Algorithmic Bilinguals

April 25, 2024

Abstract

This study tests the hypothesis that generating value from data, algorithms, and AI requires domain experts who can interact with these technologies. This decentralization of technical expertise stands in contrast to other business technologies for which the complementary skills are primarily embodied in IT specialists and it is due to the difficulties 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 algorithmic expertise from domain experts, 2) technical human capital in frontier firms has become more dispersed across non-technical occupations, and 3) the market assigns higher value to firms' algorithm investments when they have made these complementary workforce adjustments, indicating the presence of valuable intangible assets that can yield a stream of future productivity benefits from AI and data science investments. Finally, I show that the recent advance of no-code and natural language tools, that make it easier for workers in non-technical occupations to perform technical work, accelerates these changes. Implications for training, education, and algorithmic decision-making are discussed.

Keywords: human capital, jobs, algorithms, AI literacy, IT intangibles, future of work, IT complements, reskilling

1 Introduction

The impact of algorithmic decision-making on organizations is a topic of growing interest (Rock, 2019; Wu et al., 2019; Agrawal et al., 2018; Zolas et al., 2021). Research in this area has focused on the labor reallocation effects of AI and automation technologies (Acemoğlu and Restrepo, 2016; Autor and Salomons, 2018; Brynjolfsson et al., 2018; Raj and Seamans, 2018; Eloundou et al., 2023), as well as demonstrating that these technologies are not simply labor displacing (Agrawal et al., 2019; Gregory et al., 2022). Instead, they are also likely to generate new jobs and new types of jobs (Bessen, 2019; Autor et al., 2022), and a key theme of the recent literature in this area is how humans can be most effective when working alongside algorithms (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 use of algorithms by organizations is the decentralization of "algorithmic expertise" among the firm's domain experts. To do so, it focuses on two types of human capital. *Algorithmic expertise* refers to skills related to the use of tools that take in data and produce decision output in the pursuit of business goals.[1] This definition of "algorithms" includes data science and AI tools but excludes tools like databases or web technologies which do not explicitly make decisions. *Domain expertise* refers to the knowledge required to work in a specialized field such as nursing, sales, marketing, or accounting. Prior work suggests that the effective application of algorithms may be unique in its demand for both domain and technical expertise (Collins, 2004), 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).

Employing domain experts with algorithmic expertise differs from a workforce structure in which technical expertise is centralized in specialized technology (IT) workers, and in contrast to job displacement arguments, it focuses on organizational adaptation to technological change that occurs at the sub-occupational level (Spitz-Oener, 2006). It also underscores the argument that as a general purpose technology, AI will increasingly become a central component of all work. To explain these changes, this paper develops theory that builds on the literature on the economics of job design and it considers how algorithms differ from other information technologies like databases or websites (Smith, 1776; Becker and Murphy, 1992; Dessein and Santos, 2006; Teodoridis, 2017; Lindbeck and Snower, 2000). It generates hypotheses related to i) how firms adapt jobs when using algorithms, ii) how the relevant expertise will be decentralized among non-technical occupations, and iii) how these adjustments are likely to be associated with organizational performance.

These hypotheses are tested using databases on corporate hiring and employment. The first database, which captures a "near-universe" of job listings issued by US firms between 2013 and 2016, has been used in prior work on the changing skill requirements of jobs (Deming and Kahn, 2018; Acemoglu et al., 2022) and to track the spread of new technologies (Goldfarb et al., 2023). The second database is a seven-year panel of how technology skills have diffused across occupations from 2015 to 2021 in a large sample of public firms.[2] This latter database is combined with administrative data on the knowledge content of occupations from the Bureau of Labor Statistics O*NET database and with corporate financial data from the Compustat-Capital IQ database.

The analysis produces four findings. First, using the job listings data, I show that employers were increasingly searching for algorithmic expertise across a broad array of occupations in a pattern that more closely resembles general-purpose office software skills (e.g. word processing tools) than more technical skills like database administration. By 2016, only one-third of the skills that reflected algorithmic expertise were embedded in listings for IT workers. They were particularly likely to be embedded in listings requiring domain expertise and the combination of these was particularly important where workers are expected to make decisions.

These findings from the job listings data reflect changes in employer preferences, but not whether the workforce met these changes in demand. The second analysis demonstrates similar changes in the corporate employment data for an overlapping seven year panel (2015-2021). In pub-

[1] I use the term "expertise". Whether jobs require literacy or expertise with the technology is an important distinction 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.

lic firms, algorithmic expertise increasingly spread among business-facing decision-makers which is in contrast to what we observe for other information technologies. These two findings are principally descriptive, but a third finding, which can be interpreted causally under certain assumptions discussed below, is that software innovations that lower the costs of working with algorithms, such as no-code tools, increased the likelihood that employers bundled algorithmic and domain expertise.

The fourth and final analysis turns to the organizational level and indicates that the decentralization of algorithmic expertise, in tandem with the complementary technological investments, are generating productive, intangible assets for public firms. Financial markets assign higher value to investments in AI and data science when this expertise is distributed among the firm's business decision-makers who know how to apply these technologies to business goals. This interaction is particularly important for AI technologies in the later part of the panel. Robustness tests suggest that the higher values that markets assign to these assets are unique to the combination of investment in algorithms and the decentralization of algorithmic expertise. No similar patterns are evident for other technologies or for employee expertise in other technical skill categories.

This study contributes to two streams of academic literature. First, with its focus on employers, it contributes to a literature identifying management complements to investment in new information technologies (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 view of IT as a technology that automates "routine" tasks but the application of algorithmic technologies to contexts where decision rules are not easily mapped to software has renewed the discussion on how IT affects firms' labor demand (Brynjolfsson et al., 2018). In doing so, this paper contributes to an emerging literature that examines management practices that complement investment 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 adoption of algorithmic decision-making will shape the future of work, which is becoming increasingly important 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 on technical skills has focused on the IT workforce (Ang et al., 2002; Mithas and Krishnan, 2008; Wiesche et al., 2019), but there has been some work on the implications of technical skills for broader workforce outcomes Atasoy et al. (2016); Deming and Noray (2020). The absence of more work in this area is notable given the growing demand from students and workers from all backgrounds for "coding" and other technical skills. These findings, therefore, contribute to our understanding of how the human-algorithm connection will shape the demand for skills as employers embrace these technologies.

2 Theory and Hypothesis Development

2.1 Algorithmic and domain expertise

The economics literature argues that tasks are organized into jobs according to three factors: specialization, coordination, and adaptation. How algorithmic technologies like AI and data science affect the design of jobs may depend on these three factors. Specialization allows for productivity gains, as in the medical field where AI algorithms analyze medical images, freeing radiologists to concentrate on complex cases (Smith, 1776). AI can lower coordination costs by enhancing the synchronization of interdependent tasks, exemplified by supply chain management where predictive models can 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). An example is when marketing professionals leverage AI to create personalized advertising based on consumer behavior data. These adaptive capabilities are particularly valuable where the application of local knowledge is critical.[3]

Balancing specialization and coordination has been a central theme of the literature on IT and jobs because new technologies incentivize employers to adjust the mix of skills *within* occupations (Spitz-Oener, 2006). Although specialization yields greater productivity in many contexts, Lindbeck and Snower (2000) theorize that task-based complementarities in knowledge-rich jobs have shifted work away from specialization towards more "holistic" forms of work in which workers handle a diversity of tasks. Multi-task work raises productivity when there are informational complementarities among tasks because productivity in one task can be interdependent with activity levels in others (Postrel, 2002).[4] Computers complement educated workers because by automating routine tasks, they raise the productivity of front-line workers who can balance a diverse set of tasks (Autor et al., 2003; Berman et al., 1994; Bresnahan et al., 2002; Bartel et al., 2007).

The balance between specialization and coordination has broad implications for IT and job design, but this analysis focuses on data-driven workflows which are increasingly common in organizations and where effective synthesis of domain and technical expertise has been a persistent challenge. For example, consider one well-recognized and standardized process used to balance data modeling decisions with business objectives, "CRISP-DM" (Cross Industry Standard Process for Data Mining) (Wirth and Hipp, 2000; Chapman et al., 2000).[5] This 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. Domain expertise in CRISP-DM is conceptualized as residing outside of technical expertise and being 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 have identified coordination costs across stakeholders as a key weakness of this paradigm (Saltz, 2021).

[3]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.

[4]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.

[5]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.

Both the importance and the difficulties of embedding domain expertise in a data-driven decision process may be further amplified when using data science and AI technologies because these technologies i) directly produce decisions as output and ii) these decisions are not always cognitively "routine" (Brynjolfsson et al., 2018; Agrawal et al., 2018). For non-routine decisions, there can be high costs to separating technical and domain expertise, as in the case of a radiologist working with AI-based diagnostic recommendations.

2.2 Hypothesis development

Reflecting this tension, prior work has confronted the challenge of injecting domain expertise into the data-driven workflow (Mao et al., 2019; Choudhury et al., 2020; Park et al., 2021). This work argues that the iterative nature of data exploration, experimentation, and learning required for data science favors generalists, who have a diversity of skills, over specialists (Colson, 2019). The emphasis on decision-making, the need for iteration, and the non-routine mapping between the input data and output decisions sets these technologies apart from other information technologies where the output (e.g., websites, cloud storage, databases) is an input into an employee's decision-making process rather than a decision itself. A salient example is the "data scientist" job title itself, which by definition, combines technical and statistical skills with domain expertise (Davenport and Patil, 2012; Provost and Fawcett, 2013). The importance of domain expertise for effective data science has been discussed online[6], in industry panels[7], and in the press (Oostendorp, 2019).

But beyond data scientists, workers who can couple domain expertise with technical skills are becoming important to many algorithmic decision-making contexts (Jha and Topol, 2016).[8] Users of machine learning tools in high-stakes contexts must evaluate trade-offs when choosing which data to include in a model, how to construct model features, or how to assign value to the costs of prediction errors (Kleinberg et al., 2018; Cowgill, 2018; Cowgill et al., 2020). Research situated in pharmaceutical industries has shown the importance of embedding the relevant human capital in downstream occupations (Wu et al., 2019) and in healthcare, Jha and Topol (2016) and Lebovitz et al. (2022) describe the challenges healthcare workers face when interpreting the accuracy of machine learning tools and output. The first hypothesis tested in this paper is:

> ***H1***: *Skills related to algorithms are more likely to be bundled with domain expertise than those related to other information technologies.*

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

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

[8]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.

This hypothesis, based in the balance between specialization and coordination, is about whether these skills are likely to be bundled, but does not make predictions about which occupations will receive this bundle. Adaptation provides a context in which to theorize about the control of task bundles in work environments. An instructive example here is "typing pools" which existed solely to provide typing services within the organization. Over time, the typing task eventually became part of the knowledge worker's job because local adaptation is important when creating text documents.[9] Similarly, where domain expertise helps with localized decision-making, organizations may prefer that non-technical domain experts – like those in finance and human resources – receive these skills. Conversely though, AI can also substitute for some forms of domain expertise in areas such as foreign language translation which may move the bundle away from domain experts. Which occupations get the bundle, therefore, is ultimately an empirical question.

> ***H2****: Algorithmic skills are likely to be bundled among non-technical occupations where domain expertise and decision-making are both important.*

These trade-offs are not static. The job design considerations discussed above depend upon the costs of becoming proficient with algorithmic tools. If the costs of acquiring technical skills are high, it will be difficult and expensive to find domain experts with technical expertise and employers may forego any productivity gains associated with bundling these skills. On the other hand, the barriers to use for many tools is falling as producers compete to speed adoption of their products in the workplace. Examples of this are prevalent, and include the embedding of complex logic in standardized software packages (Rock, 2019), the proliferation of no-code tools like Tableau, and most recently, the growing conversational abilities of large language models like OpenAI's *Data Analyst GPT* that enable workers to do data analysis with virtually no coding background.

> ***H3****: Algorithmic skills are more likely to diffuse into non-technical occupations as the cost of acquiring these skills falls.*

Firms may have strong incentives to pursue these workforce changes. Prior work has shown that productivity-enhancing workforce adjustments are needed to realize financial returns to IT investments (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 computing technologies that can perform routine tasks, the literature has shown that allocating decision authority to front-line decision makers yields higher productivity levels (Bresnahan et al., 2002), and particularly in turbulent environments where the value of decisions depends on rapidly changing external conditions (Mendelson and Pillai, 1998; Pavlou and El Sawy, 2006; Black and Lynch, 2001; Bresnahan et al., 2002).

For AI and data science technologies that make non-routine decisions, investors may anticipate greater value from firms that also employ personnel who understand how these tools can be applied to business goals. These adjustments may be costly, in the form of employers having to navigate more competitive labor markets to hire workers with these skills, but higher values would reflect

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

the production of valuable intangible assets that the market expects to eventually yield a stream of benefits. The literature referenced 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.

> ***H4***: *Financial markets assign higher value to investments in algorithms when the complementary skills are dispersed among occupations that make business decisions.*

The next section describes the databases used to test these four hypotheses.

3 Data sources and key measure construction

3.1 Key data sources

To test these relationships, I use data sources which provide information on a) how employers are adapting jobs to algorithms and b) how the skill composition of the workforce is changing in response. I supplement these with financial data from public firms to assess how these technologies and workforce changes are connected to the value that investors assign to firms.

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 listings to measure when skills first appear in online job ads and how skills co-occurred in these listings with other skills.

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 collects and standardizes 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, job title, educational requirements, certifications required, and skill expectations for each job. Several studies have used these data to study labor markets (Hershbein and Kahn, 2018; Deming and Kahn, 2018), including how AI related skills spread across jobs and industries (Acemoglu et al., 2022; Goldfarb et al., 2023).

Lightcast associates each listing with a BLS O*NET code and employers are tagged with a North American Industry Classification Systems (NAICS) industry. Job openings list skills, such as

[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.

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 are not the same as job "requirements". Employers can omit skills from listings, some skills may be assumed but not listed, and successful candidates may not need all of the skills in a listing. Nonetheless, employers are likely to be thoughtful about the skills they place in listings because including or omitting a skill can attract or repel the wrong type of applicant.

The data collection process raises questions about which industries and occupations it covers. However, prior academic work has provided thorough information on the sampling properties of the data, so I do not duplicate those comparisons here.[11] Key findings from these comparisons are that these listings data are over-represented in computer and mathematical occupations, as well as management, health care, business, and financial occupations. They are a less robust indicator of job vacancies in blue-collar occupations.

3.1.2 Corporate employment database

The corporate employment data were collected from Revelio Labs, a workforce intelligence company.[12] Their databases are constructed from a variety of 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 analysis sample 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 present comparisons of these data with administrative data from the Bureau of Labor Statistics. Like the Lightcast data, these data are over-sampled in management, business, and technology occupations and under-sampled in areas such as agriculture and manufacturing which is consistent with the greater use of online professional platforms 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 specific technical skills, like "machine learning", diffuse across occupations and employers. Moreover, the data contain CUSIP identifier codes which allow employer records to be merged with external financial databases such as the Compustat-Capital IQ data.

[11]See, for example, Appendix A of Deming and Kahn (2018) who make thorough 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).

[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 after 2022, the skill data are imputed, rather than collected.

3.1.3 Supplementary data sources

To identify occupations requiring domain expertise, 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 financial and employment data from Compustat-Capital IQ, which was collected through the WRDS data service.

3.2 Construction of sample and key measures

3.2.1 Sample construction

For the job listings analysis, the sample includes all listings in the data set from the months ranging from January 1, 2014 to June 1, 2016 for a total of 30 months of job listings data. As documented in other work (Deming and Kahn, 2018), these data more heavily represent workers in white-collar occupations, and particularly workers in IT and management occupations. The number of listings for any given month ranges from just under 2 million listings to up to 2.5 million listings for a total sample across the 30 months of 60,769,351 listings. However, as detailed below, the regression-based analyses on this data restrict this sample to a single month and to job listings with a specific set of skills which significantly lowers the sample size for those analyses.

The Revelio workforce sample, which forms the core of this analysis, includes firm-occupation-year-skill counts for the years 2015 to 2021. To join these figures with financial data from Capital IQ, the sample is limited to public firms, producing a sample size of 7,198 firm-years. Table 1 reports summary statistics for a single year (2018) of this panel. The statistics are reported in logs because they are included in logs in the multi-variate regressions. Firms in this sample are large, with an average market value of over 57 billion dollars and almost 30,000 employees. The average firm in this sample has around 1,000 IT workers. Table 2 shows the distribution of these firms across NAICS 2-digit industries. Although there are firms in every major sector, the Manufacturing, Information, and Finance and Insurance industries together comprise almost 90% of the overall sample. The construction of key variables used in the analysis is described below.

[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.

3.2.2 Algorithmic expertise (ALG and ϕ^{ALG})

A key challenge when converting data on workforce skills into measures for empirical analysis is the mapping of skills to skill groups.[19] Recent published papers that uses large quantities of archival, digitally collected workforce data have used manual mappings of skills into technology categories. For example, Abis and Veldkamp (2024) manually assign skills to "Data Management", "Analysis", "Old Technology", and "AI" categories and Goldfarb et al. (2023) select a cluster of skills related to machine learning technologies for their analysis. Deming and Kahn (2018) curate words and phrases in the Lightcast data associated with different job skills, including cognitive, social, character, and computer categories. The literature on the impact of AI technologies on labor displacement has also generated their own rubrics for measurement (Brynjolfsson et al., 2018).

This analysis takes a similar approach, relying on the grouping of base-level skill categories into higher-level categories for technology measurement. However, it uses categorizations generated by the data providers themselves, who use clustering methods to group skill categories into different technology areas like "data science", "AI", or "Big data". Appendix B delineates the skill categories that fall into each of the technological categories used in this analysis. Examples of these skill categories include *machine learning*, *business analytics*, *julia*, and *natural language processing* and these categories themselves contain more detailed skills. For instance, *machine learning* is a skill category that includes skills like "deep learning" and "supervised learning" within it. Skills in the AI and data science categories are used to develop indicators of algorithmic expertise at the individual level and the organizational level.

At the worker level, a record (job listing) is denoted as having (requiring) algorithmic expertise (ALG, a binary indicator) if it has at least one skill that falls into this category.

At the organizational level, measures of algorithmic expertise are constructed by identifying major occupation groups that are in the top quartile of all occupations in terms of their domain expertise requirements and the importance of decision-making for that job.[20] Then, for these decision-makers within the organization, $\%ALG$ is computed as the fraction of these employees with algorithmic expertise. Finally, for each organization i in year t, a standardized measure of algorithmic expertise (ϕ^{ALG}) is computed as:

$$\phi_{it}^{ALG} = \frac{(\%ALG_{it} - MEAN(\%ALG))}{STD(\%ALG)} \tag{1}$$

Organizations where algorithmic expertise is decentralized among the firm's decision-makers have higher ϕ^{ALG} values. For the robustness tests, similar measures are constructed for other technological categories (e.g., ϕ^{CLOUD}).

[19] Indeed, because of growing interest in the "future of work", the construction of taxonomies that makes sense of emerging sources of skills data 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.

[20] These occupations are: '11' (Management), '13' (Business and Financial), '17' (Architecture and Engineering), '19' (Life, Physical, and Social Science), '23' (Legal) , '29' (Healthcare Practitioners and Technical).

3.2.3 Domain expertise (DOM) and Decision making (DMK)

Jobs are encoded according to whether they require domain expertise, defined as having knowledge of a specific discipline. The measurement of requirements for domain expertise in a job is constructed to be consistent with the measurement of algorithmic expertise. A binary indicator of whether a job requires domain expertise (DOM) takes the value 1 if an employee reports having at least one type of domain knowledge in their skill set where the list of potential domain knowledge areas is extracted from O*NET, which in its dictionaries, identifies the possible knowledge domains with which US-based jobs may require familiarity.[21] These domains are extracted from the "Knowledge" table in O*NET, which delineates "organized sets of principles and facts applying in general domains."[22] From the full list, *Computers and Electronics*, *Engineering and Technology*, *Telecommunications*, and *Mathematics* were removed because they overlap with measures of algorithmic expertise.[23]

The importance of *Decision-making* for an occupation (DMK) is also retrieved from the O*NET database which provides this measure on a scale of 1 through 7 for each six-digit occupation.

3.2.4 Additional job characteristics

Beyond algorithmic and domain expertise, the analysis uses indicators of skills related to *cognitive*, *social*, *character*, and *management* job attributes. The construction of these attributes was based on prior work that uses the Lightcast data source to construct these measures (Deming and Kahn, 2018). As with measures of algorithmic and domain expertise, records are coded as needing these attributes if the listing contains at least one related skill.[24]

3.2.5 Employers' IT investments

Obtaining consistent, firm-level measures of IT investment spanning multiple years has been a persistent challenge in the academic literature (Tambe and Hitt, 2012). IT investments are not consistently recorded on balance sheets, so scholars have leveraged alternative sources to create

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

[22] 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.*

[23] It is useful to contrast this approach with one in which jobs are identified as requiring domain expertise based solely on 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.

[24] 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.

proxy measures, such as hardware investment measures collected by marketing surveys, IT keywords referenced in legal filings, and IT employment or salaries (Lichtenberg, 1995; Brynjolfsson and Hitt, 1996; Tambe, 2014). The rationale behind the last approach is that human capital is the largest component of a firm's digitization investment and it has become even more important for AI and data science investment because much of that software stack is open-sourced, leaving no investment trail, and because much of the hardware is cloud-based and poorly measured by instruments that record the firm's owned servers and PCs.[25] Conversely, most frontier software requires technical expertise to install and maintain, so quantities of complementary, technical human capital may be the most accurate available proxy measure of a firms' technology investments.

This paper uses this approach to generate measures of firms' technology assets. It follows prior work where proxy investment measures are constructed as quantities or intensities of skills relevant to the technological domain (Tambe and Hitt, 2012).[26] This view of the dichotomy between technical workers and other non-technical occupations is similar to prior work that considers the employment of technically skilled workers as the main investment into the construction of digital assets that can be subsequently deployed by an organization to achieve its business goals (Hall et al. (2000) calls this "e-capital"). Investments in aggregate IT or its sub-categories (e.g., *Alg IT*, *AI*, *Data science*) are computed as the quantity of relevant skills in the firms' IT workforce in a given year. Because the firm-level regressions include employment measures, the "stock" of skills in a technological domain can be interpreted as the intensity of investment in that domain.[27]

3.2.6 Financial variables, assets, and industry classification

The Compustat-Capital IQ data are used to construct employer-year measures for total market value, employment, industry classification, the value of PPE (property, plant, and equipment), and other assets. As discussed above, the use of Capital IQ financial data necessitates limiting the sample to public firms. Industry variables for these firms are retrieved at the four-digit NAICS level (North American Industry Classification System). Total market value is 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.

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

[26] Like most firm-level measures, this approach records investments with measurement error. See Appendix B for a brief discussion.

[27] The main findings are robust to an alternative construction of this measure based on quantities of technical workers with at least one skill in the relevant domain which has a slightly different interpretation (e.g. quantities of AI engineers, rather than the intensity of AI skills in the tech workforce). Those results are not shown due to space constraints but are available upon request.

4 Results

4.1 Model-Free Evidence

4.1.1 The growth of algorithmic expertise in job listings

Figure 1b illustrates growth in the rate at which algorithmic skills appearing in listings within the sample period spanning the years 2013 to 2016. Each x-axis tick corresponds to one month and the y-axis is the coefficient estimate (β) from the logistic regression $ALG_i = \beta_t t_i + \epsilon_i$ where t is a vector of dummy variables for months since January 2013 and ALG indicates whether an algorithmic skill appears in a job listing. This figure indicates that the likelihood of an algorithmic skill appearing in a listing rises steeply during the early part of the sample before flattening out.

Figure 2a shows the extent to which specific technology skills, including but not limited to algorithmic skills, are bundled together with domain expertise in job listings from one month of the Lightcast data (January 2016). Skills with higher values (reaching further to right) tend to appear in jobs that also list domain expertise in their requirements. Skills colored dark blue are those that correspond to algorithms. Algorithmic skills are more commonly bundled with domain expertise and along this metric, appear to have more in common with skills like Excel and ERP systems that are commonly used by business-facing occupations. In Figure 2b, a higher value indicates skills more likely to appear outside IT occupations. Algorithmic skills, colored in dark blue, are more commonly found in job listings for non-IT occupations than other technical skills. Skills related to predictive analytics, data science, and data analysis are only slightly less dispersed than skills related to the Microsoft Office Suite, which is consistent with the claim that employers are increasingly bundling algorithmic skills in occupations where domain expertise is embedded. These figures indicate that employers are increasingly searching for algorithmic skills outside of IT specialist occupations.

4.1.2 Algorithmic expertise in business occupations

Job openings are valuable because they (i) indicate employer preferences and (ii) immediately reflect changes being made by employers. In that sense, they serves as a leading indicator of labor market changes. However, they cannot say whether the job listings indicate hard requirements or a "wish list" from employers, or whether the vacancies requiring these skills are even ultimately filled. Next, I turn to corporate workforce data to investigate whether changes in employer preferences were met by the workforce. The results from this analysis are shown in the four quadrants of Figure 3.

Figure 3a illustrates how the dispersion of different technical skills in non-technical occupations changes over time in these firms. The y-axis is the intensity with which a skill appears in these occupations and levels are depicted relative to their 2008 values. The trend for algorithmic skills, depicted in blue, indicates steady growth in the rate at which AI and data science skills have penetrated these occupations, which is consistent with the demand-side findings from Figure 1b. By 2021, these skills appeared in 10% more non-IT occupations than they did in 2010. In contrast, skills in technologies related to networks and the cloud became increasingly specialized. Fewer workers in these non-technical occupations needed skills related to these technologies. The incidence of mobile skills remained flat.

Figure 3b shows how the organizational measure of algorithmic skill defined in Equation 1 (ϕ^{ALG}) varies across industries and time. Unsurprisingly, it is highest in the Information, Professional Services, and Finance industries. This is consistent with evidence on the prevalence of these technologies in these industries as reported in the press (Lohr, 2024). Retail has climbed rapidly reflecting the growing use of consumer data for prediction. Levels are lower in Healthcare although they have been climbing, reflecting the growing use of data science and AI in healthcare.

Figure 3c depicts annual changes in ϕ^{ALG} where firms are separated into quartiles according to their market values in the final year of the sample (2021). ϕ^{ALG} is highest in higher value firms and the differences are largest between the top and bottom quartile in the earlier years of the sample, as would be consistent with an environment where workers with these skills are a scarce resource that higher value firms are better positioned to attract. In the last few years of the sample, however, ϕ^{ALG} converges across quartiles, suggesting that supply-side adjustments have made it easier for employers with less resources to attract these workers. The fourth quadrant (Figure 3d), using data from the final year of the sample, plots firms' investment in algorithmic technologies against ϕ^{ALG} where the size of the bubble reflects the market value of the firm. The largest circles, colored in blue, are those commonly referred to as "big-tech" firms. We can see that firms tend to contemporaneously invest in these technologies and in employing business-facing workers who have the skills to apply these technologies to business problems.

4.2 Correlation tests between algorithms and domain expertise in job listings

The model-free evidence presented above suggests that both employers and non-technical workers have been shifting towards greater algorithmic expertise. To formally test the hypotheses presented in the theoretical discussion, I embed these measures in a regression framework.

Figure 4a summarizes tests of whether algorithmic skills are more likely than other technologies to appear in job listings requiring domain expertise. The sample is a single month of data (Jan 2016), and the unit of observation i is the job listing. The dependent variable is a binary indication of whether a listing i requires the applicant to have job attributes related to domain expertise.

$$DOM_i = \alpha_A ALG_i + \alpha_D DAT_i + \alpha_N NET_i + \gamma_i + \epsilon_i \tag{2}$$

The right-hand side includes measures of whether the job listing reflects a need for algorithmic skills as well as, for comparison, skills in two comparison technologies, databases and networks. For comparison, other models substitute as the dependent variable i) cognitive attributes, ii) social skills, iii) character, or iv) management skills. Equation 2 also includes a vector of control variables (γ) that includes detailed job title[28], 4 digit NAICS industry, and a measure of the logged total number of skills in the listing where i indexes the listing.

[28]The data provider standardizes job title at a granular level. Exemplar job titles are "Inventory Clark" and "Ux Developer".

Figure 4a depicts estimates of α_A from Equation 2. There are positive correlations between algorithmic skills and domain expertise. Because job-title fixed effects are included, these correlations mean that algorithmic skills are particularly likely to be bundled in jobs requiring domain expertise (**Hypothesis 1**). The full form of the regression estimates, which can be found in Appendix C, indicates that correlations with database tasks are negative, which is consistent with that class of skills being more specialized within IT work. We can also observe positive correlations between the use of algorithmic technologies and cognitive skills and negative correlations with management-related job attributes. This negative relationship suggests employers are not bundling people leadership (character and management) with algorithmic skills.

The next analysis turns to the question of where in the data decision-making pipeline this combination of skills, domain expertise and algorithmic skills, is valuable: i) data management, ii) data modeling, iii) analytics, iv) visualization, v) decision-making, and vi) presentation. In the logistic regression used to evaluate this relationship, *DATA.TASK* corresponds to any of these pipeline stages, generating six separate regressions.

$$DATA.TASK_i = \beta_D DOM_i + \beta_A ALG_i + \beta_{DA}(DOM_i \times ALG_i) + \gamma_i + \epsilon_i \tag{3}$$

Figure 4b illustrates β_{DA} computed from these regressions. The estimates on *Presentation* and *Decision-making* indicate that employers are increasingly searching for candidates with algorithmic expertise and domain expertise where workers also need to make decisions. In contrast, this combination of skills is negatively correlated with tasks related to *Data management* and *Data modeling*. These tasks require highly specialized technology skills that may call for specialists rather than generalists who can incorporate domain expertise while performing the task.

4.3 Algorithms and domain expertise in business occupations

Turning towards workforce composition, Table 3 reports correlations between whether significant decision-making required in non-technical occupations and whether algorithmic expertise has diffused into them, accounting for firm, occupation, and year differences. All regressions are at the firm-occupation-year level using a panel spanning the years 2015 to 2021. Column (1) indicates that occupations that require greater decision-making are more likely to report having algorithmic skills (t=3.50) (**Hypothesis 2**). This remains true after including firm fixed-effects (column (2)) which mitigates the concern that firms that rely more heavily on algorithms simply have more workers who make decisions.

The relationships discussed to this point are principally correlational, but columns (3) and (4) test the proposition that a fall in the cost of using these tools accelerates the diffusion of these skills across occupations. Columns (3) and (4) separate algorithmic skills according to whether the skill corresponds to a "no-code" technical tool. No-code tools are a minority of algorithmic skills so the main-effect is negative (t=-28.88). Column (4), however, indicates that tools with lower costs of use accelerate more quickly into decision-making occupations (t=4.29) (**Hypothesis 3**). This finding is supportive of a causal relationship under the assumption that these tools essentially perform the same function as their code-based counterparts. If these two categories of tools perform different functions, it could reflect demand-based differences and threaten a causal interpretation.

Columns (5) and (6) aggregate the data by occupation and test the following relationship at the organizational level: $\phi_{it}^{ALG} = ALG_{it} + \gamma_{it} + \epsilon_{it}$. A higher ϕ^{ALG} measure indicates that a greater fraction of workers in the firm's non-technical occupations require domain knowledge and algorithmic expertise, i and t are the firm and year respectively, and γ is a vector of controls that account for differences in size, assets, employment, and industry. The estimates show correlations between algorithms and ϕ^{ALG}, after conditioning on other firm characteristics like size and industry. This also persists after firm fixed-effects are included (t=2.34), which means that firms that rely on algorithms have a greater degree of algorithmic expertise in decision-making occupations, even after accounting for other static sources of firm differences. The findings in columns (5) and (6) are consistent with the assertion that employers see value in coupling algorithmic investments with personnel who have the skills to understand how these algorithms can be applied to non-technical business objectives.

4.4 Algorithms, business occupations, and financial value

Another test of whether these investments are complementary is by looking at how they are valued by financial markets. The main regression is a test of a firm's market value on its assets. Prior work has used a hedonic market value framework to decompose a firm's value into its component parts and to test whether financial markets can be used to uncover the presence of valuable but otherwise intangible assets that contribute to firms' productive capacities (Brynjolfsson et al., 2002). Market value is also a particularly useful dependent variable for an analysis of AI and data science returns because firms require time to adjust new technologies to their workflow and the literature suggests that firms are not yet consistently realizing value from data science and AI investments. Investors, however, assign value to assets based on the future stream of benefits they will produce. The key regression used to test for the presence of workforce complements to algorithmic decision-making is:

$$\begin{aligned} Log(MV)_{it} =& Log(AT)_{it} + Log(PPE)_{it} + Log(IT)_{it} + \\ & Log(ALG)_{it} + \phi_{it}^{ALG} + (Log(ALG)_{it} \times \phi_{it}^{ALG}) + \gamma_{it} + \epsilon_{it} \end{aligned} \tag{4}$$

In equation 4, i indexes the firm and t is year. MV is the firm's market value, and PPE and AT are capital and other assets, respectively. IT is a proxy measure of the firm's aggregate IT investment, ALG is a proxy measure of the firm's investment in algorithms, ϕ^{ALG} is the standardized organizational measure of algorithmic expertise, and γ_{it} is a vector of fixed-effects including year, employment size and depending on the specification, industry at the four-digit NAICS level or employer fixed-effects.

Table 4 reports results from estimating Equation 4 on a seven year panel of public firm investments (2015-2021). Columns (1) through (3) in Table 4 have year and 4-digit industry fixed-effects. Column (1) reports results of market value on measures of IT investment, assets, capital (PPE), and employment. All of these variables are entered in logs. As has been found in prior work, the estimates suggest that financial markets assign economic value to investment in digital technologies (t=3.79). The next column adds a measure of specific investment in algorithmic technologies. After adding this measure into the regression, the coefficient on general IT capital falls to zero which suggests that the market returns to IT investment are principally from firms that invest at the frontier, as represented in this panel by AI and data science investments (t=3.93).

Column (3) reports estimates from the full form of equation 4 that includes the interaction between algorithms and ϕ^{ALG}. The main effect of algorithms is similar to column (2) (t=3.80) but the interpretation of the interaction term is that these investments are valued an additional 15% higher in firms where ϕ^{ALG} is one standard deviation higher (t=2.67) (**Hypothesis 4**). These estimates support the hypothesis that firms contemporaneously investing in ALG and ϕ^{ALG} are building valuable intangibles that will be useful for producing a stream of AI goods and services in the future. Column (4) substitutes firm effects for industry effects. Including firm effects drives the coefficients on algorithms, ϕ^{ALG}, and the interaction term to zero. One interpretation of this finding is that there are high adjustment costs for firms trying to build these assets. Within the relatively limited range of this panel, unobserved differences across firms explain most of the heterogeneity in firms' abilities to build these assets. One way to further probe this argument is to separate extensive and intensive margins of investment for this technology. Column (5) substitutes a measure indicating whether ϕ^{ALG} is above or below the mean for firms in the sample. The results are similar to those in column (3) (t=3.00). This binary construction eliminates most within-firm variation and shows that across-firm differences are more important for these estimates than firm's abilities to change their asset mix within the relatively small number of years covered in this sample.

These analyses suggest that within the last decade, (i) employers have been adjusting candidate search to find domain experts with expertise in algorithms, (ii) skills related to algorithms spread to business-facing occupations, and (iii) employers that made these investments jointly with matching technological investments realized higher market values, suggesting the presence of valuable intangible assets in these firms. Together, this evidence supports the primary conclusion of the paper that greater level of technical skill in a firm's business and management layer is a valuable complement to its use of algorithms in the decision-making process.

The average effects computed in the sample above are likely to obscure heterogeneity along different dimensions. Figure 5 reports estimates on the main interaction term for Equation 4 where the sample is split by employment size category (separated by tercile). These estimates indicate that the financial rewards associated with investment in workers with these skills are higher for larger firms in the sample. This may reflect the higher coordination costs faced by workers in large firms or alternatively, it may reflect the advantages that larger firms enjoy when hiring such workers or installing these technologies. The first explanation can be expected to persist over time, but the second could fade as the supply side of the market adjusts as suggested by Figure 3c.

Turning towards the temporal dimension, the analysis until this point has grouped data science and AI investments into a single category even though the frontier of IT investment has rapidly moved from data science to AI during the course of the panel. Table 5 separates these investments and divides the panel into an earlier and later time period comprised of the years before and after 2018. The first two columns suggest a relationship between data science investment and the firm's market value during the early period (t=5.21). There is no correlation with AI in this period, however. In the later period though, the estimates on both AI (t=3.19) and Data Science (t=2.95) are positive and significant suggesting a greater role for AI investments in driving value in the later period. Columns (3) and (4) introduce interaction terms with organizational skills for each of these technologies (ϕ^{DS} and ϕ^{AI}). In the early period, neither coefficient suggests meaningful correlations. The estimates from the more recent period, however, suggest that a one standard deviation higher ϕ^{AI} measure raises the value of AI investment by one-third (t=3.18). Notably, the same pattern does not emerge for data science which may suggest that the bundling with domain expertise matters more for AI, as a fully-automated decision maker.

4.5 Robustness tests

The market value regressions in the prior section provide primarily descriptive evidence that firms' use of algorithms is benefited by a broad base of algorithmic expertise. The broader literature from which it derives has primarily focused on the estimation of component values from an aggregate value measure rather than the development of causal interventions. However, this section reports additional tests to provide evidence that these correlations are likely due to the hypothesized theoretical relationships rather than omitted variable bias. Figure 6 summarizes these tests.

Figure 6a alters the type of technology investment. It shows the coefficient estimates on the interaction term on $(ALG_{it} \times \phi_{it}^{ALG})$ from Equation 4 except that it substitutes investment in other technologies. The theoretical discussion indicates that we should only observe meaningful correlations for this organizational skill term with algorithmic investment, not investments in other technologies. This is the pattern we observe in Figure 6a. Neither of the coefficients on the interaction terms created with other technology measures exhibit meaningful correlations. This pattern supports the argument that a correlation between market value and the interaction term between AI skill and AI investment is not simply reflecting another type of heterogeneity that would be picked up by other measures of technology investment, such as differences in general financial resources, free cash flow, or overall digital intensity.

Figure 6b performs a similar comparison, again using the specification in Equation 4 as a baseline, but instead of altering the technology investment measure, it retains AI as the technology measure for all interaction terms and varies the construction of ϕ^{AI}. Instead, it generates this dispersion using the prevalence of network and cloud technology skills in these non-technical occupations, rather than AI skills. Again, we only observe correlations with market value when investment in AI technologies is accompanied by AI skills in the decision-making layer of the workforce.

As a final placebo test, Figure 6c returns to the original construction of ϕ_{AI} but it only uses occupations *lowest* in their decision-making importance, rather than those that are highest. This comparison reveals that correlations with market value are stronger when constructing this skill measure from occupations where decision-making is critical. In sum, the results in Figure 6 indicate that correlations with market value only appear at the confluence of AI investment and AI skills in the business and management workforce. We should be cautious before interpreting this relationship causally, but it at least suggests that investments in algorithms and in the workers who can apply them to business decisions are disproportionately accumulating in high value firms. We do not observe similar patterns of resource accumulation for other types of technology or technical skills in employees who are not instrumental to the firm's decision-making.

5 Managerial Implications

Adoption of algorithmic decision-making, and particularly predictive AI applications, has been difficult and uneven. The evidence from this analysis suggests that the human capital of data-driven firms differs from firms that lag in this domain. This implies that firms face considerable adjustment costs when adopting these technologies. High adjustment costs imply competitive rents for firms that have successfully found the right mix of workers.

A corollary is that the costs of using AI and data science technologies are continuously falling. Conversational interfaces driven by generative AI, for instance, represent a shift in how knowledge workers interact with information technologies, making them more accessible and user-friendly than ever before. A fall in the costs of using these tools suggests that employers can 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, no-code and generative AI tools can democratize technical skill, enabling a more diverse range of employees to contribute to areas that were once the exclusive domain of technical specialists. This can lead to more innovative environments that emphasize the productive combination of human creativity and computational power. Managers, in turn, may 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 skill has been shown to have economic attributes that differentiate it from other types of expertise. For instance, frontier technical skills derive significant productivity benefits from geographic agglomeration (Saxenian, 1996). Moreover, rapid technological depreciation 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 some form of technical expertise, 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 could suggest a curricular reorientation. Although technical skills will continue to remain important for specialized workers in IT-producing industries, there may be greater emphasis from IT-using industries 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 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 expertise is becoming broadly dispersed across domain experts in effective organizations, ii) that this dispersion is due to complementarities that arise between algorithmic skill and domain expertise, and iii) that the market assigns higher value to firms that concurrently make these workforce adjustments while investing in algorithmic tools. In doing so, it documents one early but important facet of the workforce transformation occurring to support the use of algorithms in the organization.

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

Beyond these limitations, there is significant scope for future work in this area. We are at the beginning of a 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 much to be learned about how to design organizations so that humans can effectively work with algorithms. Although this paper considers one facet of workforce transformation, complements to algorithmic technologies will be wide-ranging. These will likely include even more sweeping changes to workforce skills, as well as other non-labor investments to support these capabilities. Firms' information capabilities will also continue to evolve and algorithms will become easier to deploy as software and tools progress, 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 adoption stage, there is 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 requires 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

Abis, S. and Veldkamp, L. (2024), 'The changing economics of knowledge production', *The Review of Financial Studies* **37**(1), 89–118.

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.

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* .

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.

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* .

Hall, R. E., Cummins, J. G. and Lamont, O. A. (2000), 'E-capital: The link between the stock market and the labor market in the 1990s', *Brookings Papers on Economic Activity* **2000**(2), 73–118.

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.

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.

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 scientists work 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.

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* .

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.

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

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. and Hitt, L. M. (2012), 'The productivity of information technology investments: New evidence from IT labor data', *Information systems research* **23**(3-part-1), 599–617.

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

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.

Table 1: Summary statistics for regression panel variables (2018)

Variable	Units	Mean	Std. Dev.	N
Log(Market value)	Millions (USD)	9.120	1.83	1250
Log(Assets)	Millions (USD)	8.503	2.18	1250
Log(PPE)	Millions (USD)	5.754	2.57	1250
Log(Employment)	Thousands (Employees)	2.218	1.44	1250
Log(IT)	Skill count	7.065	1.92	1250
Log(Networks)	Skill count	3.370	2.04	1250
Log(Databases)	Skill count	5.062	1.86	1250
Log(Algorithms)	Skill count	3.551	1.76	1250
Log(Data science)	Skill count	3.342	1.69	1250
Log(AI)	Skill count	2.032	1.78	1250
ϕ^{ALG}	Standardized Value	0.061	0.88	1250

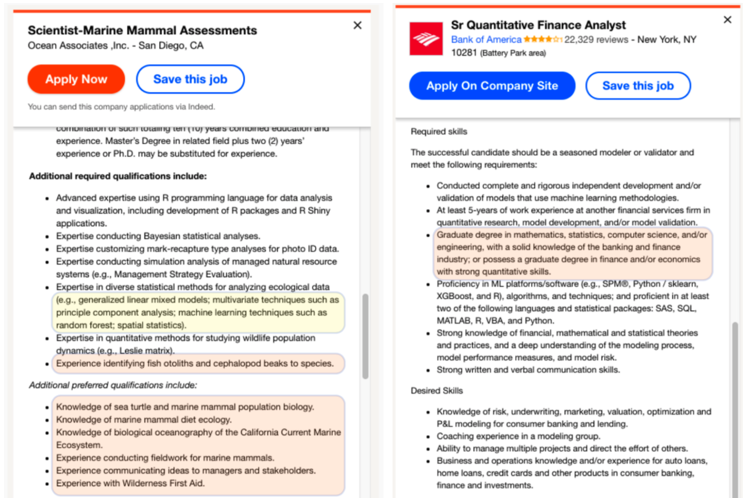
Table notes: This table reports summary statistics for firms in the 2018 cross-section of the regression panel constructed from the workforce data. The year 2018 was chosen as the midpoint in the panel window (2015-2021). The data source for the first four rows [*Market Value, Assets, PPE, Employment*] is the Capital IQ database available through Wharton Research Data Services (WRDS). The measures in the last five rows [*IT, Data, Network, Algorithms, AI, Data science,* ϕ^{ALG}] are constructed from the Revelio workforce database.

Table 2: Industry distribution of corporate workforce sample (2018)

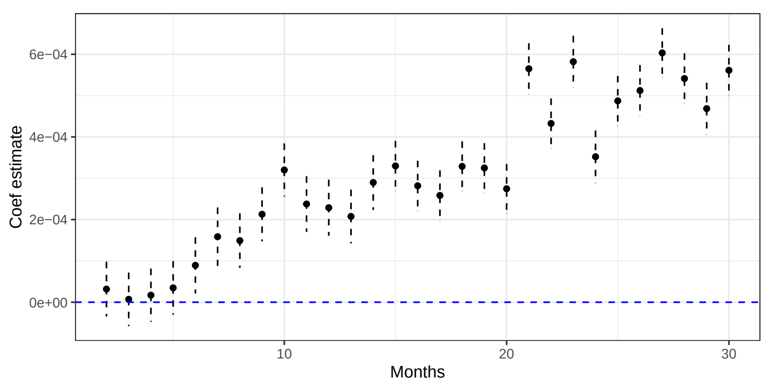
NAICS 2	Sector	N
11	Agriculture, Forestry, Fishing and Hunting	1
21	Mining, Quarrying, and Oil and Gas Extraction	29
22	Utilities	34
23	Construction	10
31-33	Manufacturing	360
42	Wholesale Trade	34
44-45	Retail Trade	42
48-49	Transportation and Warehousing	34
51	Information	268
52	Finance and Insurance	268
53	Real Estate and Rental and Leasing	26
54	Professional, Scientific, and Technical Services	60
56	Administrative and Support and Waste Management Services	25
61	Educational Services	6
62	Health Care and Social Assistance	24
71	Arts, Entertainment, and Recreation	3
72	Accommodation and Food Services	19
81	Other Services (except Public Administration)	1

Table notes: This table reports the distribution of firms across NAICS 2 digit industries in the 2018 cross-section of the regression panel. N = 1,244.

Figure 1: The growth of algorithmic skills in job listings



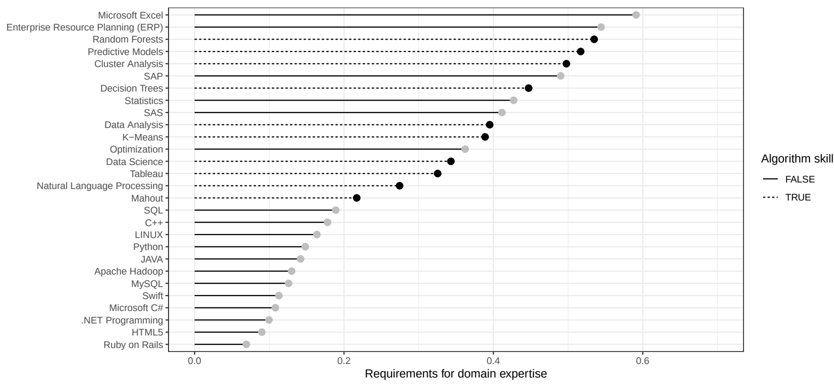
(a) Sample listings with algorithmic and domain expertise



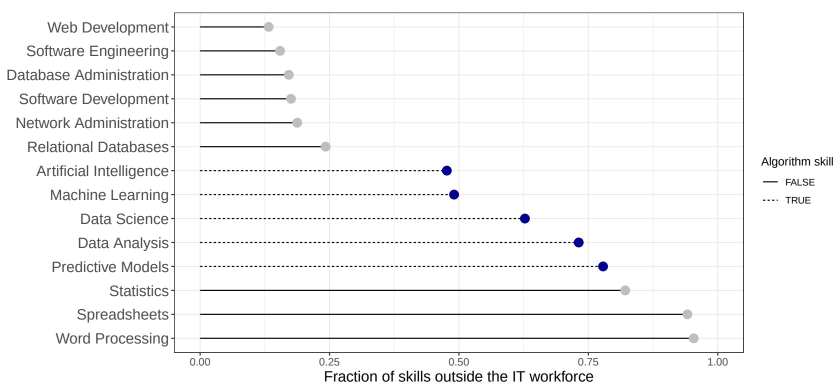
(b) Likelihood an algorithmic skill appears in a non-IT listing

Figure notes: Figure (a) shows two sample listings for jobs requiring 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 coefficient estimates and standard error bars on the regression $ALG_i = \beta month_i + \epsilon_i$ for the listings in the months covered by the Lightcast data (where coefficients reflect differences from the Jan 2013 baseline month), i indexes job listings, and ALG takes the value 1 if a listing contains an algorithmic skill and 0 otherwise. N=60,769,351. Standard error bars show the 95% confidence interval.

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



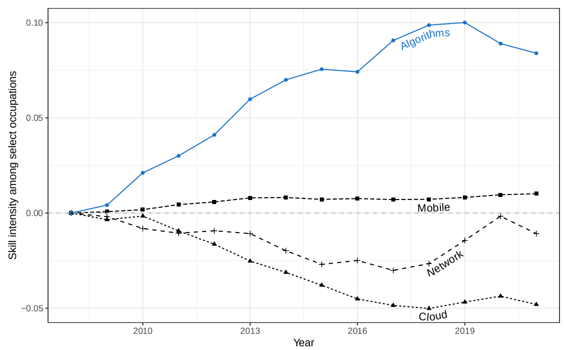
(a) Bundling of technical skills with domain expertise in job listings



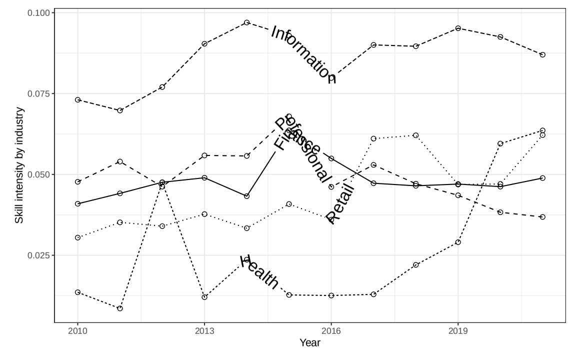
(b) Fraction of skills that appear in job listings for non-IT occupations

Figure notes: Figure (a) indicates the extent to which different technologies are bundled with domain expertise for any of these skills appearing in one month of the job listing data (January 2016) (N=763,986). Skills in dark blue (dashed line stems) are algorithms and all other technologies are shown in gray (solid line stems). Longer bars in this figure (reaching further to the right) indicate a skill that is more likely to be bundled with domain expertise. Figure (b) indicates the fraction that a technology appears in listings that are in a non-IT occupation. The sample is restricted to listings in one month (January 2016) of the sample data with skills in one of the shown areas (N=263,256). Skills in dark blue (dashed lines) are those associated with algorithms. A value closer to one means that a skill is more likely to appear in non-IT occupations.

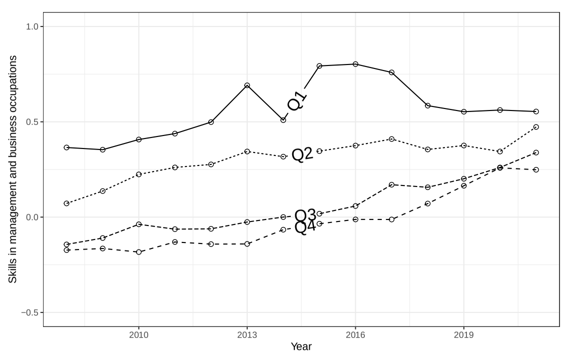
Figure 3: Changes in the locus of technical skills in organizations from 2008-2021



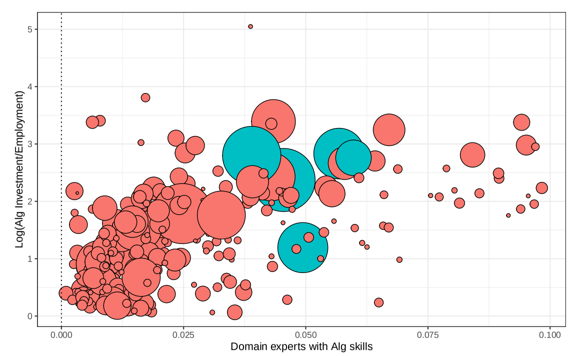
(a) Changes in skills in organizations



(b) Algorithmic skill measure by industry



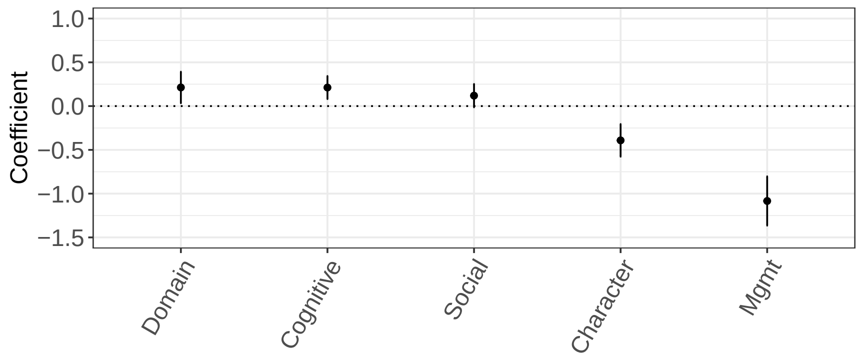
(c) Convergence in skill measure across firm size quartiles



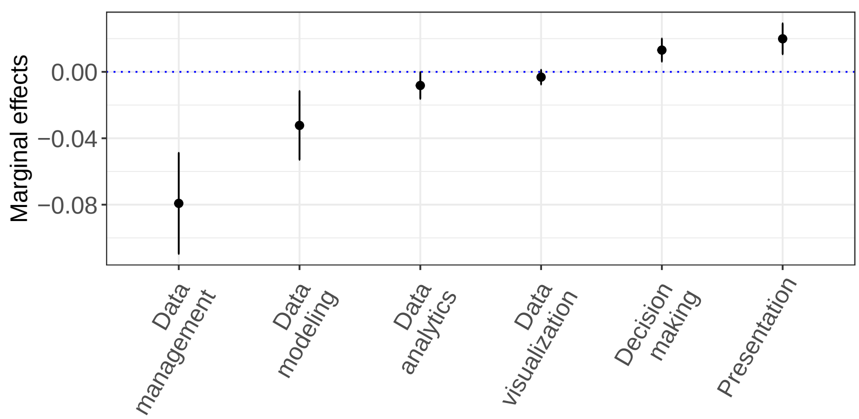
(d) Organizational accumulation of skills (AI)

Figure notes: Figure (a) illustrates changes in a measure of prevalence of different technical skills over time among the firm's domain experts. All trend lines represent changes from their values in the base year (2008). Figure (b) illustrates changes in the algorithmic skill measure over the course of the panel where firms are separated into different industries. Figure (c) shows changes in this measure in where firms are divided into quartiles by size. Figure (d) plots ϕ^{ALG} (x-axis) against AI investment levels (y-axis). The sizes of the bubbles in this figure indicate the firm's market value.

Figure 4: Characteristics of job listings with algorithmic skills



(a) Correlations between algorithmic skills and different job skills



(b) Marginal effects of domain expertise on different tasks in the data pipeline

Figure notes: Figure (a) depicts correlations between various skills needed on-the-job and algorithmic skills appearing in job listings. Each vertical bar is a coefficient estimate from a separate regression of the form $SKILL_i = \alpha_{ALG} ALG_i + \alpha_{DATA} DATA_i + \alpha_{NET} NETWORK_i + Log(No.\ Skills)_i + \gamma_i + \phi_i + \epsilon_i$ where for each of the five different regressions, $SKILL$ is one of $DOMAIN$, $COGNITIVE$, $SOCIAL$, $CHARACTER$, or $MANAGEMENT$, i indexes the listing, γ and ϕ are occupation and industry fixed-effects respectively, and $Log(No.\ Skills)$ is the logged number of skills in the listing. The point estimate shown is the coefficient on α_{ALG} from each regression and the vertical bars indicate 95% confidence intervals. Figure (b) reports results from a test of which tasks in the data pipeline are most likely to need a combination of both algorithmic and domain expertise. The logistic regression is $TASK_i = \beta_{DA}(DOM_i \times ALG_i) + \beta_D DOM_i + \beta_A ALG_i + Log(No.\ Skills)_i + \epsilon_i$ where DOM_i and ALG_i are binary variables indicating that a listing requires domain expertise or algorithmic expertise and the data tasks can be one of either *Data management*, *Data modeling*, *Data visualization*, *Decision making*, *Data analytics*, or *Presentation.* The estimate that is presented is the marginal effect of the β_{DA} coefficient from the logistic regression. Standard error bars show the 95% confidence interval.

Table 3: Conditional correlations between job attributes and algorithmic skills

	ALG				ϕ^{ALG}	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
DMK × NOCODE				0.120***		
				(0.028)		
DMK	0.072***	0.113***		0.028		
	(0.021)	(0.020)		(0.017)		
NOCODE			−0.924***	−1.381***		
			(0.032)	(0.100)		
Log(Occupational count)	0.451***	0.506***	0.269***	0.266***		
	(0.012)	(0.012)	(0.008)	(0.008)		
Log(Employment)	−0.005	0.079***	0.062***	0.062***	−0.049**	0.225**
	(0.012)	(0.021)	(0.011)	(0.011)	(0.023)	(0.109)
Log(Alg IT)					0.032***	0.095***
					(0.006)	(0.025)
Log(Assets)					0.020	0.058
					(0.020)	(0.049)
Log(PPE)					0.013	0.014
					(0.015)	(0.037)
Fixed-effects						
Firm FE		Yes	Yes	Yes		Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
O*NET 2 FE	Yes	Yes	Yes	Yes		
Industry FE (NAICS 3)	Yes				Yes	
Fit statistics						
R^2	0.276	0.378	0.320	0.321	0.107	0.428
Observations	160,614	160,614	321,228	321,228	8,280	8,279

Table notes: In the first four columns, observations are at the firm-occupation-year level at the level of the 6-digit Standard Occupational Classification (SOC) occupation and the regression model is $ALG_{ijt} = DM_{ijt} + Log(Occ\ Count)_{ijt} + Log(Employment)_{it} + \gamma_{ijt} + \epsilon_{ijt}$ where i is the firm, j is the occupation, and t is the year and γ and ϕ are industry, firm, year, and occupational fixed-effects. *Decision-making* is an indicator of the importance of decision-making for the occupation as recorded in the O*NET database for that 6-digit SOC occupation. *Occupational count* is the number of workers in that firm-occupation-year combination. *Employment* indicates firm employment levels and does not vary at the occupational level. *NOCODE* is an indicator of whether the skill is related to a "no-code" technology. Standard errors are clustered on the employer. For the last two columns, the regression model is $\phi_{it}^{ALG} = Log(Assets)_{it} + Log(Alg)_{it} + Log(PPE)_{it} + Log(Employment)_{it} + \epsilon_{it}$ where observations are at the firm-year level. *Assets*, *Employment*, and *PPE* are firm level measures from the Capital IQ database. ***p<.01, **p<.05, *p<.10.

Table 4: OLS regressions of algorithmic investment and organizational skills on market value

			Log(Market Value)		
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Log(Assets)	0.724***	0.720***	0.720***	0.563***	0.720***
	(0.057)	(0.058)	(0.058)	(0.036)	(0.058)
Log(PPE)	0.089**	0.089**	0.088**	0.081***	0.089**
	(0.045)	(0.044)	(0.044)	(0.026)	(0.044)
Log(IT)	0.049***	0.000	0.000	0.036***	0.000
	(0.015)	(0.018)	(0.018)	(0.012)	(0.018)
Log(Employment)	−0.001	0.000	0.001	0.228***	0.000
	(0.047)	(0.046)	(0.046)	(0.046)	(0.046)
Log(Alg IT)		0.049***	0.047***	−0.007	0.042***
		(0.016)	(0.015)	(0.007)	(0.015)
ϕ^{ALG}			−0.010	0.000	
			(0.008)	(0.007)	
Log(Alg IT) × ϕ^{ALG}			0.008***	0.001	
			(0.003)	(0.003)	
$\phi - HIGH$					−0.019
					(0.017)
Log(Alg IT) × $\phi - HIGH$					0.013***
					(0.005)
Fixed-effects					
Firm FE				Yes	
Year FE	Yes	Yes	Yes	Yes	Yes
Industry FE (NAICS 4)	Yes	Yes	Yes		Yes
Fit statistics					
R^2	0.888	0.889	0.889	0.974	0.889
Observations	7,229	7,229	7,229	7,229	7,229

Table notes: This table reports regressions of how workforce skill composition relates to firms' market value on the full firm panel which ranges from 2015-2021. The regression model is $Log(MV)_{it} = Log(Assets)_{it} + Log(PPE)_{it} + Log(IT)_{it} + Log(Employment)_{it} + Log(Alg)_{it} + \phi_{it}^{ALG} + (Log(Alg)_{it} \times \phi_{it}^{ALG}) + \epsilon_{it}$ where observations are at the level of the firm-year. Columns (1), (2), and (3) all include year and 3-digit NAICS fixed effects but add progressively more variables. Column (4) uses the same specification as (3) but substitutes 4-digit NAICS controls. Column (5) substitutes firm-level fixed-effects. Standard errors are clustered on employer. N=7,198. ***p<.01, **p<.05, *p<.10.

Figure 5: $Alg \times \phi^{ALG}$ by employment size tercile

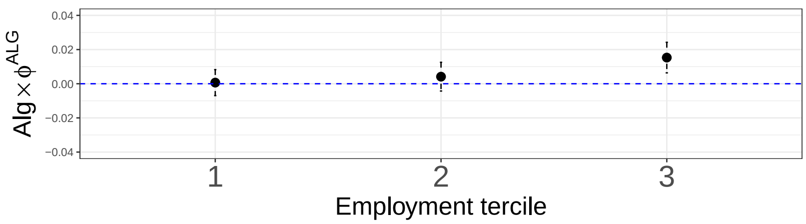


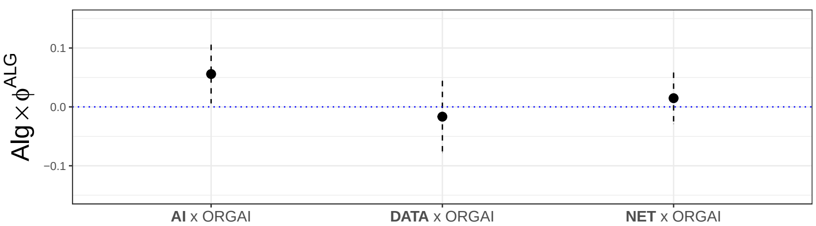
Figure notes: The y-axis indicates the coefficient on the interaction term between algorithm investment and ϕ^{ALG} from the main specification used in column (4) of Table 4 where the sample is divided into terciles by employment size. The sample size in each regression is approximately one-third the sample size used in column 4 of Table 4. The x-axis divides firms by employment quartile where "1" is the smallest firms in the sample and "3" is the largest firms in the sample. Standard error bars indicate the 95% confidence interval.

Table 5: Separating AI and data science investment over the panel (2015-2021)

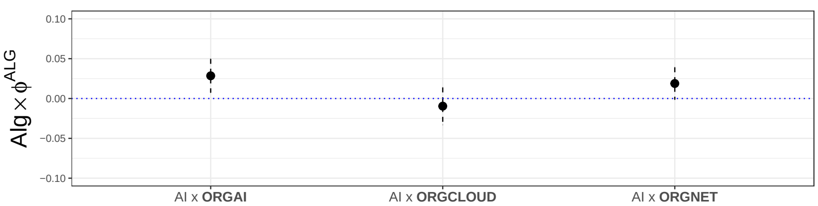
DV	Log(Market Value)			
Years	2015-2017	2018-2021	2015-2017	2018-2021
Model:	(1)	(2)	(3)	(4)
Variables				
Log(Assets)	0.733***	0.696***	0.733***	0.695***
	(0.051)	(0.044)	(0.051)	(0.043)
Log(PPE)	0.086**	0.073*	0.085*	0.064*
	(0.043)	(0.037)	(0.043)	(0.037)
Log(AI)	0.038	0.104***	0.040	0.098***
	(0.025)	(0.031)	(0.025)	(0.029)
Log(Data Science)	0.125***	0.105***	0.124***	0.095**
	(0.021)	(0.037)	(0.021)	(0.040)
Log(IT)	−0.103***	−0.103**	−0.103***	−0.090**
	(0.029)	(0.040)	(0.029)	(0.039)
Log(Employment)	0.000	−0.014	0.001	−0.001
	(0.048)	(0.047)	(0.048)	(0.043)
ϕ^{AI}			0.015	−0.068**
			(0.016)	(0.026)
ϕ^{DS}			−0.018	0.004
			(0.022)	(0.045)
Log(AI) × ϕ^{AI}			−0.008	0.031***
			(0.006)	(0.011)
Log(Data Science) × ϕ^{DS}			0.004	0.013
			(0.006)	(0.011)
Fixed-effects				
Year FE	Yes	Yes	Yes	Yes
Industry FE (NAICS 3)	Yes	Yes	Yes	Yes
Fit statistics				
R^2	0.903	0.853	0.903	0.856
Observations	4,033	2,475	4,033	2,475

Table notes: This table reports regressions of how algorithmic investment and skill composition measures relate to the market value of public firms across the first and second halves of the panel where algorithmic investment is broken into separate AI and data science investment and skill categories. Observations are at the firm-year level. The first and third columns use observations from the years 2015 to 2017 and the second and fourth columns use observations from 2018 to 2021. ϕ^{AI} and ϕ^{DS} are constructed in the same way as ϕ^{ALG} in Table 4 except on the restricted set of AI or data science skills, respectively. Standard errors are clustered at the firm level. ***p<.01, **p<.05, *p<.10.

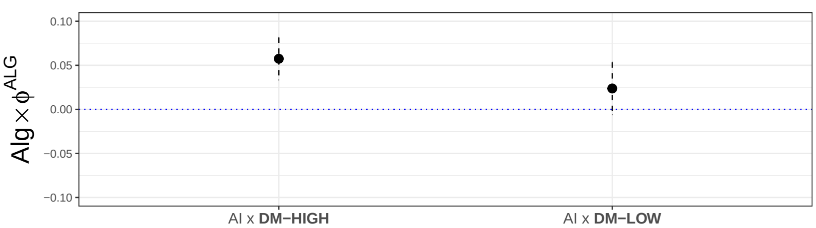
Figure 6: Placebo tests using alternative constructions of ϕ^{AI}



(a) Comparisons with investment in other technologies



(b) Comparisons with organizational measures of other technical skills



(c) Comparisons with measures of technology and other technical skills

Figure notes: The top facet illustrates placebo tests from interacting ϕ^{AI} with investment in other technology classes for the main regression panel. (N=1,308). It displays interaction terms from the regression in column (4) of Table 4 ($Log(MV)_{it} = Log(Assets)_{it} + TECH_{it} + \phi_{it}^{ALG} + (TECH_{it} \times \phi_{it}^{ALG}) + \gamma_{it} + \epsilon_{it}$). The marker on the left is for AI investment, the middle one is for networks, and the one towards for the right is relational databases. The middle facet takes a similar approach but uses measures of organizational skill dispersion in other technologies. These measures are based in AI, cloud, and networks. N=8,947. The bottom facet separates occupations used to construct the skill measure into those where decisions are important (on the left) and decision are unimportant (on the right). The standard error bars in both figures indicate 95% confidence intervals.

A Description of corporate workforce data

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

A.1 Data generating process and sampling frame

Revelio is a workforce intelligence company that federates data across a 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 contain data on the movements of an extremely large sample of US-based employees across firms, the job titles they hold, and the skills they acquire. Data on employment spells, at scale, are not otherwise collected by government agencies. They are only available through resume banks so these types of data are particularly useful for studying quantities of workers in firms with different skills and the flow of workers of different types between organizations.

On the other hand, there are some potential issues when using data sources of this type. Workers participate on professional networking sites unevenly. Moreover, workers can be selective about what information they include on these sites and what information they omit. These choices generate measurement error when these data sources are being used to understand a firm's skills or occupations. Prior work discusses some of these considerations (Horton and Tambe, 2015) but the following sections calibrate specific strengths and deficiencies in terms of coverage. Measurement error in this data set is discussed later in this appendix.

A.2 BLS-SOC share comparisons

The distribution of Revelio workers across occupations is shown in Figure 7a. Figure 7b 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 is provided 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 Information Technology 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 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.

A.3 NAICS Industry comparisons

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

A.4 Geographic (state) comparison

A final comparison, shown in the bottom panel (Figures 7e and 7f) is state-level comparisons. This comparison evaluates the reported geographic location of workers in the Revelio data set with the distribution of workers across US states. Unsurprisingly, we can see that states with significant industry representation for finance and technology (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 larger role in the state economy.

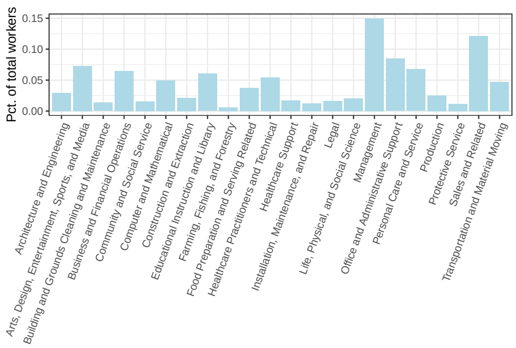
A.5 Discussion

In sum, when we consider the spread of algorithmic technologies into occupations, industries, and geographies, workers in the Revelio data set are likely to be over-representative of those information-intensive industries, occupations, and sectors that are likely to be most impacted by these technological changes.

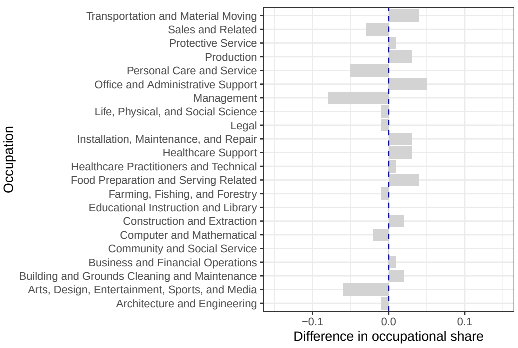
Having greater quantities of workers in this database from some sectors and occupations will affect the precision of the measurement, but this may fall into under normal, random measurement error if those workers who do report their skills are not very different from the ones who do. The number of workers in the database from each Fortune 500 firm is large though, so this type of measurement error should not be very large. Even in underrepresented occupations and industries, the database should produce a high-quality signal of the skill content of a profession.

A less innocuous issue is that the reporting of skills themselves may be inconsistent. Workers in some occupations and industries may be more inclined to report these skills on their profiles. They may consistently report skills that are likely to lead to future employment opportunities, but inconsistently report skills that the market does not deem to be particularly valuable. This can impact the interpretation of the magnitudes of the coefficients in the main regressions (e.g. market return to a marginal database engineer), although it should not impact the sign and direction of these estimates.

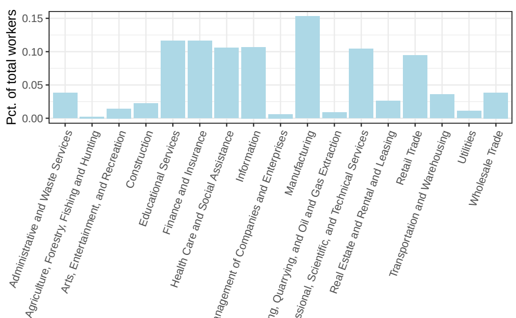
Figure 7: Revelio data distributions



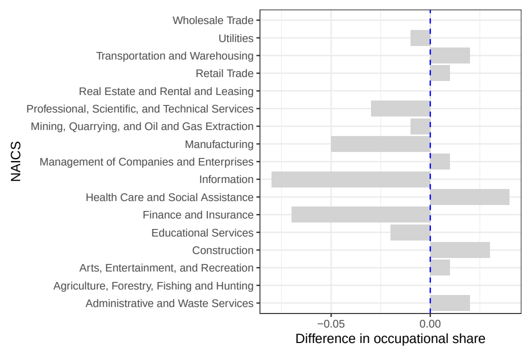
(a) Occupational code (SOC)



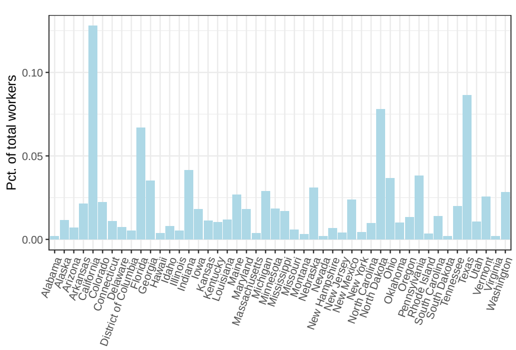
(b) Occupational code (SOC)



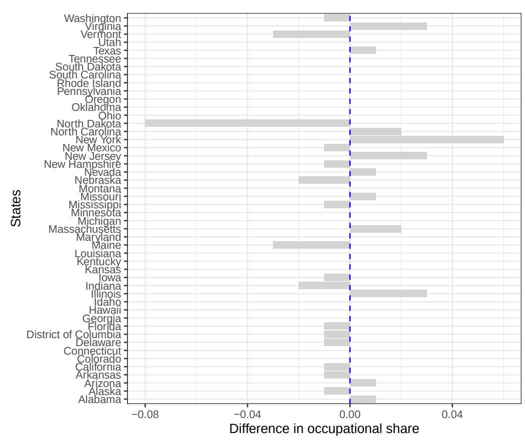
(c) Industry (NAICS)



(d) Industry (NAICS)



(e) States



(f) States

Notes: These three figures illustrate the difference in compositional shares between the Revelio and BLS data sets. The top row is comparison of occupations. The second row is comparison across NAICS industries. The third row is comparisons across states. The length of each bar for plots in the second column is computed as the difference in the share that the worker category accounts for in the Revelio data and in the administrative 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.

B Categorizing skills into technological areas

A key measurement task for this analysis is to generate a taxonomy of skills, either as embedded in job listings or reported by employees on their profiles, that enables measurement of technological expertise. This requires construction of a mapping from granular skills to the broader technological areas to which they are related. For instance, skills such as “Oracle DB” and “MySQL” both indicate expertise with relational database technologies. To construct this taxonomy in the *workforce data*, I leverage an existing structure from Revelio that categorizes skills into technological groups. This data provider uses data clustering techniques to categorize skills into a taxonomy. This approach combines skills into common groups if they inhabit a similar area of the skill landscape after clustering. The ensuing technological clusters are then assigned labels by the provider. The skills that appear in each of the key technology categories, as constructed by the data provider, are shown below. It is important to note that each of the skills shown below are one of 1,500 keyword skill categories that contain sub-skills within them. For instance, ’machine learning’ may include skills within it like ’classification’, ’clustering’, or ’deep learning’. The skill name itself is not an indicator of a hierarchy. For instance, the ’Tableau’ category also contains ’Microsoft BI’ and ’Qlik’, which are competitor no-code tools. Therefore, each of the technological areas shown below is a mapping of categories which themselves are a grouping of keywords.

Skills in the *job listings* data are organized under a separate taxonomy. However, to maintain consistency across the analysis, I harmonize the skills in the job listings data with the technological categories included in the workforce intelligence data. For instance, an ’Algorithms’ category was created from the job listings data by identifying skills in the job listings that had a match with one of the skills in the equivalent category in the workforce data. Matches were made manually, to account for minor differences in case or how skill names were standardized by the different providers.

B.1 Technology categories created from skill categories in the workforce data

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

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

Big Data Technologies. 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

Relational Databases. 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

Cloud & Mobile Technologies. microsoft azure, windows azure, amazon services/aws, cloud-computing, cloud computing, amazon web services (aws), cloud applications, vmware, openstack, vmware esx/vmware infrastructure/v-sphere;; android, objective-c/ios development, mobile device management, wireless technologies, wireless communications systems, mobile application development, swift/xcode, android development/android sdk

Network Administration. 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

General Information Technology. 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

B.2 Technology categories created from skills in the job listings data

ALG. *Algorithms.*
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. *Relational databases & Big data.*
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

NET. *Web & Networks.*
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 Additional results

In this section, we present a fuller discussion for some of table results discussed in an abbreviated manner in the main text. Table B.1 presents results corresponding to the coefficient estimates depicted in Figure 4. The table reports results from the full form of the regression which is:

$$ATTR_i = \beta_A ALG_i + \beta_D DAT_i + \beta_N NET_i + \gamma_i + \epsilon_i$$

Figure 4 in the main text shows that algorithmic skill is correlated with domain expertise and cognitive skill. There are negative correlations with *Character* and *Management*, after conditioning on job title and industry. The regression also includes measures of skills related to databases (DATA) and network administration (NET). Database management is negatively correlated with all dependent variables which is probably a by-product of the skill-intensive nature of that position. Network administration exhibits relatively weak correlations with all of these job attributes.

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	1.548^{***}	0.142^{**}	-0.481^{***}	0.129^{**}	-1.260^{***}
	(0.066)	(0.064)	(0.088)	(0.063)	(0.137)
$DATA$	-0.833^{***}	-0.194^{***}	-0.173^{**}	-0.209^{***}	-0.646^{***}
	(0.070)	(0.055)	(0.072)	(0.056)	(0.084)
NET	-0.423^{***}	0.033	0.079	-0.036	-0.009
	(0.055)	(0.041)	(0.057)	(0.042)	(0.066)
Log(No. of Skills)	1.680^{***}	1.540^{***}	1.885^{***}	1.800^{***}	2.247^{***}
	(0.047)	(0.035)	(0.053)	(0.038)	(0.064)
Constant	-6.461^{***}	-4.779^{***}	-7.233^{***}	-5.890^{***}	-7.965^{***}
	(0.162)	(0.111)	(0.174)	(0.120)	(0.203)
Job Title FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes
Observations	24,888	24,888	24,888	24,888	24,888
Log Likelihood	-8,951.652	-13,366.200	-7,837.757	-12,908.540	-6,061.045
Akaike Inf. Crit.	18,315.300	27,144.390	16,087.510	26,229.070	12,534.090

Table notes: This table reports results from the logit regression $ATTR_i = \beta_A ALG_i + \beta_D DAT_i + \beta_N NET_i + \gamma_i + \epsilon_i$. It estimates conditional correlations between algorithmic skills and other 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.