or how to assign value to the costs of different prediction errors (Kleinberg et al., 2018; Cowgill, 2018; Cowgill et al., 2020). Research on pharmaceutical industries has indicated the importance of embedding the relevant human capital in downstream occupations to achieve successful innovation outcomes (Wu et al., 2019), and in healthcare, Lebovitz et al. (2022) describes the challenges arising with interpreting the accuracy of machine learning tools and Jha and Topol (2016) argues the

[4]Poll results from 2014 suggest that it is the most common methodology used for data mining and data science projects, with about half of the respondents reporting using CRISP-DM and the other half divided over other methods. See https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html, last visited on Jan 4 2023.

[5]For example, see Is domain knowledge necessary for a data scientist? Accessed on March 11, 2019.

[6]A video of one such industry panel is captured here: `https://youtu.be/qKcUsIqoSHE`.

[7]The educational community has also started to respond to these changes. For instance, the notion that data-driven employers increasingly demand "bilingual" workers (i.e. individuals who have both technical skills and domain expertise) was underscored by an announcement from MIT on their investment in a new College for Artificial Intelligence. The goal of the college, said L. Rafael Reif, the president of M.I.T., is to "educate the bilinguals of the future." He defines bilinguals as people in fields like biology, chemistry, politics, history and linguistics who are also skilled in the techniques of modern computing that can be applied to them. Additionally, it is expected that the "bilingual" graduates who emerge from this new College — combining competence in computing and in other fields — will be of enormous value to employers. *New York Times*, Oct 15, 2018. MIT Plans College for Artificial Intelligence, Backed by $1 Billion.

importance of medical experts acquiring the skills required to understand predictive model output. The first hypothesis is that these forces, which are greater for algorithmic technologies, are pushing organizations to favor bundling of algorithmic skills in jobs outside the IT workforce.

***H1****: Algorithmic skills are more likely than other business information technologies to be bundled with domain expertise.*

## 2.2 Which workers receive the bundle?

This leaves open the question of which occupations receive this bundle of skills. The literature argues that tasks are organized into jobs according to three factors: specialization, coordination, and adaptation, so how AI affects job roles depends on how it affects these factors. Specialization allows for Smithian productivity gains as in the medical field where AI algorithms analyze medical images, freeing radiologists to concentrate on complex cases (Smith, 1776). In terms of coordination, AI enhances the synchronization of interdependent tasks, exemplified by supply chain management where predictive analytics efficiently align inventory, supply, and delivery schedules (Becker and Murphy, 1992). A third factor is adaptation, with AI excelling in tailoring tasks to local information (Dessein and Santos, 2006). Marketing professionals, for instance, might leverage AI to create personalized advertising based on consumer behavior data. AI's adaptive capabilities may be particularly valuable for tasks where the application of local knowledge is critical to the decision.[8]

***H2****: Algorithmic skills are more likely to appear in occupations where local adaptation is important.*

These forces provide context in which to theorize about the control of information and task bundles in work environments. An instructive historical parallel is typing pools, where workers were once organized into pools to produce typing services.[9] The typing task eventually became part of the knowledge worker's job because local adaptation is important when creating documents. In the same way, if domain expertise helps with local adaptation and the costs of acquiring the relevant technical expertise falls, organizations may choose for domain experts – like those in finance and human resources – to receive these skills. On the other hand, AI introduces the potential for decentralized control. AI can substitute for some forms of domain expertise, reducing the need for domain expertise in areas such as foreign language proficiency, due to AI-powered translation tools. Here, the technology itself can provided the services once provided by domain experts. Therefore, who gets the bundle is an empirical question.

***H3****: Algorithmic skills are more dispersed across business occupations than other information technologies.*

[8] There are parallels for these arguments in the construction of teams. Using academic publication data, Teodoridis (2017) shows that a decrease in the cost of acquiring new technical knowledge changes the optimal mix of expertise when constructing diverse teams.

[9] I am grateful to Anna Salomons for suggesting this comparison.

These considerations are not static. The job considerations discussed above are closely related to the costs of acquiring technical knowledge, which are constantly changing. As the costs of some types of technical expertise falls, employers will bundle it with knowledge workers.

***H4***: *Algorithmic skills are more likely to be bundled in the jobs of domain experts as the cost of using algorithmic tools falls.*

### 2.3 Workforce transformation and market value

Similarly, for investments in tools like AI and data science, firms may not recognize value from these technologies unless they have the people in place who can apply these technologies to valuable decision contexts. Prior work has shown that workforce adjustments like these are needed to realize financial returns to investments in different technologies (Black and Lynch, 2001; Bresnahan et al., 2002; Caroli and Van Reenen, 2001; Bresnahan et al., 2002; Bartel et al., 2007; Bloom et al., 2012). For technologies that can perform routine tasks, allocating decision authority to front-line decision makers yields higher productivity levels (Bresnahan et al., 2002), making IT investments valuable in turbulent environments where the value of decisions depends on rapidly changing external conditions (Mendelson and Pillai, 1998; Pavlou and El Sawy, 2006; Tambe et al., 2012; Black and Lynch, 2001; Bresnahan et al., 2002).

Similarly, firms invest in algorithms may also need to design jobs to colocate algorithmic skills with knowledge workers in order to realize greater value. These higher values reflect valuable intangible assets that the market expects to eventually yield a stream of productive benefits. The literature above suggests that the application of data science and AI in a production context, by introducing new challenges related to coordinating domain expertise with effective data modeling, analysis, and application, amplify the productivity benefits that arise when hiring employees that can synthesize both types of knowledge.

***H5***: *Financial markets assign higher value to algorithmic investment when algorithmic skills are dispersed among business-facing occupations.*

## 3 Data sources and key measure construction

### 3.1 Key data sources

#### 3.1.1 Job listings database

When employers have job openings, they post details online on their corporate web sites or on job boards. These listings identify the employer and the job title, the geographic location of the position, the skills and education sought from candidates, offered wages, and other fields relevant to the search process. I use this data to measure when specific skills begin to appear in online job ads and how skills co-occured in these listings with other skills.

Job listings data have been used in several papers on changing workforce skill requirements (Todd et al., 1995; Slaughter and Ang, 1996; Gallivan et al., 2002; Lee and Han, 2008). This study uses data from Lightcast, a labor market analytics firm that 1) uses software to crawl a "near-universe" of online job postings and 2) uses natural language processing to parse skills and other job information.[10] This data provider uses proprietary software to collect and standardize data from over 17,000 job boards and corporate web sites, and these data are processed to ensure that a job listing is not counted multiple times if an employer posts it several places on the web. The processed data include posting date, job location (metropolitan area), employer name, job title, educational requirements, certifications required for the position, and skill expectations for each job. A growing number of studies use this data source to study labor markets (Hershbein and Kahn, 2018; Deming and Kahn, 2018; Modestino et al., 2019), including how AI related skills spread across jobs and industries (Acemoglu et al., 2022; Goldfarb et al., 2023).

Lightcast associates each listing in the database with a BLS O*NET code and the employer in a listing is tagged with a North American Industry Classification Systems (NAICS) industry. Job openings list skills, such as *Python*, *Random Forest*, *Chemistry*, *Supply Chain*, *Accounting*, *Data Science*, *Teamwork*, or *Communication* which are standardized using a skill dictionary maintained by Lightcast. These skill data should not be interpreted as "requirements". Employers can omit skills from listings, some skills may be assumed rather than listed, and successful candidates may not need all of the skills in a listing. Nonetheless, employers are thoughtful about the skills they put in listings because including or omitting a skill can attract or repel the wrong type of applicant.

The data collection process raises questions about its coverage. Prior academic work has provided thorough information on the sampling properties of the data and compared it with administrative data sets, so I do not duplicate those comparisons here.[11] Key findings from these comparisons are that these job listing data over-represent in computer and mathematical occupations, as well as management, health care, business, and financial occupations, but they represent IT workers particularly well. They are a less robust indicator for job openings in blue-collar occupations.

### 3.1.2 Corporate employment database

The corporate employment data were collected through a partnership with the workforce intelligence company Revelio Labs.[12] Their databases are constructed from a variety of data sources including online career profiles and federal databases.[13] These data are similar in their informational content to that posted on online professional networks such as LinkedIn and they cover a large fraction of white-collar work in the US. The data cover both public and private US firms but the sample used in

[10] Until June of 2022, Lightcast was known as "Burning Glass Technologies" and is referred to as such in much of the prior work that has used this data set. In this paper, for consistency, we use the name Lightcast throughout, including when referencing the use of these data in prior papers.

[11] See, for example, Appendix A of Deming and Kahn (2018) who make comparisons of these Lightcast data with administratively collected data sources.

[12] See https://www.reveliolabs.com/

[13] Scholars have argued that the lack of firm-level data on workforce skills is a constraint for understanding how firms are adjusting to technological change (Frank et al., 2019; Raj and Seamans, 2018).

this study is limited to public firms so that they can be connected with financial market data. This data source has been less widely used in the literature than the job listings data, so in Appendix A, I report detailed comparisons of these data with administratively sampled workforce data from the Bureau of Labor Statistics. We can see from these comparisons that like the Lightcast data, these data are oversampled in management, business, and technology occupations and undersampled in areas such as agriculture and manufacturing which is consistent with the greater use of professional networking sites in knowledge-intensive occupations.[14]

These workforce data are used to generate measures of annual firm-occupation-skill employment activity from 2008 through 2021.[15] This panel records how skills, like "machine learning", related to different technologies diffuse across occupations and employers. Moreover, the data contain CUSIP identifier codes for employers, and so employers can be joined with external firm-level financial databases such as the Compustat-Capital IQ data (described below).

#### 3.1.3 Supplementary data sources

To create job expertise measures, the O*NET codes in the job listing data are connected with the Occupational Information Network (O*NET) content model published by the Bureau of Labor Statistics.[16] The O*NET database has been very widely used in academic research,[17] is government administered, collected by surveying occupational experts, and provides information on employment, wages, and the work content of US jobs. The O*NET taxonomy reports work requirements including the knowledge required for different occupations.[18] Finally, some analyses also use firms' financial data from Compustat-Capital IQ, which was collected through the WRDS data service.

### 3.2 Key measures

#### 3.2.1 Algorithmic and technological expertise

The key unit of analysis for the paper is worker skills. A challenge when analyzing large volumes of skills data is the development of classifications that can provide meaning to groups of skills.[19] This absence of standardized taxonomies on the skill content of jobs is reflected in the existing academic

[14]The provider notes some potential issues with the reporting of skills in the data. The profile data is combined from multiple sources which gather publicly available profiles, but around May 2021, user skills disappeared from the majority of public profiles. The provider imputes (predict) skills after that data, and notes whether the skill on a profiles is reported or imputed. However, I do not know the imputation algorithm, and so I limit the analysis sample to the years through 2021 only.

[15]The data provider notes in their documentation that the skill data are imputed, rather than collected, from 2022 onwards.

[16]See https://www.onetonline.org.

[17]Notable examples include (Autor et al., 2003).

[18]The O*NET data is periodically revised to reflect the changing structure of the US workforce. Although it was revised in 2019, I use the version from before this revision to match the O*NET codes in my version of the Lightcast data, which were based on the taxonomy before the O*NET revision took place.

[19]Indeed, precisely because of the growing interest in the "future of work", the construction of taxonomies that can make sense of emerging sources of skills data and inform career development pathways is an active and ongoing area of research among businesses and information agencies. For example, see recent efforts by Nesta in the UK or Lightcast.

literature, where there have been notable efforts to develop meaningful taxonomies around IT skills (Lee et al., 1995; Niederman et al., 2016). Empirical papers that study large quantities of archival, digitally collected skill data have used manually constructed mappings of skills to conceptual measures (Deming and Kahn, 2018). Even foundational papers in the economics literature in this area have required the authors to use their own discretion (or those of colleagues or experts) to identify which skills in a database are most relevant to their phenomenon of interest (Autor et al., 2003). The small number of papers in the emerging literature on the impact of AI technologies have also generated taxonomies based on their own judgment (Brynjolfsson et al., 2018).

This paper takes a simpler approach, which is to rely on the categorizations provided by the data providers. The data providers have engaged in extensive efforts to use data-mining to group skills together into different business technology areas, such as “data science”, “AI”, or “Big data”. We group data science and AI skills together and refer to them as “algorithmic” tools and we conduct comparisons between these technologies and other business technologies that do not fall into this category. We focus on data science and AI because, as discussed above, they are the key inputs to the recent wave of technologies that directly make decisions, and our theoretical arguments are based on the coordination costs that arise when these automated decision-makers are directly integrated into a production context. Appendix C details the specific skills fall into each of the focal categories used in this analysis.

Measures of algorithmic literacy indicate whether a job listing or employment profile have a skill that falls within our algorithms category, which includes AI and data science. The measure of algorithmic literacy is constructed as the fraction of employees in business-facing occupations who have technical skills in this category.

### 3.2.2 Domain expertise

Jobs are also encoded according to whether they require domain expertise, which as discussed above, is defined as “knowledge of a specific, specialised discipline or field”. The measurement of domain expertise in jobs is encoded consistently with algorithmic expertise. It is a binary measure where jobs are coded as requiring domain expertise if a domain-related skill appears in a listing. The list of potential domain-related skills jobs can require is extracted from the O*NET database, which curates a comprehensive list of all of the possible domains with which US-based jobs may require workers to know.[20] These potential knowledge domains are extracted from the “Knowledge” table in O*NET which delineates “organized sets of principles and facts applying in general domains.”[21] From the full list, *Computers and Electronics*, *Engineering and Technology*, *Telecommunications*, and *Mathematics* were removed because they potentially overlap with measures of algorithmic expertise.

[20]See `https://www.onetonline.org/find/descriptor/browse/Knowledge/`.

[21]The domain categories identified in the O*NET knowledge set are *Administration and Management, Biology, Building and Construction, Chemistry, Clerical, Communications and Media, Customer and Personal Service, Design, Economics and Accounting, Education and Training, English Language, Fine Arts, Food Production, Foreign Language, Geography, History and Archeology, Law and Government, Mechanical, Medicine and Dentistry, Personnel and Human Resources, Philosophy and Theology, Physics, Production and Processing, Psychology, Public Safety and Security, Sales and Marketing, Sociology and Anthropology, Therapy and Counseling, and Transportation.*

It is useful to contrast this approach with one in which jobs are identified as requiring domain expertise based solely on their titles. This approach would place the restriction on our analysis that jobs with the same title cannot differ in the knowledge they require. Relaxing this restriction is important for this analysis because it allows for an analysis of the diffusion of new skills into occupations (i.e. sub-occupational change) rather than changes to the occupational mix which is central to the arguments in this paper.

#### 3.2.3 Additional job characteristics

Beyond algorithmic and domain expertise, some analyses include indicators of skills related to *cognitive*, *social*, *character*, and *management* job attributes. The construction of these job attributes was based on prior work using the same data source and were constructed using the methods reported in that paper (Deming and Kahn, 2018). As with algorithmic and domain expertise, jobs are coded as requiring these attributes if the listing contains a related skill.[22]

#### 3.2.4 Employers' technology investments

Obtaining consistent, firm-level measures of IT investment that span multiple years has been a persistent challenge in the academic literature. IT investments are not consistently recorded on balance sheets, so scholars have had to leverage alternative data sources to create proxy measures, such as hardware investment measures collected by marketing firms through surveys, IT mentions recorded in legal filings, and more recently, investment into complementary IT skills. The rationale behind using labor expense to measure IT investment is that 1) it is a large component of a firm's technology investment and 2) it is growing because much frontier software is open source (economically free) and much of the hardware used is cloud-based and therefore not well measured by instruments that record the firm's owned servers and PCs.[23] However, installing frontier software requires technical expertise, so the wages paid to technical personnel or the quantities of technological workers employed by a firm may be a good proxy measure of a firms' technological investment, even if the software is free or hosted in the cloud. Studies using labor expense as a proxy measure of the firms' IT investment include aggregate numbers of IT workers employed by the firm or when measuring investment in specific technologies like machine learning, workers that have a specific technical skill (Lichtenberg, 1995; Brynjolfsson and Hitt, 1996; Tambe, 2014).

This paper takes the approach used in this literature. The employer workforce data described above allow construction of employer-year measures of the quantities of IT workers at a firm that

[22]Deming and Kahn (2018) construct these job attribute measures based on whether a listing has a skill related to the attribute. These skills, as reported in Table 1 of that paper, are: *cognitive* [problem solving, research, analytical, critical thinking, math, statistics], *social* [communication, teamwork, collaboration, negotiation, presentation], *character* [organized, detail oriented, multitasking, time management, meeting deadlines, energetic], and *management* [project management, supervisory, leadership, management (not project), mentoring, staff]. Deming and Kahn (2018) also include *writing*, *customer service*, *financial*, *computer*, and *software* job attributes in their analysis but those attribute families are not included in this analysis.

[23]These arguments likely hold particularly true for modern AI model expense which is increasingly based on an open-source stack and run in cloud data centers.

have skills in different technology categories. Measures of firms' technology investments, then, are operationalized as quantities of IT workers with technical skills in areas like data science, machine learning, and cloud technologies. These are interpreted as proxy measures of the intensity of firms' investments in each of these areas.[24]

#### 3.2.5 Financial variables, assets, and industry classification

From the Compustat-Capital IQ data, measures are constructed at the employer-year level of total market value, employment, industry classification, the value of PPE (property, plant, and equipment), and other assets. The use of Capital IQ financial data necessitates limiting the sample to public firms. Industry variables for these firms are retrieved at the three-digit NAICS (North American Industry Classification System) level. Total market value was computed as described in an existing literature relating intangible assets to firm value (e.g. see appendix describing variable construction in Brynjolfsson et al. (2002)). It is computed as the value of equity at the end of the fiscal year plus the value of preferred stock plus total debt which represents the total worth of a firm as assessed by the financial markets. Assets are computed as total assets minus PP&E.

## 4 Results

### 4.1 Model-Free Evidence

Figure 1b illustrates growth in the incidence of algorithmic skills appearing in listings within a three-year sample period spanning the years 2013 to 2016. Each x-axis tick in this figure is one month and the y-axis is the coefficient estimate ($\beta$) from the logit regression $ALG_i = \beta_t t_i + \epsilon_i$ where $t$ indexes months since January 2013 and $ALG$ indicates whether an algorithmic skill appears in a job listing. The series of estimates indicates that the likelihood that an algorithmic skill appears in a listing in this sample grows steadily in the earlier part of the sample before flattening out in the later part of the sample period.

Figure 2b shows the extent to which specific, common technical skills are bundled with domain expertise. Skills associated with higher values (further to right) are more often found in jobs that also list domain expertise in their requirements, and skills in dark blue are those that fall under the algorithmic category. This comparison shows that skills associated with the use of algorithms and data science are more commonly bundled with domain expertise. In this regard, these skills have more in common with skills like Excel and ERP systems that are commonly used by business-facing occupations. Figure 2a illustrates the extent to which skills associated with a broader class of categories appears in non-IT occupations. Higher values indicate skills more likely to appear in a broader variety of occupations. These figures indicate that algorithmic skills, those colored in dark blue, are less concentrated by occupation than other technical skills. Skills related to predictive analytics, data science, and data analysis are particularly dispersed and only slightly less

[24]Like most firm-level measures, this approach records the firms' investments with some degree of measurement error.

so then general skills related to the Microsoft Office Suite which are commonly used by workers across all knowledge-based occupations. This finding is consistent with the claim that employers are increasingly bundling algorithmic skills in occupations where domain expertise can be found.

Data on job openings are valuable because they (i) indicate employer preferences and (ii) can reflect immediate adjustments by employers and serve as a leading indicator of labor market changes. However, these data cannot tell us whether the posted listings indicate hard requirements or a "wish list" from employers, or whether the vacancies that require these skills are ever filled. Therefore, I turn to corporate workforce data to investigate whether changing employer preferences manifested in workforce changes.

Figure 3a illustrates how the dispersion of technical skills in business-facing occupations has been changing over time within these firms. In this figure, the y-axis is the intensity with which a skill appears in these occupations and all levels are depicted relative to their base rates in 2008. The trend line for AI and data science skills is shown in blue. These trend lines indicate that there has been steady growth in the rate at which AI and data science skills have penetrated these occupations. By 2021, skills related to these technologies appear in about 10% more occupations than thy did in 2010. Infrastructure technologies like investments in networks and the cloud became increasingly specialized. Fewer workers in business occupations needed the skills related to these technologies. The incidence of mobile skills in this worker sample remained flat in these occupations.

Figure 3b in the top right quadrant shows levels of this $ORGSKL$ measure across different industries. Unsurprisingly, it is highest in the Information and Professional Services industries, which include technology and finance firms. This is consistent with recent evidence on the prevalance of these technologies in these two industries (Lohr, 2024). Retail has climbed rapidly, perhaps reflecting the growing use of consumer data for prediction. Levels of this skill measure are lower in the Arts and Health industries although they have been climbing steadily in Healthcare reflecting the growing use of AI in healthcare domains.

Figure 3c in the bottom left quadrant depicts changes in the $ORGSKL$ measure from year-to-year, where firms are separated into quartiles according to their market values in 2021, the final year of the sample. This figure suggests that this figure is highest in higher value firms, and that it diverges in the first two-thirds of the sample, as might be expected, for instance, if workers with this combination of skills are a scarce employee resource that higher value firms can better attract. In the last few years of the sample, however, this figure appears to begin converging again across different quartiles, suggesting that supply of these workers may have been adjusting to demand.

The bottom right quadrant (Figure 3d), using data from the last year of the sample, plots firms' investment in AI technologies against the $ORGSKL$ measure. From this plot, we can see that firms tend to concurrently invest in both the workers that can install these technologies and business-facing workers with the complementary skills to apply these technologies to business domains. The largest circles, colored in blue, are those commonly referred to as "big-tech" firms.

### 4.2 Regression tests of changes in firms' hiring patterns

To test the proposed hypotheses, I investigate if domain expertise is more likely to accompany the use of algorithmic tasks in job listings. The general form of the logistic regression used to evaluate these correlations is the following:

$$DOM_i = \beta_{SKL} ALGSKL_i + \gamma_i + \epsilon_i \tag{1}$$

Figure 4a illustrates estimates ($\beta_{SKL}$) from regressions of individual data pipeline activities on domain skills to investigate where domain expertise might be most important when using data science tools. The coefficient estimates on analytics are almost as high as presentation and higher than decision making which is consistent with the statement that data science and analysis tools require domain expertise. By contrast, data management and modeling are negatively correlated with domain expertise so these tasks are confined to specialist jobs in the vacancies in the sample. The unit of observation $i$ in this model is the job listing. The dependent variable is a binary indication of whether a listing $i$ requires the applicant to have a form of domain expertise ($DOM$). The measures on the right-hand side indicate if the listing requires skill in any of the three technology categories. It also includes a vector of control variables ($\gamma$) that includes job title, industry, and a measure of the logged number of skills in the job ad.

Figure 4b reports the results of Equation 1 where different technical skills are placed on the right-hand side, and different job attributes, including domain expertise, are the dependent variable. In this figure, we observe positive correlations between algorithmic skill and domain expertise. Because these tests include job-title fixed effects, these correlations indicate that algorithmic technologies tend to be bundled in jobs requiring domain expertise in a way that has not occurred with other data technologies (**H1**).[25] We also see positive correlations between the use of algorithmic technologies and cognitive job attributes and negative correlations with management-related job attributes. It is reasonable that for high-level managers, and jobs that require a high-degree of personal-facing skills, it may be less critical for workers to be able to integrate data and decision-making.

### 4.3 Regression tests of complementarities with firms' human capital

Before discussing market value relationships, Table 2 reports correlations between firms' investments in algorithms and our main skill dispersion measure. Column (1) indicates that after controlling for size, assets, and aggregate IT investment, employers are coinvesting in algorithms along with business-facing workers who have algorithmic skills (**H3**). The correlations are negative for employment size and aggregate IT investment. The employment size figure, conditional on assets, likely reflects organizations with large front-line workforces (e.g. customer service, retail) which should lower the skill dispersion measure since it is taken over all of a firm's employees. Column (2) includes

[25] By contrast, correlations with database management tasks are negative, consistent with that task being specialized within IT work. The full form of these regressions can be found in Appendix B.

employer fixed-effects. Here, most of the correlations with the skill measure are absorbed, indicating that most of these measures reflect heterogeneity in employers. However, the main correlation of interest, between investment in algorithms and business-facing workers with algorithmic skills, persists, which means that firms in our sample are concurrently changing these inputs over the course of our panel.

Table 3 embeds workforce skill measures, along with technology investment measures into a regression framework that tests if the market assigns higher value to firms that concurrently invest in these two factors. Market value is a useful indicator for two reasons. First, firms need time to adjust new technologies to their production context, and most evidence suggests that firms are not yet realizing value from many of their investments in AI and analytics. Secondly, examining market value has the benefit that the value of workforce investments captured in rising market value can be interpreted as intangible assets, which are valuable to the firm and should have implications for future productivity. This approach is similar to that used in prior work to test whether workforce complements build new intangibles and raise the returns to broader investments in information technology (Brynjolfsson et al., 2002). The form of the regression is:

$$Log(MV)_{it} = Log(Assets)_{it} + TECH_{it} + ALGSKL_{it} + (TECH_{it} \times ALGSKL_{it}) + \gamma_{it} + \epsilon_{it} \qquad (2)$$

In this model, $i$ indexes the firm and $t$ indexes the year, $TECH_{it}$ is an indicator of investment in different technologies measured as described earlier, $ALGSKL_{it}$ is the measure of skill dispersion, and $\gamma_{it}$ is a vector of fixed-effects including year, industry at the three-digit NAICS level, and depending on the specification, employer fixed-effects. Columns (1) through (3) in Table 3 have industry and year fixed-effects but do not include employer fixed-effects. Column (1) is a baseline regression showing correlations between IT investment and market value. The coefficient estimate is similar to prior work that reports estimates from market value regressions. In column (2), adding a measure of algorithmic skills shows that firms that have invested in technical workers with data science and AI skills have a higher market value. The coefficient estimate on IT is no longer significant after adding this measure, which indicates that the $ALG$ measure separates out frontier firms for which technological investment is expected to generate capabilities that the market rewards.

Column (3) reports estimates from the full form of Equation 2, which includes the interaction terms between the technology and skill measures. These estimates indicate that the organizational skill measures are associated with higher levels of market value. Importantly, these workforce skills are necessary complements to investing in algorithms. Moreover, investments in these technologies themselves, without the supporting workforce engineering exhibit no significant correlations with market value (**H5**). This pattern of estimates suggests that firms that invest in algorithmic decision-making and are adjusting the skill content of their workforce in the way described in this paper are building valuable intangibles (t=1.90). Column (4) is the same specification, except that it includes employer fixed-effects, which should absorb any unobservable and time-invariant heterogeneity. Some of the positive coefficients on aggregate technical investment disappear, which suggests that part of the estimate on the IT investment measure reflects firm-level heterogeneity. However, the positive coefficient estimate on the interaction term for algorithmic technologies is robust to including firm fixed-effects. It falls only slightly in magnitude. The pattern of results in this table indicates that firms that invest in algorithms unlock the value of these investments when they are able to disperse the human capital related to algorithms throughout the firm. In a final set of comparisons, Figure 5 reports the estimate on the main interaction term fom Equation 2 where the sample is split into different size categories according to employment. This figure indicates that the market rewards that firms receive from concurrently investing in these two factors are highest for the largest firms in the sample.

Taken together, these analyses suggest that in the last decade, (i) employers adjusted hiring practices to attract domain experts with expertise in algorithms, (ii) human capital related to algorithmic tools spread to business-facing occupations, and (iii) employers that made these investments jointly with matching technological investments realized higher market values, indicating that the presence of valuable intangible assets in these firms. Together, these three pieces of evidence support the primary conclusion of the paper that greater level of technical skill in a firm's domain experts is a valuable complement to its use of algorithmic decision-making.

### 4.4 Robustness tests

The evidence suggests investments in data science complement algorithmic skills in business and management occupations. We can also conduct robustness tests that evaluate whether we find similar results when substituting measures based in different technologies or in different skills. Figure 6a shows the coefficient estimates on the interaction term on $(TECH_{it} \times ORGSKL_{it})$ from Equation 2 when using $TECH$ measures based in AI (which is the same as in our main regression), which we compare with measures based in infrastructure technologies like networks and databases. If workforce changes in business-facing occupations are particularly important for AI technologies, we should only observe meaningful correlations for this skill-based interaction term with AI investment, not with investments that firms make in other technologies. We do observe such a pattern in Figure 6a. Neither of the coefficients on the other interaction terms are significantly different than zero. This pattern of estimates supports the argument that a correlation between market value and the interaction term between AI skill and AI investment is not simply reflecting the type of heterogeneity that would be picked up by other measures of technology investment. In other words, the correlations we observe between value and the interaction between skills and AI investment are specific to AI, not general technology investment.

Figure 6b performs a similar comparison, again using the specification in Equation 2, but instead of altering the technology measure, it retains AI as the technology measure for all regressions and varies the skill measure within the organization. It reports the coefficient estimate on the interaction term where the $ORGSKL$ measure in business-facing occupations is generated using the prevalence of network and database technology skills in this segment of the workforce, rather than AI skills. Again, we see that the market value correlations are only different from zero when investment in AI technologies is accompanied by AI skills in this layer of the workforce. General tech acumen dispersed within the business layer of the workforce does not appear to raise returns to AI investments.

## 5 Managerial Implications

Adoption of algorithmic decision-making, and particularly predictive AI applications, has been difficult and uneven. Our evidence suggests that the human capital of leading, data-driven firms may differ in important ways from that of firms that are lagging in this domain. If investments in algorithms require substantial coomplementary workforce change, it may suggest considerable adjustment costs for firms seeking to adopt these management practices. High adjustment costs imply higher levels of concentration for AI and algorithmic investment and competitive rents for firms that have successfully installed the right workforce complements.

A corollary is that the costs of using data science technologies are continuously falling. Generative AI applications, for instance, represent a significant shift in how knowledge workers interact with technology, making it more accessible and user-friendly than many previous information technologies. One of the most notable aspects of this shift is the ease with which employees can engage in technical tasks using natural language interfaces. This reduces the time and effort required to learn and use programming languages, making technology more accessible to a broader range of professionals. A fall in the costs of performing this work suggests that employers may accelerate the rate at which data analysis tasks are pushed to domain experts.

The implications of this shift for both managers and educators can be significant. From a management perspective, generative AI tools can democratize technical skills, enabling a more diverse range of employees to contribute to areas that were once the exclusive domain of specialists. This can lead to more innovative work environments that emphasize the productive combination of human creativity and computational power. Managers will need to adapt by focusing less on specific technical skills when hiring and more on general problem-solving abilities and adaptability.

Another challenge for managers is that technical expertise has economic attributes that differentiate it from other forms of expertise. For instance, frontier technical skills are known to derive significant productivity benefits from geographic agglomeration (Saxenian, 1996; Fallick et al., 2006). Moreover, rapid depreciation of technologies changes the economics of professions in which technical human capital plays an important role, which has implications for topics like gender diversity and skilled immigration that routinely attract scrutiny from legislators and managers. If a growing number of occupations requires expertise with technology and algorithms, it may have implications for the structure of labor markets for these professions.

For educators, the falling costs of technical skill acquisition associated with no-code and generative AI technologies suggests a curricular reorientation. Although technical skills will continue to remain important for specialized workers, there may be a greater emphasis on understanding how to effectively interact with AI tools, interpret their outputs, and apply critical thinking to leverage AI-generated content. Educators will need to focus more on educating students about how to guide and evaluate AI output, rather than just how to perform the tasks that AI can now handle. The results in this study suggest that this type of education will be needed for all majors, not just technical majors. Institutions that have not traditionally been as focused on providing technical skills to students, such as business schools, have observed a surge in interest in demand for courses teaching data, analytics, and AI technologies (Eisenmann, 2013; Lohr, 2017; Guetta and Griffel, 2021; Becker, 2023). This study suggests that these changes may be an appropriate response to a labor market that will increasingly demand algorithmic bilinguals.

## 6 Conclusions

This paper provides evidence from two different data sources that i) algorithmic literacy is becoming broadly dispersed across business occupations, ii) that this dispersion is occurring due to complementarities that arise between technical skill and domain expertise, and iii) that the market assigns higher value to firms that concurrently makes these workforce adjustments while investing in algorithmic tools. In doing so, it documents one important but early facet of the workforce transformation that is occurring around algorithmic technologies.

Nonetheless, there are several limitations of this analysis that are worth noting. The data analyzed here provide limited visibility into the degree and nature of the expertise required by workers and the analysis is limited to the narrow question of how a a specific category of skills are bundled into jobs. For example, the data do not record when domain experts require deep expertise with a technology or instead, when interactional expertise, required to engage with developers and builders of these tools, would be sufficient. These findings also leave open a number of important and challenging questions about how to restructure decisions around AI technologies and where firms should place oversight of algorithmic decisions.

There is significant scope for future work in this area. By most accounts, we are at the beginning of a very large wave of investment in technologies that convert data into decisions, and research about this phenomenon and the workforce transformation that will be required to accompany these changes is in its infancy. There is a great deal to be learned about how to design organizations so that humans can effectively work together with algorithms. Although this paper considers one facet of workforce transformation, complements to AI and data science technologies will be wide-ranging. They will likely include more sweeping changes to workforce skills, as well as other non-labor factors needed to drive these capabilities (Rock et al., 2024). Firms' information capabilities will also continue to evolve and algorithms will become easier to deploy as better software and tools become available, which will lower the costs of adoption and further accelerate the diffusion of these technologies into new jobs.

Indeed a key limitation of this paper, like most research on technology and work, is that it takes a static view. At this early stage of adoption, there is still relatively little evidence that the use of these technologies has broad labor market consequences (Acemoglu et al., 2022). Stronger causal evidence of the impact of these workforce changes on performance may require allowing firms more time to adapt to this new mode of production. Additionally, new technologies for data collection, analysis, prediction, and visualization will offer improved capabilities to generate insights. As this boundary pushes forward, it will continue to change markets for these skills, and continue to raise new questions about how employers should integrate algorithms into the workflow.

# References


Acemoglu, D., Autor, D., Hazell, J. and Restrepo, P. (2022), 'Artificial intelligence and jobs: Evidence from online vacancies', *Journal of Labor Economics* **40**(S1), S293–S340.

Acemoğlu, D. and Restrepo, P. (2016), 'The race between machines and humans: Implications for growth, factor shares and jobs'.

Agrawal, A., Gans, J. and Goldfarb, A. (2018), *Prediction machines: The simple economics of artificial intelligence*, Harvard Business Press.

Agrawal, A., Gans, J. S. and Goldfarb, A. (2019), 'Artificial intelligence: The ambiguous labor market impact of automating prediction', *Journal of Economic Perspectives* **33**(2), 31–50.

Ang, S., Slaughter, S. and Yee Ng, K. (2002), 'Human capital and institutional determinants of information technology compensation: Modeling multilevel and cross-level interactions', *Management Science* **48**(11), 1427–1445.

Atasoy, H., Banker, R. D. and Pavlou, P. A. (2016), 'On the longitudinal effects of it use on firm-level employment', *Information Systems Research* **27**(1), 6–26.

Autor, D., Chin, C., Salomons, A. M. and Seegmiller, B. (2022), New frontiers: The origins and content of new work, 1940–2018, Technical report, National Bureau of Economic Research.

Autor, D., Levy, F. and Murnane, R. J. (2003), 'The skill content of recent technological change: An empirical exploration.', *Quarterly Journal of Economics* **118**(4).

Autor, D. and Salomons, A. (2018), 'Is automation labor share-displacing? productivity growth, employment, and the labor share', *Brookings Papers on Economic Activity* p. 1.

Babina, T., Fedyk, A., He, A. X. and Hodson, J. (2022), 'Firm investments in artificial intelligence technologies and changes in workforce composition', *Available at SSRN 4060233* .

Bartel, A., Ichniowski, C. and Shaw, K. (2007), 'How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills', *The Quarterly Journal of Economics* **122**(4), 1721–1758.

Bayer, S., Gimpel, H. and Markgraf, M. (2022), 'The role of domain expertise in trusting and following explainable ai decision support systems', *Journal of Decision Systems* **32**(1), 110–138.

Beane, M. (2019), 'Shadow learning: Building robotic surgical skill when approved means fail', *Administrative Science Quarterly* **64**(1), 87–123.

Becker, G. S. and Murphy, K. M. (1992), 'The division of labor, coordination costs, and knowledge', *The Quarterly Journal of Economics* **107**(4), 1137–1160.

Becker, S. (2023), 'How MBA programs are embracing data in their curricula', *Fortune* .

Berman, E., Bound, J. and Griliches, Z. (1994), 'Changes in the demand for skilled labor within US manufacturing: Evidence from the annual survey of manufactures', *The Quarterly Journal of Economics* **109**(2), 367–397.

Bessen, J. (2019), 'Automation and jobs: When technology boosts employment', *Economic Policy* **34**(100), 589–626.

Black, S. E. and Lynch, L. M. (2001), 'How to compete: The impact of workplace practices and information technology on productivity', *Review of Economics and Statistics* **83**(3), 434–445.

Bloom, N., Sadun, R. and Reenen, J. V. (2012), 'Americans do IT better: US multinationals and the productivity miracle', *American Economic Review* **102**(1), 167–201.

Bresnahan, T. F., Brynjolfsson, E. and Hitt, L. M. (2002), 'Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence.', *The Quarterly Journal of Economics* **117**(1).

Brynjolfsson, E. and Hitt, L. (1996), 'Paradox lost? firm-level evidence on the returns to information systems spending', *Management science* **42**(4), 541–558.

Brynjolfsson, E., Hitt, L. M. and Yang, S. (2002), 'Intangible assets: Computers and organizational capital', *Brookings papers on economic activity* **2002**(1), 137–181.

Brynjolfsson, E., Jin, W. and McElheran, K. (2021), 'The power of prediction: Predictive analytics, workplace complements, and business performance', *Business Economics* **56**(4), 217–239.

Brynjolfsson, E., Mitchell, T. and Rock, D. (2018), What can machines learn and what does it mean for occupations and industries?, *in* 'AEA Papers and Proceedings'.

Caroli, E. and Van Reenen, J. (2001), 'Skill-biased organizational change? Evidence from a panel of British and French establishments', *Quarterly Journal of Economics* pp. 1449–1492.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. et al. (2000), 'Crisp-dm 1.0: Step-by-step data mining guide', *SPSS inc* **9**(13), 1–73.

Choudhury, P., Starr, E. and Agarwal, R. (2020), 'Machine learning and human capital complementarities: Experimental evidence on bias mitigation', *Strategic Management Journal* **41**(8), 1381–1411.

Collins, H. (2004), 'Interactional expertise as a third kind of knowledge', *Phenomenology and the Cognitive Sciences* **3**(2), 125–143.

Colson, E. (2019), 'Why data science teams need generalists, not specialists', *Harvard Business Review* .

Cowgill, B. (2018), 'Bias and productivity in humans and algorithms: Theory and evidence from resume screening', *Columbia Business School, Columbia University* **29**.

Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N. and Chaintreau, A. (2020), Biased programmers? or biased data? A field experiment in operationalizing AI ethics, *in* 'Proceedings of the 21st ACM Conference on Economics and Computation', pp. 679–681.

Davenport, T. H. and Patil, D. (2012), 'Data scientist: The sexiest job of the 21st century', *Harvard business review* **90**(5), 70–76.

Deming, D. J. and Noray, K. (2020), 'Earnings dynamics, changing job skills, and stem careers', *The Quarterly Journal of Economics* **135**(4), 1965–2005.

Deming, D. and Kahn, L. B. (2018), 'Skill requirements across firms and labor markets: Evidence from job postings for professionals', *Journal of Labor Economics* **36**(S1), S337–S369.

Dessein, W. and Santos, T. (2006), 'Adaptive organizations', *Journal of Political Economy* **114**(5), 956–995.

Dixon, J., Hong, B. and Wu, L. (2021), 'The robot revolution: Managerial and employment consequences for firms', *Management Science* **67**(9), 5586–5605.

Eisenmann, T. (2013), 'Should MBAs learn to code?', *Harvard Business Review* .

Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2023), 'Gpts are gpts: An early look at the labor market impact potential of large language models', *arXiv preprint arXiv:2303.10130* .

Fallick, B., Fleischman, C. A. and Rebitzer, J. B. (2006), 'Job-hopping in Silicon Valley: Some evidence concerning the microfoundations of a high-technology cluster', *The Review of Economics and Statistics* **88**(3), 472–481.

Faraj, S., Pachidi, S. and Sayegh, K. (2018), 'Working and organizing in the age of the learning algorithm', *Information and Organization* **28**(1), 62–70.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E. et al. (2019), 'Toward understanding the impact of artificial intelligence on labor', *Proceedings of the National Academy of Sciences* **116**(14), 6531–6539.

Gallivan, M., Truex III, D. P. and Kvasny, L. (2002), An analysis of the changing demand patterns for information technology professionals, *in* 'Proceedings of the 2002 ACM SIGCPR conference on Computer personnel research', ACM, pp. 1–13.

Goldfarb, A., Taska, B. and Teodoridis, F. (2023), 'Could machine learning be a general purpose technology? a comparison of emerging technologies using data from online job postings', *Research Policy* **52**(1), 104653.

Gregory, T., Salomons, A. and Zierahn, U. (2022), 'Racing with or against the machine? evidence on the role of trade in europe', *Journal of the European Economic Association* **20**(2), 869–906.

Guetta, D. and Griffel, M. (2021), 'Python: The new MBA must have', *ORMS Today* .

Hershbein, B. and Kahn, L. B. (2018), 'Do recessions accelerate routine-biased technological change? Evidence from vacancy postings', *American Economic Review* **108**(7), 1737–72.

Horton, J. J. and Tambe, P. (2015), 'Labor economists get their microscope: Big data and labor market analysis', *Big data* **3**(3), 130–137.

Jha, S. and Topol, E. J. (2016), 'Adapting to artificial intelligence: Radiologists and pathologists as information specialists', *Jama* **316**(22), 2353–2354.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S. (2018), 'Human decisions and machine predictions', *The Quarterly Journal of Economics* **133**(1), 237–293.

Kumar, P. and Sharma, M. (2022), 'Data, machine learning, and human domain experts: none is better than their collaboration', *International Journal of Human–Computer Interaction* **38**(14), 1307–1320.

Lebovitz, S., Lifshitz-Assaf, H. and Levina, N. (2022), 'To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis', *Organization Science* **33**(1), 126–148.

Lee, C. K. and Han, H. (2008), 'Analysis of skills requirement for entry-level programmer/analysts in Fortune 500 corporations', *Journal of Information Systems Education* **19**(1), 17.

Lee, D. M., Trauth, E. M. and Farwell, D. (1995), 'Critical skills and knowledge requirements of IS professionals: A joint academic/industry investigation', *MIS quarterly* pp. 313–340.

Levina, N. and Xin, M. (2007), 'Research note-comparing IT workers' compensation across country contexts: Demographic, human capital, and institutional factors', *Information Systems Research* **18**(2), 193–210.

Lichtenberg, F. R. (1995), 'The output contributions of computer equipment and personnel: A firm-level analysis', *Economics of innovation and new technology* **3**(3-4), 201–218.

Lindbeck, A. and Snower, D. J. (2000), 'Multitask learning and the reorganization of work: From Tayloristic to Holistic organization', *Journal of labor economics* **18**(3), 353–376.

Lohr, S. (2017), 'Where non-techies can get with the programming', *THE NEW YORK TIMES, April* **4**.

Lohr, S. (2024), 'Generative a.i.'s biggest impact will be in banking and tech, report says', *THE NEW YORK TIMES, Feb* .

Mao, Y., Wang, D., Muller, M., Varshney, K. R., Baldini, I., Dugan, C. and Mojsilović, A. (2019), 'How data scientistswork together with domain experts in scientific collaborations: To find the right answer or to ask the right question?', *Proceedings of the ACM on Human-Computer Interaction* **3**(GROUP), 1–23.

Mendelson, H. and Pillai, R. R. (1998), 'Clockspeed and informational response: Evidence from the information technology industry', *Information Systems Research* **9**(4), 415–433.

Mithas, S. and Krishnan, M. S. (2008), 'Human capital and institutional effects in the compensation of information technology professionals in the united states', *Management Science* **54**(3), 415–428.

Modestino, A. S., Shoag, D. and Ballance, J. (2019), 'Upskilling: do employers demand greater skill when workers are plentiful?', *Review of Economics and Statistics* pp. 1–46.

Niederman, F., Ferratt, T. W. and Trauth, E. M. (2016), 'On the co-evolution of information technology and information systems personnel', *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* **47**(1), 29–50.

Oostendorp, N. (2019), 'Radical change is coming to data science jobs'.

Park, S., Wang, A. Y., Kawas, B., Liao, Q. V., Piorkowski, D. and Danilevsky, M. (2021), Facilitating knowledge sharing from domain experts to data scientists for building nlp models, *in* '26th International Conference on Intelligent User Interfaces', pp. 585–596.

Pavlou, P. A. and El Sawy, O. A. (2006), 'From IT leveraging competence to competitive advantage in turbulent environments: The case of new product development', *Information systems research* **17**(3), 198–227.

Postrel, S. (2002), 'Islands of shared knowledge: Specialization and mutual understanding in problem-solving teams', *Organization science* **13**(3), 303–320.

Provost, F. and Fawcett, T. (2013), 'Data science and its relationship to big data and data-driven decision making', *Big data* **1**(1), 51–59.

Raj, M. and Seamans, R. (2018), Artificial intelligence, labor, productivity, and the need for firm-level data, *in* 'The economics of artificial intelligence: An agenda', University of Chicago Press, pp. 553–565.

Rock, D. (2019), 'Engineering value: The returns to technological talent and investments in artificial intelligence', *Available at SSRN 3427412* .

Rock, D., Tambe, P. and Wang, Z. (2024), 'Deep capital for deep learning', *Working Paper* .

Saltz, J. S. (2021), Crisp-dm for data science: Strengths, weaknesses and potential next steps, *in* '2021 IEEE International Conference on Big Data (Big Data)', IEEE, pp. 2337–2344.

Sambasivan, N. and Veeraraghavan, R. (2022), The deskilling of domain expertise in ai development, *in* 'Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems', pp. 1–14.

Saxenian, A. (1996), *Regional advantage*, Harvard University Press.

Slaughter, S. and Ang, S. (1996), 'Employment outsourcing in information systems', *Communications of the ACM* **39**(7), 47–54.

Smith, A. (1776), *From The Wealth of Nations.*

Spitz-Oener, A. (2006), 'Technical change, job tasks, and rising educational demands: Looking outside the wage structure', *Journal of labor economics* **24**(2), 235–270.

Tambe, P. (2014), 'Big data investment, skills, and firm value', *Management Science* **60**(6), 1452–1469.

Tambe, P., Hitt, L. M. and Brynjolfsson, E. (2012), 'The extroverted firm: How external information practices affect innovation and productivity', *Management Science* **58**(5), 843–859.

Tambe, P., Ye, X. and Cappelli, P. (2020), 'Paying to program? Engineering brand and high-tech wages', *Management Science* .

Teodoridis, F. (2017), 'Understanding team knowledge production: The interrelated roles of technology and expertise', *Management Science* **64**(8), 3625–3648.

Todd, P. A., McKeen, J. D. and Gallupe, R. B. (1995), 'The evolution of IS job skills: A content analysis of IS job advertisements from 1970 to 1990', *MIS Quarterly* pp. 1–27.

White, R. W., Dumais, S. T. and Teevan, J. (2009), Characterizing the influence of domain expertise on web search behavior, *in* 'Proceedings of the second ACM international conference on web search and data mining', pp. 132–141.

Wiesche, M., Joseph, D., Thatcher, J., Gu, B. and Krcmar, H. (2019), 'Research curation: IT workforce', *MIS Quarterly* .

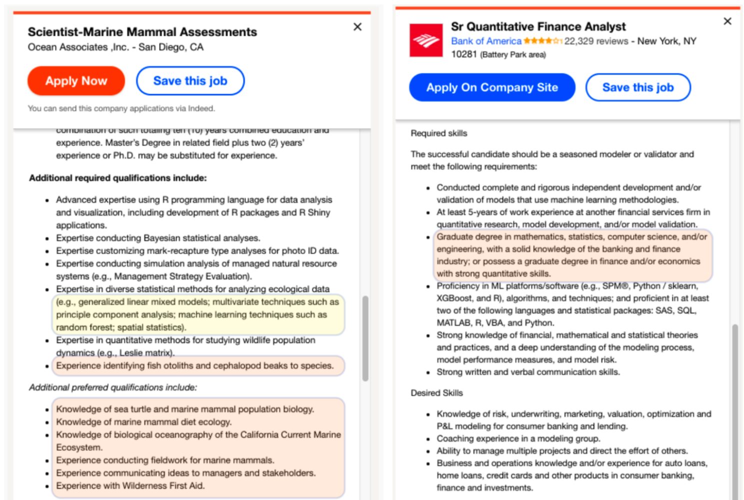
Wirth, R. and Hipp, J. (2000), Crisp-dm: Towards a standard process model for data mining, *in* 'Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining', Vol. 1, Manchester, pp. 29–39.

Wu, L., Lou, B. and Hitt, L. (2019), 'Data analytics supports decentralized innovation', *Management Science* **65**(10), 4863–4877.

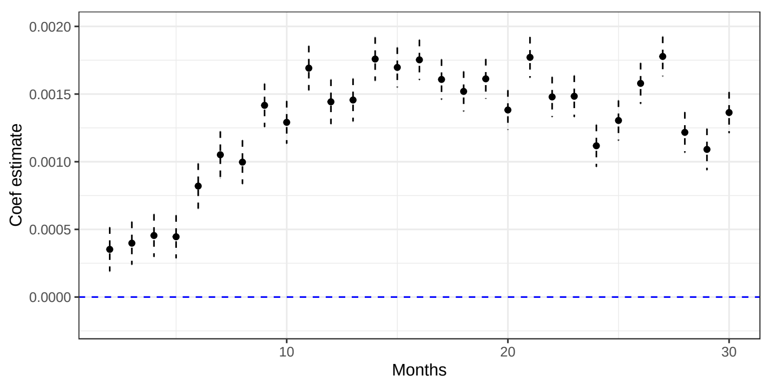
Xue, M., Cao, X., Feng, X., Gu, B. and Zhang, Y. (2022), 'Is college education less necessary with AI? Evidence from firm-level labor structure changes', *Journal of Management Information Systems* **39**(3), 865–905.

Zolas, N., Kroff, Z., Brynjolfsson, E., McElheran, K., Beede, D. N., Buffington, C., Goldschlag, N., Foster, L. and Dinlersoz, E. (2021), Advanced technologies adoption and use by us firms: Evidence from the annual business survey, Technical report, National Bureau of Economic Research.

Figure 1: The growth of algorithmic skills in job listings



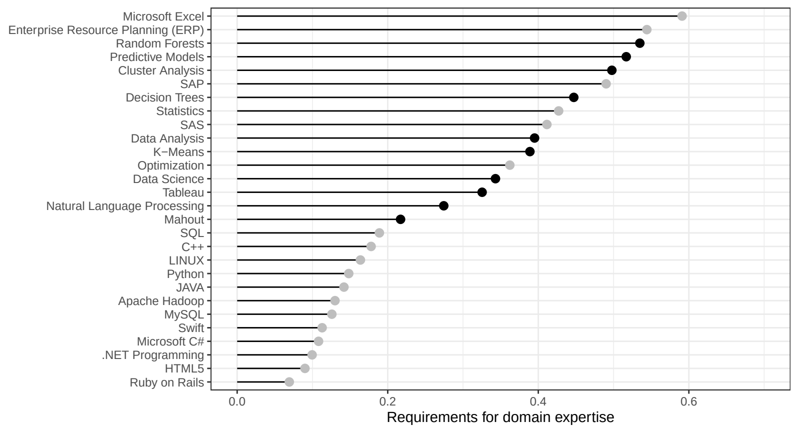

(a) Sample listings with algorithmic and domain expertise



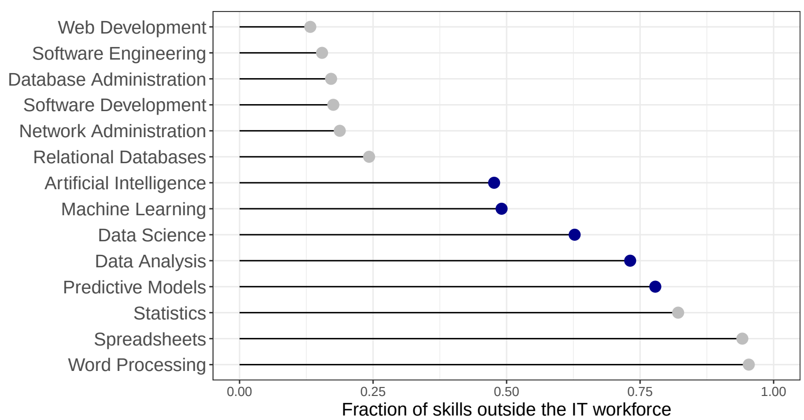

(b) Likelihood an algorithmic skill appears in a job listing

*Figure notes*: Figure (a) shows sample listings for jobs that require familiarity with both algorithmic tools (highlighted in yellow) and domain expertise (highlighted in orange), related in these examples to marine biology and finance. These listings and screenshots were extracted from the website Indeed.com. Figure (b) shows the coefficient estimates and standard error bars on the regression $ALG_i = \beta month_i + \epsilon_i$ for each of the months from 2013 onwards (omitting January of 2013), where $i$ indexes job listings and $ALG$ indicates whether an algorithmic skills appears in a listing.

Figure 2: Algorithmic skills, domain expertise, and job listings



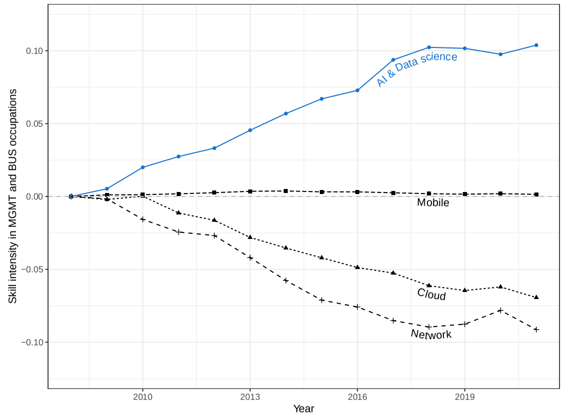

(a) Bundling of technical skills with domain expertise



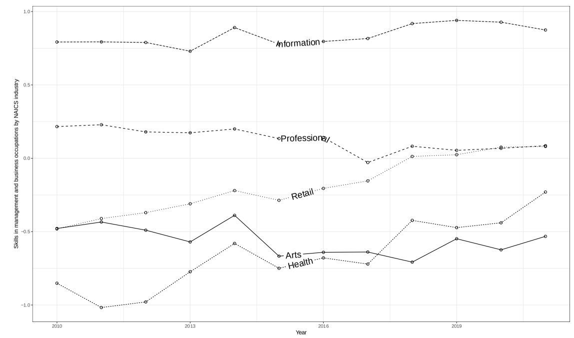

(b) Fraction of skills that appear outside IT job listings

*Figure notes*: Figure (a) indicates the extent to which different technical skills are bundled with domain expertise. Skills in dark blue are those associated with the algorithmic category.Figure (b) indicates the fraction of times a skill appears in listings outside an IT occupation where a value close to 1 means that skills is almost always appearing in job listings for non-IT occupations.

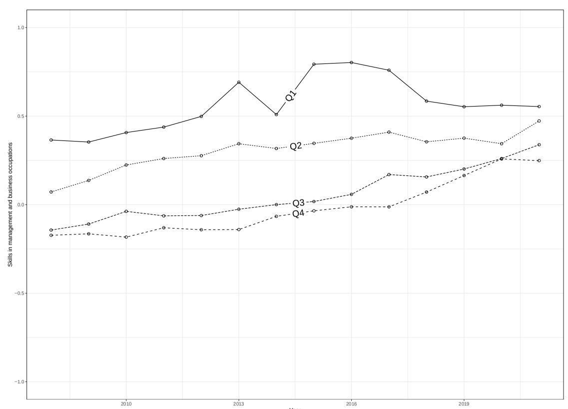
Figure 3: Changes in the locus of technical skills in organizations from 2008-2021



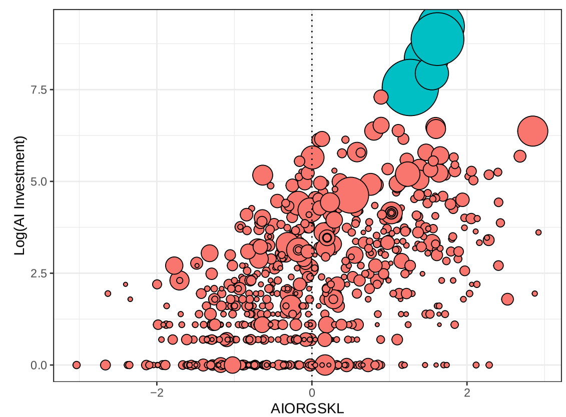

(a) Changes in skills in organizations




(b) Skills by industry



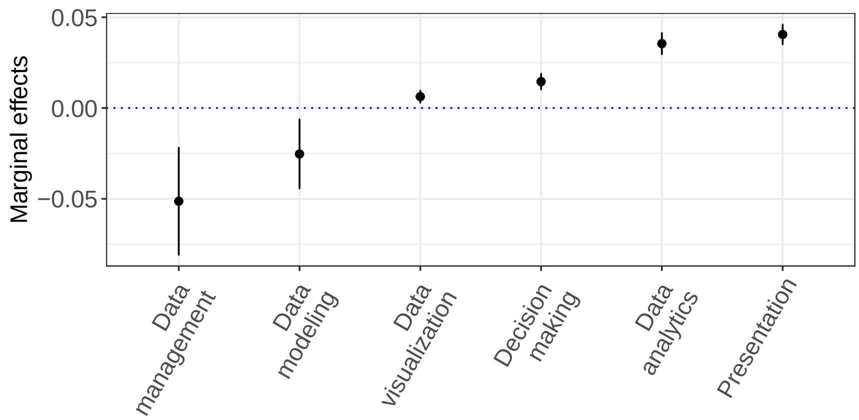

(c) Convergence across size quartiles



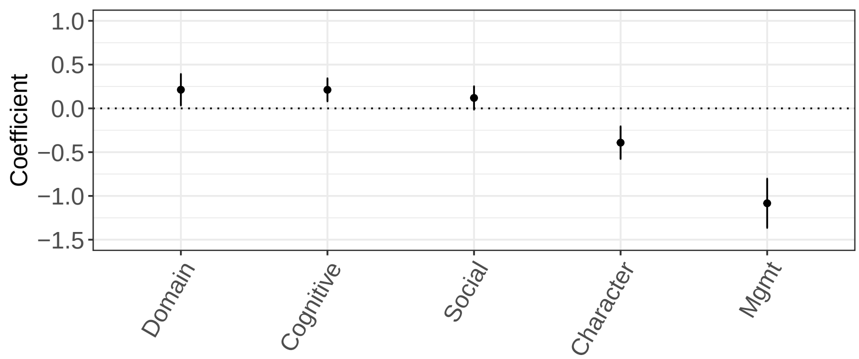

(d) Organizational accumulation of skills

*Figure notes*: Figure (a) illustrates changes in the skill measure over time when constructed using different technologies. All trend lines are changes from theie values in the base year (2008). Figure (b) illustrates changes in this skill measure over the course of our panel in different industries. Figure (c) shows changes in this measure in organizations over time where firms are divided into four different quartiles by size. Figure (d) plots the organizational AI skill measure (x-axis) against its AI investment levels (y-axis). Bubbles are sized by firm's market value.

Figure 4: Characteristics of job listings with $ALG$ skills




(a) Marginal effects on data pipeline tasks on domain expertise




(b) Job characteristics and algorithmic skills

*Figure notes*: This figure depicts correlations between various categories of job characteristics and algorithmic skills appearing in job listings. Each vertical bar is a separate regression of the form $CATEGORY_i = \alpha_{ALG}ALG_i + \alpha_{DATA}DATA_i + \alpha_{NET}NETWORK_i + Log(Skills)_i + \gamma_i + \phi_i + \epsilon_i$ where CATEGORY is one of DOMAIN, COGNITIVE, SOCIAL, CHARACTER, or MANAGEMENT, $i$ indexes the listing, $\gamma$ and $\phi$ are occupation and industry fixed-effects respectively, and $Log(skills)$ is the log of the number of skills in the listing. The point estimate that is shown is the coefficient on $\alpha_{ML}$ from each regression and the vertical bars indicate 95% confidence intervals.

Table 1: Summary statistics for firms (2018 cross-section)

| Variable | Units | Mean | Std. Dev. | N |
|---|---|---|---|---|
| Market value | Thousands (USD) | 46456.1 | 189437.4 | 1509 |
| Assets | Thousands (USD) | 48694.0 | 230567.6 | 1509 |
| Prop., Plant, and Equip. | Thousands (USD) | 3447.3 | 11102.9 | 1509 |
| Employment | Thousands | 24.7 | 78.4 | 1509 |
| Data capital | Tech skill count | 1130.3 | 7323.7 | 1509 |
| Network capital | Tech skill count | 276.7 | 1622.5 | 1509 |
| AI capital | Tech skill count | 59.2 | 411.8 | 1509 |
| IT capital | Tech skill count | 6.3 | 19.3 | 1509 |
| *ALGSKL* | Standardized Value | 0.0 | 1.1 | 1509 |

*Table notes*: This table reports summary statistics for firms in the 2018 cross-section of the workforce sample. The year 2018 was chosen because it is the midpoint in the panel window.

Table 2: Conditional correlations between technology investment and employee skills

| | Skill dispersion | |
|---|---|---|
| Model: | (1) | (2) |
| *Variables* | | |
| Log(Alg) | 0.385*** | 0.111*** |
| | (0.019) | (0.014) |
| Log(Assets) | 0.021 | −0.025 |
| | (0.018) | (0.024) |
| Log(PPE) | −0.018 | 0.005 |
| | (0.022) | (0.015) |
| Log(IT) | −0.503*** | 0.011 |
| | (0.049) | (0.033) |
| Log(Employment) | −0.265*** | 0.027 |
| | (0.039) | (0.037) |
| *Fixed-effects* | | |
| Firm FE | | Yes |
| Year FE | Yes | Yes |
| Industry FE (NAICS 3) | Yes | |
| *Fit statistics* | | |
| $R^2$ | 0.533 | 0.937 |
| Observations | 10,464 | 10,464 |

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

*Figure notes*: This table reports regressions of how workforce skill composition relates to firms' market value. The regression model is $Log(Market\ Value)_{it} = Log(Assets)_{it} + Log(ITInv)_{it} + Log(ALGSKL)_{it} + \epsilon_{it}$ where the observations are at the level of the firm-year.

Table 3: Relationship between technology investment, skill dispersion, and market value

| | Log(Market Value) | | | |
|---|---|---|---|---|
| Model: | (1) | (2) | (3) | (4) |
| *Variables* | | | | |
| Log(Assets) | $0.767^{***}$ | $0.759^{***}$ | $0.754^{***}$ | $0.662^{***}$ |
| | (0.036) | (0.037) | (0.037) | (0.028) |
| Log(PPE) | $0.154^{***}$ | $0.154^{***}$ | $0.149^{***}$ | $0.096^{***}$ |
| | (0.031) | (0.031) | (0.030) | (0.018) |
| Log(IT) | $0.140^{***}$ | −0.010 | $0.075^{**}$ | 0.004 |
| | (0.022) | (0.028) | (0.029) | (0.026) |
| Log(Employment) | −0.020 | −0.020 | 0.021 | $0.198^{***}$ |
| | (0.039) | (0.037) | (0.033) | (0.043) |
| Log(Alg) | | $0.086^{***}$ | 0.020 | $0.031^{***}$ |
| | | (0.015) | (0.019) | (0.012) |
| *ALGSKL* | | | $0.085^{**}$ | 0.023 |
| | | | (0.040) | (0.034) |
| Log(Alg) × *ALGSKL* | | | $0.036^{***}$ | $0.026^{**}$ |
| | | | (0.008) | (0.011) |
| *ALGSKL* × Log(IT) | | | −0.016 | −0.034 |
| | | | (0.024) | (0.022) |
| *Fixed-effects* | | | | |
| Firm FE | | | | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Industry FE (NAICS 3) | Yes | Yes | Yes | |
| *Fit statistics* | | | | |
| $R^2$ | 0.913 | 0.915 | 0.918 | 0.986 |
| Observations | 8,663 | 8,663 | 8,663 | 8,662 |

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

*Figure notes*: This table reports regressions of how workforce skill composition relates to firms' market value. The regression model is $Log(Market\ Value)_{it} = Log(Assets)_{it} + Log(ITInv)_{it} + Log(ALGSKL)_{it} + \epsilon_{it}$ where the observations are at the level of the firm-year.

Figure 5: Coefficient estimates by employer size

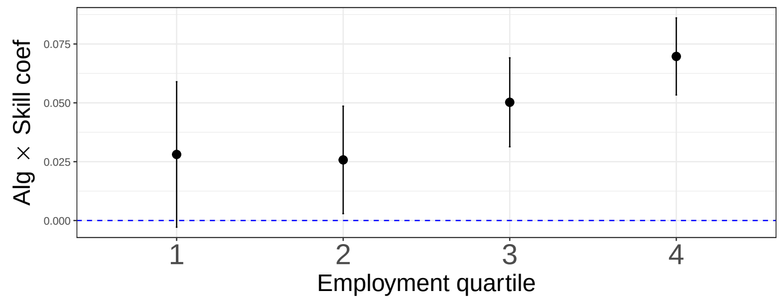
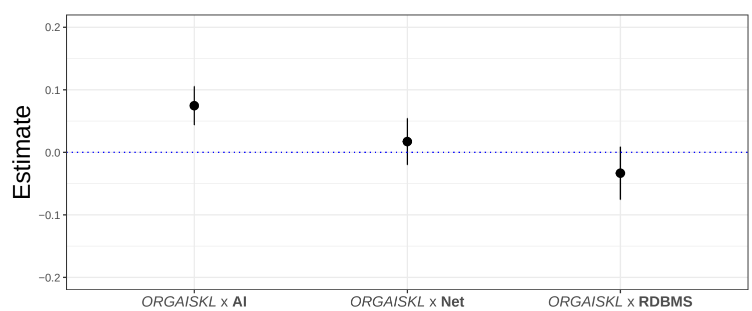
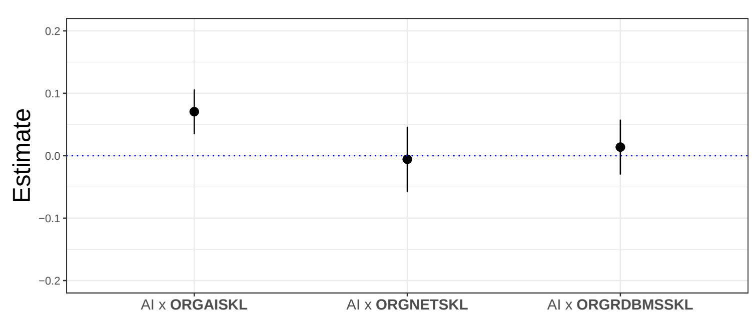



*Figure notes*: The y-axis indicates the coefficient on the interaction term between algorithm investment and skill dispersion. The x-axis divides firms by employment quartile where "1" is the smallest firms and "4" is the largest firms.

Figure 6: Robustness tests using alternative measures of $TECH$ and $ORGSKL$




(a) Comparisons with other information technologies




(b) Comparisons with other organizational skill measures

*Notes:* The top facet illustrates placebo tests when using other measures of $TECH$ in the regression $Log(MV)_{it} = Log(Assets)_{it} + TECH_{it} + ORGSKL_{it} + (TECH_{it} \times ORGSKL_{it}) + \gamma_{it} + \epsilon_{it}$. The bottom facet illustrates the coefficient on the interaction term when using other measures of $ORGSKL$.

# A Corporate workforce data

This section discusses the Revelio corporate workforce data and presents comparisons of these workforce data with administrative data sets. The purpose of this comparison is to evaluate the sampling properties of these data and discuss any limitations sampling restrictions might impose on the estimates presented above. To evaluate coverage in these data, comparisons are presented below with three different data sources: i) the occupational distribution of US workers reported by the Bureau of Labor Statistics (BLS), ii) the NAICS industry distribution of workers, and iii) how workers are distributed across US states. These are all data sets with known sampling properties.

## A.1 Data generating process

Revelio is a workforce intelligence company that federates data across a wide range of Internet sources including federal databases, professional networking sites, and job posting aggregators. This analysis relies on their workforce, position, and skill databases which capture data on the movements of a very large sample of employees across firms, the job titles they hold, and the skills they acquire.

However, there are also several potential concerns when using data sources of this type.

Prior work discusses some of the considerations when using data of these type to study labor market flows (Horton and Tambe, 2015).

## A.2 BLS share comparisons

The first set of comparisons between Revelio workers and the Standard Occupational Codes (SOC) is shown in Figure 7b. The figure presents differences in shares of the major occupational groups as reported by the BLS and represented in the Revelio data, where the assignment of workers to SOC areas in the Revelio data set is done by Revelio. The blue line indicates no (zero) difference in shares such that bars to the right (left) are those occupations where the occupation accounts for a higher (lower) proportion of workers in the BLS data than the Revelio data. From this comparison, we can see that "white-collar", knowledge-intensive occupations like management and IT work, tend to be over represented in the Revelio data set whereas front-line occupations in sectors like manufacturing, production, and transportation are underrepresented. This is not a surprise given that these data are gathered from professional networking sites on which white-collar workers tend to be over represented. The length of each bar is the difference in shares across these two data sources. The largest imbalance in occupations in Management. The difference in the share of total workers that managers account for in the Revelio data set (15%) and the BLS (7%) is about 8% percentage points.

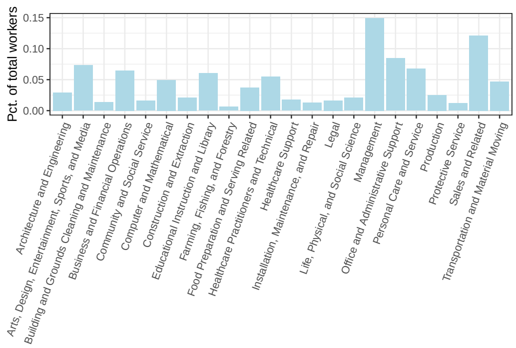
Industry level comparisons are reported in Table 7d. These industry level comparisons are conducted at the 2-digit NAICS (North American Industry Classification System) level where the underlying allocation of workers across industries is again taken from the Occupational Employment Survey data. Industry classifications in the Revelio data are generated by assigning employers to industries and are directly reported by Revelio for each employee. The share differences we can observe in this comparison are consistent with the observation that white-collar professions are overrepresented. Tech, finance, professional services, and manufacturing account for larger shares of employees in the Revelio data than in the BLS data. Healthcare and construction account for smaller shares.

A final comparison, shown in the third panel (Figure 7f), is state-level comparisons. We can see that states with significant industry representation for finance and tech (such as New York) account for a relatively larger share of workers in the Revelio data. The largest imbalance is in North Dakota, where industries like Oil extraction and Agriculture play a large role in the state economy.

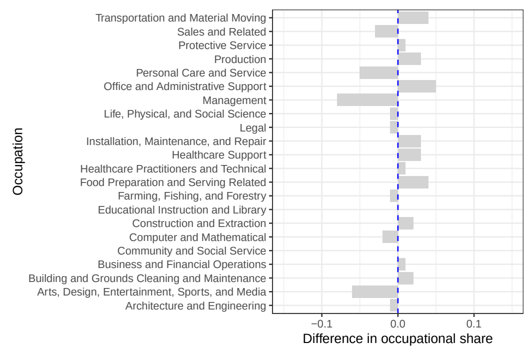
In sum, Figure **??** indicates that the workers in the Revelio data set are likely overrepresentative of the industries and sectors and types of workers that are most likely to adapt to these technological changes.

*Notes:* These three figures illustrate the difference in compositional shares between the Revelio and BLS data sets. The height of each bar is computed as the difference in the share that the worker category accounts for in the Revelio data and in the BLS data. For instance, Management workers comprise 15% of the Revelio data set and 7% of the BLS data set so the length of the bar indicates an 8% difference between the two.

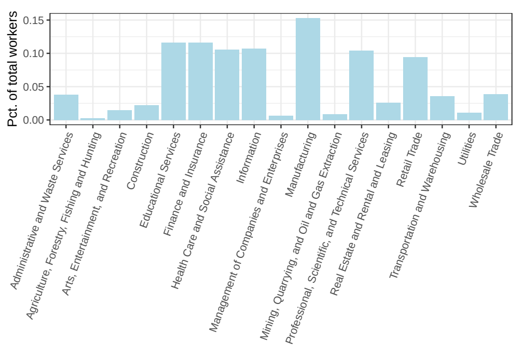
Figure 7: Revelio data distributions



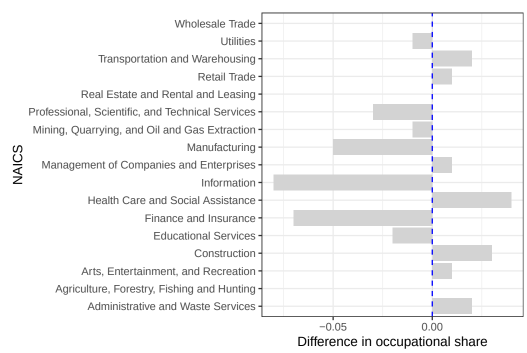

(a) Occupational code (SOC)



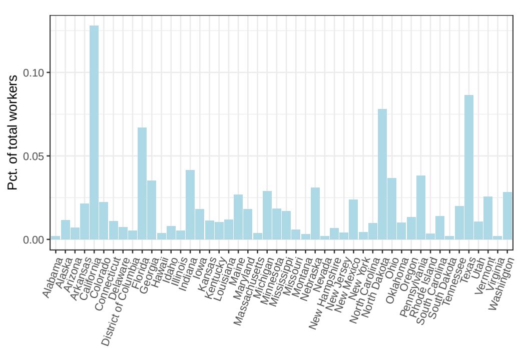

(b) Occupational code (SOC)



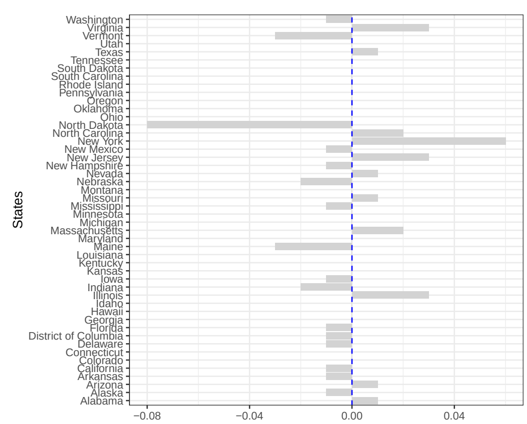

(c) Industry (NAICS)




(d) Industry (NAICS)




(e) States




(f) States

# B Additional robustness tests

In this section, we report the results of additional robustness tests.

Table B.1: Logistic regression of algorithmic tools on domain expertise and other job attributes

| | *Dependent variable:* | | | | |
|---|---|---|---|---|---|
| | Domain | Social | Character | Cognitive | Management |
| | (1) | (2) | (3) | (4) | (5) |
| *ALG* | $0.213^{**}$ | $0.119^{*}$ | $-0.391^{***}$ | $0.212^{***}$ | $-1.084^{***}$ |
| | (0.091) | (0.067) | (0.094) | (0.066) | (0.143) |
| *MANAGE* | $-0.414^{***}$ | $-0.157^{***}$ | $-0.229^{***}$ | $-0.264^{***}$ | $-0.680^{***}$ |
| | (0.078) | (0.056) | (0.074) | (0.057) | (0.085) |
| *COLLECT* | 0.002 | 0.050 | 0.088 | $-0.070^{*}$ | $-0.024$ |
| | (0.065) | (0.041) | (0.057) | (0.042) | (0.066) |
| Log(No. of Skills) | $1.559^{***}$ | $1.552^{***}$ | $1.903^{***}$ | $1.774^{***}$ | $2.237^{***}$ |
| | (0.056) | (0.036) | (0.055) | (0.038) | (0.064) |
| Constant | $-6.754^{***}$ | $-4.857^{***}$ | $-7.241^{***}$ | $-5.760^{***}$ | $-7.910^{***}$ |
| | (0.185) | (0.112) | (0.176) | (0.120) | (0.204) |
| Job Title FE | Yes | Yes | Yes | Yes | Yes |
| Industry FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 24,388 | 24,388 | 24,388 | 24,388 | 24,388 |
| Log Likelihood | -6,929.787 | -13,096.380 | -7,572.037 | -12,642.760 | -5,966.758 |
| Akaike Inf. Crit. | 14,271.580 | 26,604.760 | 15,556.070 | 25,697.520 | 12,345.510 |

*Table notes*: This table reports results from the logit regression $domain_i = ALG_i + MANAGE_i + COLLECT_i + \phi_i + \epsilon_i$. It estimates conditional correlations between algorithmic and skill based job attributes. The variable $ALG$ indicates whether the job ad includes at least one skill related to an algorithmic tool and $Log(No.of skills)$ is the log of the total number of skills in the job ad. The dependent variable indicates whether or not a job listing requires knowledge of an application domain, social skills, character, cognitive skills, and people management skills, respectively. All regressions include job title and industry fixed-effects (NAICS 4). Standard errors are shown in parentheses, with *, **, and *** denoting significance at the 10%, 5%, and 1% level, respectively.

# C Categorizing skills into technological areas

A key measurement task for this analysis is to use the skills embedded in job listings or reported by employees on their profiles as an indicator of having some measure of expertise with a technology. This requires constructing a mapping from thousands of skills to the broader technological areas to which they are related. For instance, skills such as “Oracle DB“, “MySQL“, and “Relational Databases“ are all indicative of a worker having expertise with database technologies.

This is not an easy task, as it requires some measure of discretion. Therefore, rather than construct such a taxonomy from scratch, I rely on an existing taxonomy structure overlaid by the data provider that categorizes skills into technological groups. Th data science team of this data provider uses clustering techniques to place skills into a taxonomy, and the approach combines skills into common groups if they inhabit a similar area of the skill landscape after clustering. Labeling of technological groups is done after clusters emerge in the skill landscape. The default skill clustering, used in this analysis, is at fifty groups. The skills that appear in each of these categories, as constructed by the data provider, are shown below.

The provider for the job listings data also provides a taxonomy through which to interpret the detailed skills data that appear on job listings. However, to maintain consistency across the analysis, the skills in the job listings data were mapped to technological categories by harmonizing them with the Revelio taxonomy. For example, the ’Algorithms’ category was created in the job listings data by identifying skills in the listings data that had a match with one of the skills in the same category in the workforce data. Matches were identified manually, to account for minor differences in case or how skill names were standardized by the different providers.

## C.1 Technical skills in the job listings data

**Algorithmic skills**. Machine Learning, Decision Trees, Random Forests, Recommender Systems, Mahout, Support Vector Machines, Artificial Intelligence, Predictive Modeling, Predictive Analytics, Predictive Models, Data Mining, Deep Learning, Neural Networks, K-Means, Cluster Analysis, Natural Language Processing

**Data management**. SQL, MySQL, Structured Query Language, database management, database administration, data cleaning, data extraction, database querying, Big Data, Apache Hadoop, NoSQL, MongoDB, Apache Hive, Splunk, MapReduce, PIG, Cassandra, SOLR, Sqoop

**Data collection**. Objective C, Swift, HTML5, Javascript, HTML, iOS, CSS, Cisco, Network Engineering, Network Administration, Computer Networking, Network Support, Network Concepts and Terminology, Data Communications, Network Installation, Wireless Local Area Network (LAN), Network Management System, Network Infrastructure

## C.2 Technical skills in the workforce data

**Information Technology (IT)**. software testing, software engineering/software design, software training, software documentation, software installation/laptops, software development life cycle, embedded systems/embedded software, software,software architecture, software licensing, software quality assurance, software implementation,object oriented software, software deployment, open source software, software asset management, software project management, software integration, software development life cycle (sdlc), software development, release management, unix, ftp, object oriented design, oop, c++/c, c++ language, microsoft visual studio c++, visual c++,c/c++, windows server, windows server 2008/windows server 2003, .net/asp.net, unit testing, it governance, sdlc, bash, shell, linux, object-oriented programming, it audit/cisa, assembly language, servers, user acceptance testing, it, support/server, object-oriented programming (oop), continuous integration, it infrastructure management, operating systems,visual basic for applications (vba), information technology, shell scripting/unix shell scripting, linux system administration, code review, server administration, agile testing,regular expressions, system testing/system integration testing, powershell, ldap, orm, vb.net,linux kernel, vdi, ibm rational tools, nas/enterprise storage, smtp sap, ivr, ibm iseries, asp, weblogic, dos, ibm aix,ado.net/asp.net ajax, asp.net mvc/linq/entity framework, vsam, raid, it operations

**Network**. lan-wan, lan, ssl, ssl certificates, wan, network operations, ip networking, computer networking, voice over ip (voip)/internet protocol (ip), network troubleshooting,network architecture, network security,network development, computer network operations, wireless networking, network administration, san/storage area networks/netapp, internet protocol suite (tcp/ip), tcp/ip, data mapping tcp/ip protocols, routing protocols/switching,switches/routers, routing/qos, wifi, dns/dhcp, ethernet, wireless, mpls, netcool, ccna/ccnp, putty, wimax, snmp

**Data**. master data management, spatial databases/web mapping, data warehousing/etl, database administration, database, database security, metadata/metadata management, oracle sql developer/oracle database, data entry,data quality, data acquisition, data management, data processing, data integration/data warehouse architecture, data migration, database design,data collection, db2, sql, pl/sql, mssql/ms sql/ms sql server, sql server management studio, oracle sql, sqlite, mysql/php,performance tuning/sql tuning, oracle pl/sql development,sql server, microsoft sql server, extract/transform/load (etl),sybase, t-sql/ssis/ssrs, teradata, sap hana,jsp/jdbc, edi, sqr, rdbms, oracle rac, ibm db2

**Mobile**. android, objective-c/ios development, mobile device management, wireless technologies, wireless communications systems, mobile application development, swift/xcode, android development/android sdk

**Internet**. front-end,web analytics, cascading style sheets,website development, web services, website updating/website administration, web marketing, j2ee application development/j2ee web services,web 2.0, web applications, responsive web design, web development,basic html, html + css, html scripting, xhtml, html 5/css3, html5/bootstrap, html/css, cascading style sheets (css), django, css, tomcat/jboss application server,nginx, internet protocol suite, node.js/react.js, hosting, ajax, client/server, search engine optimization (seo)/search engine marketing (sem), angularjs, jquery/jquery ui, fiddler, javascript, wireshark, java script, lamp,selenium/selenium webdriver,jsf/jpa, soap,http, backbone.js,websphere/websphere application server, .net framework/asp.net web api, soapui, typescript, jmeter

**Big Data**. distributed systems/scalability, mongodb, hive/apache pig, docker/devops, middleware, data center, centos/debian, hadoop/apache spark/mapreduce, ubuntu, server architecture, red hat linux, high performance computing, vms,socket programming, olap, soa, websphere mq, multithreading, service-oriented architecture (soa), ibm tivoli, hive/apache pig

**Cloud**. microsoft azure, windows azure, amazon services/aws, cloud-computing, cloud computing, amazon web services (aws), cloud applications, vmware, openstack, vmware esx/vmware infrastructure/vsphere

**Data Science**. data visualization, data mining, statistical data analysis, big data, data modeling, data analytics/-data science/big data analytics, marketing analytics, quantitative analytics, analytics, business analytics, predictive analytics/predictive modeling, pandas, tableau, nosql/redis, numpy, R, scala, spark, julia, pyspark

**AI**. machine learning, natural language processing, image processing/computer vision, artificial intelligence, tensorflow, pytorch, scikit-learn