Algorithmic Bilinguals

Prasanna B. Tambe*
Wharton School, U. Pennsylvania

May 27, 2024

Abstract

This study tests the hypothesis that generating value from data, algorithms, and AI requires the employment of domain experts who are skilled 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. Using two different workforce data sets, I show that 1) employers have been shifting hiring towards requiring greater algorithmic expertise from domain experts, 2) algorithmic expertise 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 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 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, these technologies 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).

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

This paper presents new theory and evidence that a complement to the effective use of algorithms by organizations is the decentralization of "algorithmic expertise" among the firm's domain experts. *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] The definition of "algorithms" used here includes data science and AI tools but excludes tools like databases or web technologies which do not explicitly produce 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 the extent to which it demands both types of human capital to be integrated (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).

The hiring of domain experts with algorithmic expertise contrasts with an employment structure in which technical expertise is primarily centralized in specialized technology (IT) workers, and underscores the argument that as a general purpose technology, AI will increasingly become a central component of many occupations. To explain these changes, this paper develops theory that builds on the literature on the economics of job design and 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). In particular, it focuses on organizational adaptation to technological change that occurs at the sub-occupational level (Spitz-Oener, 2006). 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 may be associated with organizational performance.

These hypotheses are tested using two 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 to study 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 panel of how technology skills have diffused across occupations from 2008 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 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 non-technical occupations in a pattern that more closely resembles general-purpose office software skills (e.g. word processing tools) than technical skills like database administration. By 2016, only one-third of algorithmic expertise was embedded in IT listings. These skills were particularly likely to be embedded

[1]This paper uses the term "expertise". Whether jobs require expertise or "literacy" with the technology is an important distinction with different policy implications but it is beyond the reach of the data sources used in this paper so 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.

in listings that also required domain expertise and the combination of algorithmic and domain expertise is especially important where workers make decisions.

These findings reflect changes in employer preferences, but not whether the workforce could meet these changes in demand. The second analysis utilizes workforce data to demonstrate similar changes in the corporate employment data for an overlapping seven year panel (2015-2021). In public firms, algorithmic skills increasingly spread among non-technical decision-makers in contrast to what we observe for other information technologies. A third analysis, which is intended to be causal in its interpretation, is that software innovations that lower the costs of working with algorithms, such as "no-code" tools, increase the likelihood that domain experts would have algorithmic skills.

The fourth and final analysis turns to the organizational level and demonstrates that financial markets assign higher value to investments in AI and data science when the relevant expertise is distributed among decision-makers who know how to apply these technologies to business goals. This interaction is more important for AI technologies than it is for data science. Robustness tests suggest that the higher values assigned to these assets are unique to these technologies and no similar patterns emerge for other technologies or for employee expertise in other technical skill categories.

This study contributes to two streams of academic literature. 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 prior 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). Yet, 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 and the contemporaneous reduction in the costs of using analytical and AI tools. 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 Data, algorithms, and job design

How algorithmic technologies affect job design depends on three factors: specialization, coordination, and adaptation. Specialization allows for productivity gains, as in the example of medicine where AI algorithms analyze images, freeing radiologists to concentrate on complex cases (Smith, 1776). AI can lower coordination costs by synchronizing 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), such as the case of marketing professionals using 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 shift work away from specialization towards more "holistic" 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 implications for many aspects of IT and job design, but this analysis focuses on the application of algorithms to decisions, where effective synthesis of domain and technical expertise has been a specific 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

[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, sub-occupational shifts have been less widely documented because administrative data agencies do not capture it 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 quickly in occupations that were being computerized. In that sample, within-occupational change accounted for 36% of educational upgrading.

[5]Poll results from 2014 indicate that it is the most common method used for data mining and data science projects, with about half of all 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.

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).

Both the importance and the difficulties of embedding domain expertise in a data-driven decision process are 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 confronts the challenge of how domain expertise is most effectively injected into a 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 is an input into an employee's decision-making process rather than a decision itself (e.g., websites, cloud storage, databases). An example is the creation of the "data scientist" job title itself, which 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).

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 who operate 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 costs to 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 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:

[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]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.

H1: *Employers are more likely to demand algorithmic skills from their domain experts than they are to demand other technical skills.*

This hypothesis does not predict which occupations receive this bundle of domain and technical skills. However, adaptation provides a context in which to theorize about the control of task bundles in work environments. 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. An instructive example is "typing pools" which existed solely to provide typing services within the organization. Over time, the typing task became part of the knowledge worker's job because local adaptation is important for documentation.[9] Conversely, AI may substitute for some forms of domain expertise in areas such as foreign language translation which could move the bundle away from domain experts. Determining whether domain experts or technical experts receive the bundle, therefore, is an empirical question.

H2: *Among domain experts, algorithmic skills are more likely to found in occupations where decision-making plays an important role.*

The answer to the question of who receives the bundle is not static, however, because it depends 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 workers who have acquired domain and technical expertise and employers may forego any productivity gains associated with bundling these skills together. On the other hand, the barriers to use for many tools is falling as vendors compete to speed adoption of their products in the workplace. Examples include the embedding of complex machine learning 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 no coding background.

H3: *A fall in the cost of acquiring algorithmic skills will accelerate their diffusion to the firm's domain experts.*

Job reconfiguration of the type proposed in the prior hypotheses is costly, but firms have strong incentives to pursue these 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), 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).

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

For algorithmic technologies, investors may anticipate greater returns from firms that employ personnel who understand how these tools can be applied to business goals. These adjustments can be costly if employers have to participate in more competitive labor markets to hire workers with these skills, but higher market values would reflect 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 algorithmic investments when the complementary algorithmic expertise is dispersed among domain experts.*

The next sections describe the empirical tests and databases used to test these four hypotheses.

3 Empirical tests

This section describes three tests that evaluate whether algorithmic and domain expertise are unusually valuable when bundled together: (i) whether employers are searching for algorithmic expertise among domain experts (*H1* & *H2*), (ii) whether a fall in cost of acquiring algorithmic skills accelerates their diffusion among domain experts (*H3*), and (iii) whether financial markets are rewarding employers who combine algorithmic investments with these personnel changes (*H4*).

3.1 Correlation tests of employer preferences

Job vacancy data enables tests of whether employers increasingly search for domain expertise and algorithmic skills when hiring non-technical occupations. The first regression model estimates an equation of the form:

$$DOM_i = \alpha_A ALG_i + \gamma_i + \epsilon_i \tag{1}$$

The dependent variable indicates if a job vacancy requires domain expertise and i indexes the job listing. The model estimates correlations with algorithmic expertise, holding other factors constant. The regression includes other technology variables as well as a vector of control variables (γ) that includes job title[10], four-digit industry, and the logged total skills in the listing. For comparison, estimates are presented from other models that substitute other job attributes for domain expertise. Namely, these models substitute: i) social skills, ii) cognitive attributes, iii) character, or iv) management skills on the left-hand side of Equation 1.

To provide insight into why a combination of domain and technical expertise may be particularly valuable, a second test evaluates what tasks co-occur with this combination.

[10] The data provider standardizes job title at a granular level. Exemplar job titles are "Inventory Clerk" and "UX Developer".

$$TASK_i = \beta_D DOM_i + \beta_A ALG_i + \beta_{DA}(DOM_i \times ALG_i) + \gamma_i + \epsilon_i \tag{2}$$

$TASK$ corresponds to any of a variety of tasks in the data-driven decision-making process. DOM and ALG are binary indicators of whether the vacancy requires domain and algorithmic expertise and i indexes the listing. A vector of control variables (γ) includes job title[11], four-digit industry, and the logged total skills in the listing.

3.2 Correlation tests of changes to workforce skills

To test how these employer preferences have translated to workforce changes (*H2* & *H3*), firm-level employment data are used to evaluate how algorithmic expertise is distributed at (i) the occupation level and (ii) at the firm level. At the occupation level, the regression model is:

$$ALG_{ijt} = \beta_{DMK} DMK_{ijt} + \gamma_{ijt} + \epsilon_{ijt} \tag{3}$$

In this regression, i is the firm, j is SOC 6-digit occupation, t is the year, and DMK indicates the importance of decision-making required for that occupation. This specification allows for variation in occupational characteristics, accounting for differences across firms, including employment and time trends. A variation of Equation 3 interacts DMK with an indicator of whether a technical skill is based in no-code technologies ($NOCODE$) to evaluate whether the spread of no-code tools accelerates the spread of these technologies into non-technical occupations.

Shifting focus to the firm level, I test if firms tend to co-invest in domain experts with algorithmic expertise (ϕ^{ALG}) along with algorithmic technologies.

$$\phi^{ALG}_{it} = ALG\ IT_{it} + \gamma_{it} + \epsilon_{it} \tag{4}$$

For this analysis, i and t are the firm and year respectively, and γ includes firm-level controls for size, assets, employment, and industry.

3.3 ϕ^{ALG} and firms' financial value

If these workforce adjustments generate productive assets for firms, financial investors should assign value to them (*H4*). Prior work uses market value regressions to uncover the presence of valuable but otherwise invisible (intangible) assets that contribute to the firm's productive capacity (Brynjolfsson et al., 2002). This approach may be particularly useful for analyzing AI and data science returns because firms need time to adjust to new technologies 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 regression used to test if greater dispersion of algorithmic expertise (ϕ^{ALG}) forms a complement to a firm's use of

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

algorithms is:

$$\begin{aligned} Log(MV)_{it} = & Log(AT)_{it} + Log(PPE)_{it} + Log(IT)_{it} + \\ & Log(ALG\ IT)_{it} + \phi_{it}^{ALG} + (Log(ALG\ IT)_{it} \times \phi_{it}^{ALG}) + \gamma_{it} + \epsilon_{it} \end{aligned} \tag{5}$$

In Equation 5, 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, which account for much of the firm's value. IT measures the firm's aggregate IT investment, $ALG\ IT$ is a proxy measure of the firm's investment in algorithms, and ϕ^{ALG} is the standardized measure of the decentralization of algorithmic expertise among occupations requiring domain expertise. The control variables (γ_{it}) include year, employment, and depending on the specification, industry at the four-digit NAICS level or employer fixed-effects.

4 Data sources and key measure construction

4.1 Key data sources

The empirical tests described above are informed using data sources on a) how employers are adapting jobs to algorithms and b) how the skill composition of the workforce is changing in response. These are supplemented with financial data from public firms to assess how technological and workforce changes are connected to the value that investors assign to firms.

4.1.1 Job listings database

When employers have job vacancies, they post details on their web sites or on job boards. These listings identify employer and job title, the geography of the position, the skills and education sought from candidates, wages offered, and other fields relevant to the search process. In this analysis, listings are used to measure when skills first appear in job ads and how skills co-occur in listings with other skills.

The listings are provided by 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 job information.[12] This provider collects and standardizes data from over 17,000 job boards and corporate web sites, and these data are processed to ensure a listing is not counted multiple times if it is posted in several places on the web. The processed data include posting date, metropolitan area, employer, job title, educational requirements, certifications required, and skill expectations for each vacancy. 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

[12]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.

North American Industry Classification Systems (NAICS) industry. Job openings list skills, such as *Python*, *Random Forest*, *Chemistry*, *Supply Chain*, *Accounting*, *Data Science*, *Teamwork*, or *Communication* which are standardized using a skill dictionary maintained by Lightcast. These skill data 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 industry and occupational coverage. However, prior academic work has provided information on the sampling properties of these data. See, for example, Appendix A of Deming and Kahn (2018), who conduct a detailed comparison of the Lightcast data with administrative data sources. 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 under-sample blue-collar occupations.

4.1.2 Corporate employment database

The corporate employment data were provided by Revelio Labs, a workforce intelligence company.[13] Their databases are constructed from a variety of online sources,[14] and are similar in their informational content to data posted on online professional networks such as LinkedIn. They cover a large fraction of white-collar work in the US, including both public and private US firms, but the sample used in this analysis is limited to public firms so observations can be connected with financial market data. This data source has been less widely used in the literature than the job listings, so in Appendix A, I present comparisons 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.

This database is used to generate measures of annual firm-occupation-skill employment activity from 2008 through 2021.[15] This panel provides information on how specific technical skills, like "machine learning", diffuse across occupations and employers. The records for each employer include CUSIP identifier codes which allow them to be merged with external financial databases such as the Compustat-Capital IQ data.

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

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

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

4.1.3 Supplementary data sources

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

4.2 Construction of sample and key measures

4.2.1 Sample construction

For the job listings analysis, the sample includes all listings in the data set from the months ranging from January 2014 to June 2016 for a total of 30 months of job listings data. 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 described below, the regression-based analyses on these data restrict the 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 of 7,198 firm-years. Table 1 reports summary statistics for a single year from the mid-point of this panel (2018). 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.

4.2.2 Employees' algorithmic expertise (ALG and ϕ^{ALG})

Workforce skill data are used to measure whether employees have or require algorithmic expertise. A key challenge when using workforce skills for empirical analysis is the mapping of granular skills to meaningful measures of expertise.[19] Recent published papers that use large quantities of archival,

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

[17] One notable example of its use for examining technical change is 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 earlier version to match the O*NET codes in my version of the Lightcast data, which were based on the taxonomy before the revision took place.

[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.

digitally collected workforce data have used manual mappings. 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 but 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 groups used in this analysis. For algorithms, examples of these base-level categories include *machine learning*, *business analytics*, *julia*, and *natural language processing* and each of these 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 technology categories are combined 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, algorithmic expertise measures are constructed as the fraction of workers in major occupation groups that are i) in the top quartile of all occupations in terms of their domain expertise requirements and ii) in the importance of decision-making for that job ($\%ALG$).[20] How domain expertise and decision-making importance are measured is explained in the next section. For organization i in year t, an organizational measure of algorithmic expertise (ϕ^{ALG}) is computed as:

$$\phi_{it}^{ALG} = \frac{(\%ALG_{it} - MEAN(\%ALG))}{STD(\%ALG)} \tag{6}$$

Firms in which domain experts have more algorithmic expertise have higher ϕ^{ALG} values. For some robustness tests, parallel measures are constructed for other technological categories (e.g., ϕ^{CLOUD}).

4.2.3 Domain expertise (DOM), decision-making (DMK), and other job attributes

Job-level measurement of domain expertise is constructed to be consistent with the measurement of algorithmic expertise as described above. A binary indicator (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 skills is extracted from the O*NET dictionaries, which identify all of the possible knowledge domains with which US-based jobs may require familiarity.[21] These domains are

For example, see recent efforts by Nesta in the UK or Lightcast.

[20]These O*NET major occupational groups are: 11-0000 (Management), 13-0000 (Business and Financial), 17-0000 (Architecture and Engineering), 19-0000 (Life, Physical, and Social Science), 23-0000 (Legal), and 29-0000 (Healthcare Practitioners and Technical).

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

extracted from the "Knowledge" table in O*NET, which describes "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] Similarly, the importance of *Decision-making* for an occupation (DMK) is retrieved from the O*NET database which provides this measure on a scale of 1 through 7 for each six-digit occupation.

Finally, 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]

4.2.4 Employers' technology 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 often leverage alternative sources to create 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-source, leaving no documented 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. Conversely, most frontier software requires technical expertise to install and maintain, so quantities of complementary, technical human capital may be the most accurate available accounting of a firms' technology investments.

This approach is used to generate measures of firms' technology assets. It follows prior work

[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 important to contrast this approach with one in which jobs would be identified as requiring domain expertise based solely on titles. Such an approach would impose the restriction that jobs with the same title cannot differ in their knowledge content. 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 this analysis.

[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.

where proxy investment measures are constructed as quantities or intensities of skills relevant to the technological domain (Tambe and Hitt, 2012).[25] This view of the dichotomy between technical workers and non-technical occupations is similar to work that treats 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 (i.e., $ALG\ IT$, IT, $DS\ IT$, $AI\ IT$) 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 area can be interpreted as the intensity of investment in that domain.[26]

4.2.5 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 earlier, 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 the Appendix of Brynjolfsson et al. (2002) which describes variable construction). 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.

5 Results

5.1 Model-Free Evidence

5.1.1 The growth of algorithmic expertise in non-technical job listings

Figure 1b illustrates the growth rate of algorithmic skills in non-technical job listings from 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 i is the listing, t is a vector of dummy variables for months since January 2013 when a vacancy was posted, and ALG indicates whether an algorithmic skill appears in a job listing. The likelihood of an algorithmic skill appearing in a non-technical occupation listing rises throughout the sample.

Figure 2a shows the extent to which specific technology skills, including but not limited to algorithmic skills, are bundled with domain expertise in one month of the Lightcast data (January

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

[26] 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.

2016). Algorithmic skills, colored blue, are more commonly bundled with domain expertise (reaching further to the right) and along this dimension have more in common with skills like Excel and ERP systems that are commonly used by non-technical occupations. Figure 2b indicates that algorithmic skills, again colored in dark blue, are more commonly found in job listings for non-technical occupations than are other technical skills. Predictive analytics, data science, and data analysis are only slightly less dispersed than skills related to the Microsoft Office Suite, which supports the claim that employers are increasingly bundling algorithmic skills with domain expertise.

5.1.2 Algorithmic expertise in the workforce data

Job listings indicate what employers want. However, they cannot say whether these listings indicate hard requirements or instead, are an employer "wish list", or whether the vacancies requiring these skills are even ultimately filled. Therefore, I next turn to corporate employment data to investigate whether the changes indicated by these listings are also reflected in workforce changes. These analyses are shown in the four quadrants of Figure 3.

Figure 3a shows that technical skills have become more dispersed in non-technical occupations in these firms. The y-axis is the intensity with which a skill appears in non-technical occupations, with levels depicted relative to their 2008 values. There is steady growth in the rate at which AI and data science skills have penetrated non-technical occupations, consistent with the evidence from job vacancies shown in Figure 1b. By 2021, these skills appeared in 10% more non-technical occupations than in 2010. In contrast, skills corresponding to two other categories, network and cloud technologies, became increasingly specialized, and the incidence of mobile skills remained flat.

Figure 3b shows how the measure ϕ^{ALG}, defined in Equation 6, varies across industry and time. It is highest in the Information, Professional Services, and Finance industries, which is consistent with what has been reported in the press (Lohr, 2024). Retail has climbed rapidly which may reflect the growing use of consumer data for prediction. Levels are lower in Healthcare although they have been climbing, reflecting the growing use of AI and data science in healthcare.

Figure 3c depicts annual changes in ϕ^{ALG} after separating firms into quartiles according to their market values in the final year of the sample (2021). ϕ^{ALG} is greatest in higher quartiles and the differences are largest in the earlier years of the sample, consistent with a labor market where workers with these skills are a scarce resource that higher value firms can more easily attract. In the last years of the sample, however, ϕ^{ALG} converges across quartiles, suggesting that supply-side adjustments have made it easier for employers with fewer resources to attract these workers. The fourth quadrant (Figure 3d) uses data from the final year of the sample and plots firms' investments in algorithms against ϕ^{ALG} where the bubble size reflects the firm's market value. The largest circles (colored in blue) are those commonly referred to as "big-tech" firms. We can see that firms contemporaneously invest in algorithms and in business-facing workers with algorithmic skills.

5.2 Correlational evidence from job vacancy data

From descriptive evidence, we turn towards empirical tests of our hypotheses, beginning with Equation 1. Figure 4a depicts estimates of α_A. The sample is a single month (Jan 2016), and the unit of observation i is the job listing. This test indicates that domain expertise is a predictor of algorithmic skill. The regression includes job-title fixed effects so algorithmic skills are likely to be bundled in a listing for a given title when domain expertise is included among the requirements (**Hypothesis 1**). The full set of estimates from Equation 1, which can be found in Appendix C, indicate negative correlations with data management, which is consistent with database skills being more centralized within IT occupations. We also observe positive correlations between the use of algorithms and social and cognitive skills and negative correlations with management-related job attributes. This negative relationship suggests that employers are not bundling people leadership (character and management) with algorithmic skills.

Figure 4b illustrates β_{DA} from Equation 2. The estimates on *Presentation* and *Decision-making* indicate that algorithmic and domain expertise most often appear together for positions requiring decision-making. In contrast, this combination of skills is negatively correlated with tasks related to *Data management* and *Data modeling*. These data tasks also require greater numbers of skills so they may not favor generalists who bring a diversity of skills to the task.

5.3 Correlational evidence from workforce data

Table 3 reports the results of empirical tests of Equations 3 and 4. In a reflection of Figure 4b from the job vacancy data, Column (1) indicates that occupations where decision-making is important are more likely to have algorithmic skills (t=3.50) (**Hypothesis 2**). This remains true after including firm fixed-effects (Column (2)) which eliminates the role of static firm differences in driving these relationships, such as heterogeneous demand for decision-making.

This evidence is circumstantial but columns (3) and (4) test the causal proposition that a fall in the cost of using algorithmic tools speeds the diffusion of these skills into non-technical occupations. Columns (3) and (4) separate algorithmic skills according to whether the skill corresponds to a "no-code" tool, where technologies are identified as falling into this category using GPT-4 based labeling.[27] No-code tools are a small fraction of all represented technical skills so the main-effect is negative (t=-28.88). Column (4), however, indicates that tools with lower costs of use see greater uptake in non-technical decision-making occupations (t=4.29) (**Hypothesis 3**). This finding supports a causal relationship subject to the assumption that these tools perform the same functions as their code-based counterparts. If these two types of tools perform substantially different functions, heterogeneity in occupational usage may instead reflect differences in demand.

Columns (5) and (6) aggregate the data by occupation and test if algorithmic expertise is more decentralized (ϕ^{ALG}) in firms that invest in algorithms which would be predicted if these two investments are complementary. The estimates confirm that investments in algorithms and in

[27] The use of Large Language Models for data annotation has been shown to be effective in domain-specific tasks and is becoming increasingly common in the literature (Møller et al., 2023).

ϕ^{ALG} are correlated, after controlling for characteristics like size and industry that may drive these investment patterns. These correlations are robust to including firm fixed-effects (t=2.34).

5.4 Regression evidence from market values

Table 4 reports estimates from the market value regression in Equation 5 on a seven year panel of public firm investments (2015-2021). Columns (1) through (3) include year and 4-digit industry fixed-effects. Column (1) reports results of market value on measures of IT investment, assets, capital (PPE), and employment where all of these variables are entered in logs. As prior work has also found (Brynjolfsson et al., 2002), financial markets assign economic value to investments in general IT capital (t=3.79). After adding algorithmic investment into the regression, the coefficient on general IT capital falls to zero suggesting that the market returns to IT investment are principally from frontier investments, as represented in this panel by AI and data science investments (t=3.93).

Column (3) reports estimates from the full specification shown in Equation 5 that includes the interaction term between investment in algorithms and ϕ^{ALG}. The main effect on algorithms is similar to column (2) (t=3.80) but the interpretation of the interaction is that these investments are valued an additional 15% higher in firms where ϕ^{ALG} is one standard deviation above the mean (t=2.67). This supports the hypothesis that contemporaneous investment in ALG and ϕ^{ALG} builds valuable intangible assets that will be useful for producing a future stream of AI goods and services (**Hypothesis 4**). Column (4) adds firm effects instead of industry effects. Including firm effects drives the coefficients on algorithms, ϕ^{ALG}, and the interaction term to zero. This may simply reflect limited variation in this relatively short panel, but another interpretation is that there are high adjustment costs for firms building these assets. Within the limited range of this panel, unobserved differences across firms may explain most of the heterogeneity in firms' endowments of these assets. A way to probe this argument is to separate extensive and intensive margins of investment for algorithms. Column (5) substitutes a measure indicating if 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). By eliminating most within-firm variation, they show that across-firm differences are more important for these estimates. Firms have a limited ability to change their asset mix within the years covered in this sample.

These three sets of analyses from the job vacancies, employment data, and financial markets, suggest that in the last decade, (i) employers have been adjusting search to find domain experts with expertise in algorithms, (ii) algorithmic skills have spread to domain experts in decision-making positions, and (iii) employers that match their algorithmic investments with these workforce changes realized higher market values, suggesting the presence of valuable intangible assets in these firms. This evidence collectively 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.

5.5 Effect heterogeneity and robustness tests

The average effects computed in Table 4 potentially obscure important heterogeneity. Figure 5 reports estimates on the main interaction term for Equation 5 where the sample is split by employment size (separated by tercile). The financial returns to these investments are greater for larger firms, which may reflect the higher coordination costs faced by workers in large firms or alternatively, the advantages larger firms enjoy when hiring such workers or installing these technologies. The first explanation could be expected to persist over time, but the second could fade as the supply side of the market adjusts as suggested by Figure 3c.

The analysis until this point has grouped AI and data science investments into a single category even though the frontier of IT investment has rapidly progressed from data science to AI during the years covered by the panel. Table 5 separates these technologies and divides the panel into the years before and after 2018. The first two columns suggest a relationship between data science investment and market value during the earlier 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 greater value being assigned to AI investment in the later period. Columns (3) and (4) introduce interaction terms with organizational skills for these technologies (ϕ^{DS} and ϕ^{AI}). In the early period, neither coefficient exhibits 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 suggests that the bundling of domain and algorithmic expertise matters more for AI technologies.

Figure 6 summarizes the results of robustness tests that suggest the correlations reported above are likely due to the hypothesized theoretical relationships rather than omitted variable bias. These tests focus on the interaction between AI and ϕ^{AI} in the last two years of the sample, which is the window in which these employment patterns appear to be most salient. Figure 6a retains the ϕ measure but interacts it with different technology investments (investments in data and networks, respectively). The theoretical discussion suggests we should observe the strongest correlations with algorithmic investment, which is the pattern we observe. Neither of the interaction terms created using the other technologies exhibits meaningful correlations, which indicates that correlations between market value and the interaction of ϕ^{AI} and AI investment are not simply reflecting sources of unobserved heterogeneity that would be correlated with overall technology investment, such as general financial resources, free cash flow, or overall digital intensity. Figure 6b performs a similar comparison but it uses AI investment for all interactions and alters the decentralization measure. It includes ϕ^{AI} but it also adds measures constructed in similar ways for ϕ^{NET} and ϕ^{CLOUD}. We only observe market value correlations when AI investment is accompanied by ϕ^{AI}.

In a final placebo test, Figure 6c returns to the original construction of ϕ^{AI} but constructs an alternate measure using occupations where decision-making is *lowest* in its importance. This comparison reveals that market value correlations are stronger for occupations where decision-making is important. In sum, Figure 6 indicates that correlations with market value only appear at the confluence of AI investment and AI skills among domain experts. We should be cautious before interpreting this relationship causally, but these results do suggest that investments in algorithms and in the workers who can apply them to meet business goals are disproportionately accumulating in high value firms. Similar patterns of resource accumulation do not exist for other technologies or skills or for employees who are not instrumental to the firm's decision-making processes.

6 Managerial Implications

Adoption of algorithmic decision-making, and particularly predictive AI applications, has been difficult and uneven (Zolas et al., 2021). The evidence from this analysis suggests that the human capital of data-driven firms differs from that of firms that lag in this domain. This implies firms face considerable adjustment costs when adopting these technologies, which in turn suggests competitive rents for firms that have successfully found the right mix of workers.

A caveat is that the costs of using AI and data science technologies are rapidly 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. A reduction in the costs of using these tools means that employers can more easily push data analysis tasks to domain experts. The implications of this shift for managers can be significant. For managers, 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 managerial challenge is that technical skill has 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 technical expertise is required from a growing number of occupations, 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 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 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 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). These findings suggest these changes may be an appropriate response to a labor market that will increasingly demand algorithmic bilinguals.

7 Conclusions

This paper provides evidence from two different data sources that i) algorithmic expertise is becoming broadly dispersed across domain experts in organizations at the frontier of algorithmic investment, ii) that this pattern is due to complementarities that arise between algorithmic and domain expertise, and iii) that the market assigns higher value to firms that concurrently make these workforce adjustments while investing in algorithmic tools. It documents one early but important facet of the workforce transformation occurring to support the use of algorithms in the organization.

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

Beyond these limitations, there is significant scope for future work in this area. Research about the coming wave of investment in technologies that convert data into decisions, and the workforce transformation that will be required to accompany these changes, is in its infancy. We have much to learn about how to design organizations so that humans can effectively work with algorithms. This paper considers one facet of workforce transformation but 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.

Indeed, a key limitation of this paper, like most research on technology and work, is that it takes a static view. At this early stage, there is limited evidence that the use of these technologies has had 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 frontier advances, it will continue to change markets for these skills, and continue to raise new questions about how employers should integrate algorithms into their processes.

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.

Møller, A. G., Aarup Dalsgaard, J., Pera, A. and Aiello, L. M. (2023), 'Is a prompt and a few samples all you need? using gpt-4 for data augmentation in low-resource classification tasks', *arXiv e-prints* pp. arXiv–2304.

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	1,250
Log(Assets)	Millions (USD)	8.503	2.18	1,250
Log(PPE)	Millions (USD)	5.754	2.57	1,250
Log(Employment)	Thousands (Employees)	2.218	1.44	1,250
Log(IT)	Skill count	7.065	1.92	1,250
Log(Network IT)	Skill count	3.370	2.04	1,250
Log(Database IT)	Skill count	5.062	1.86	1,250
Log(Alg IT)	Skill count	2.949	2.21	1,250
Log(Data science)	Skill count	3.342	1.69	1,250
Log(AI)	Skill count	2.032	1.78	1,250
ϕ^{ALG}	Standardized Value	0.061	0.88	1,250

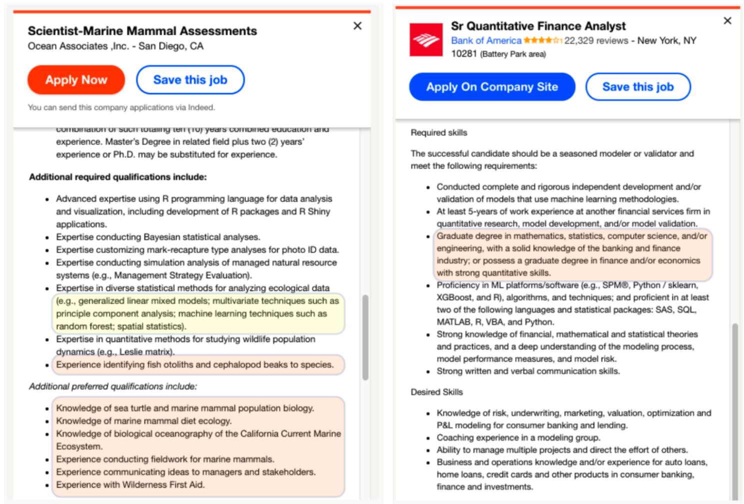
Table notes: This table reports summary statistics for firms in the 2018 cross-section of the regression panel constructed from the workforce data. The year 2018 was chosen as the midpoint in the panel window (2015-2021). The data source for the first four rows [*Market Value, Assets, PPE, Employment*]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

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

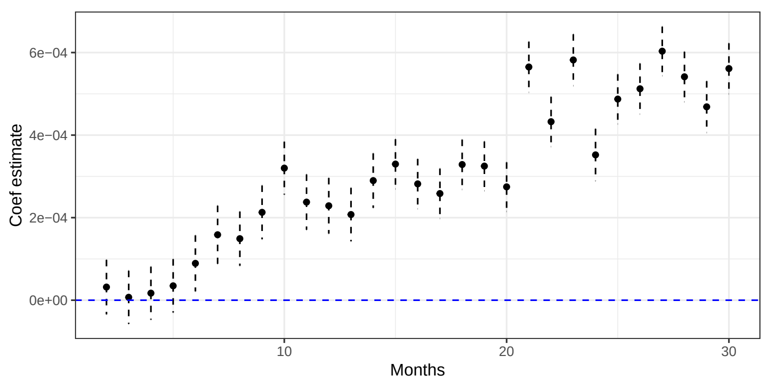
NAICS 2	Sector	N
11	Agriculture, Forestry, Fishing and Hunting	1
21	Mining, Quarrying, and Oil and Gas Extraction	29
22	Utilities	34
23	Construction	10
31-33	Manufacturing	360
42	Wholesale Trade	34
44-45	Retail Trade	42
48-49	Transportation and Warehousing	34
51	Information	268
52	Finance and Insurance	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. It uses the same cross-section of firms as in Table 1.

Figure 1: The growth of algorithmic skills in job listings



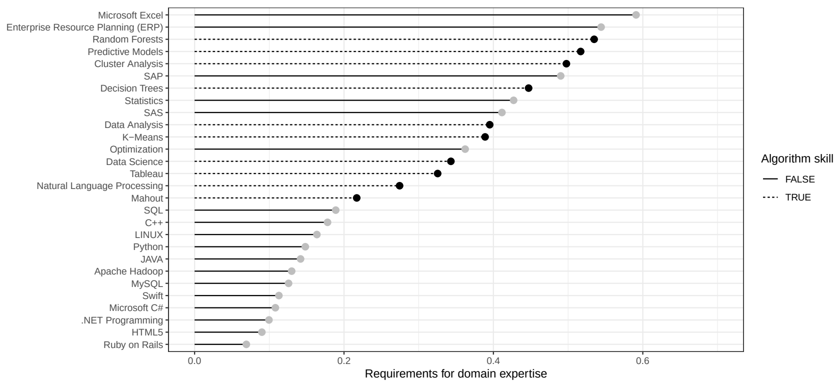
(a) Sample listings with algorithmic and domain expertise



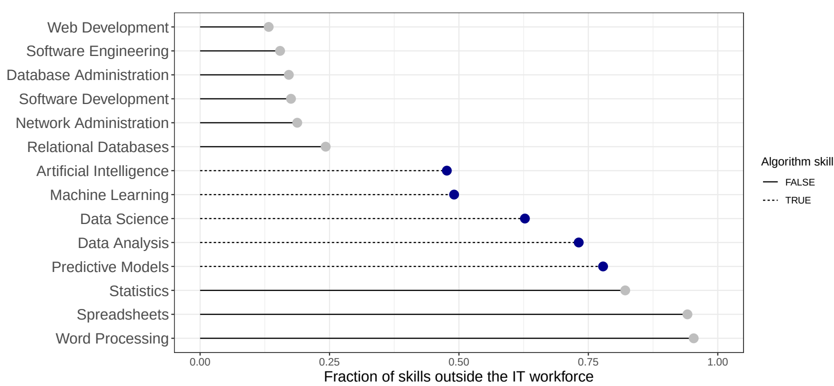
(b) Likelihood algorithmic expertise appears in a non-technical listing

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 $ALG_i = \boldsymbol{\beta}\mathbf{MONTH}_i + \epsilon_i$ for the listings in the months covered by the Lightcast data a dummy variable is included for each month in the data set and 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



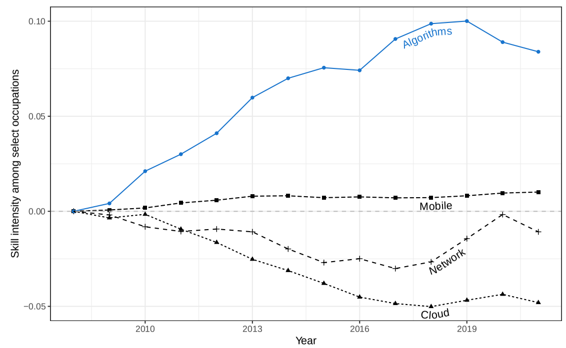
(a) Bundling of technical skills with domain expertise



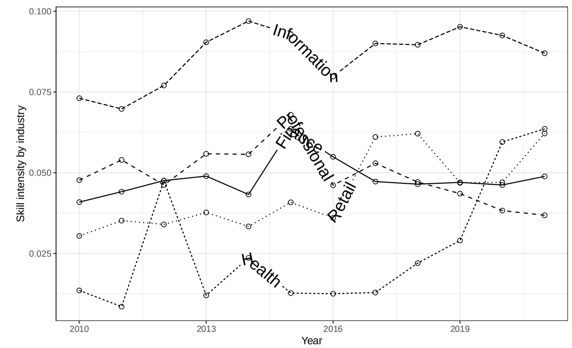
(b) Fraction of skills appearing in non-technical job listings

Figure notes: Figure (a) indicates the extent to which different information technologies are bundled with domain expertise for skills appearing in a single month (January 2016) of the job listing data (N=763,986). Skills in dark blue (dashed line stems) are in the algorithms category 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 of occurrences where a technology appears in non-IT occupation listings. The sample is restricted to listings in one month (January 2016) of the sample data with skills in one of the areas indicated (N=263,256). Skills in dark blue (dashed lines) are those in the algorithms category. A value closer to one means that a skill is more likely to appear in non-IT occupations.

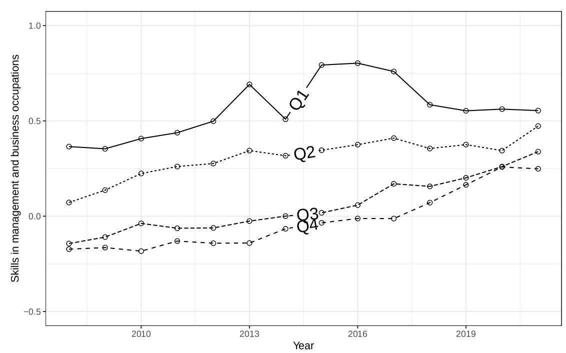
Figure 3: Changes in the locus of technical expertise in organizations from 2008-2021



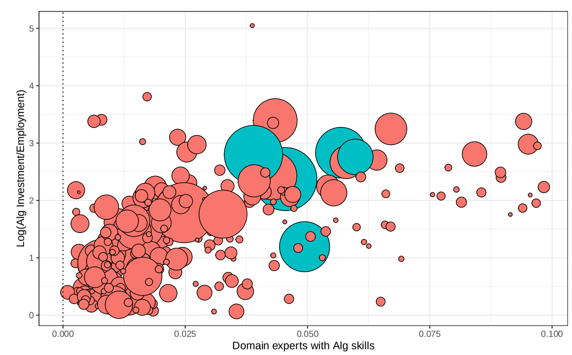
(a) Skill intensity in decision-making occupations



(b) ϕ^{ALG} by industry



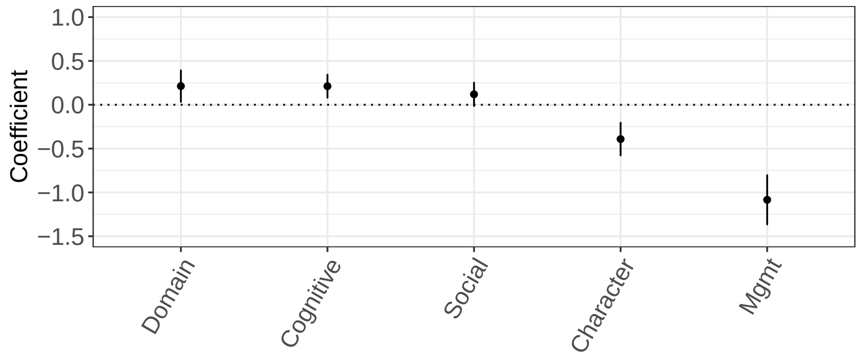
(c) Convergence in ϕ^{ALG} across firm size quartiles



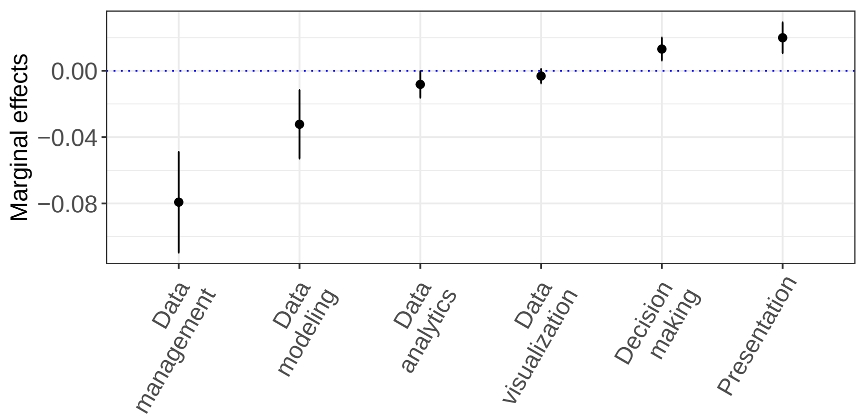
(d) Organizational accumulation of skills

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

Figure 4: Skills in non-technical job listings requiring algorithmic expertise



(a) Correlations between algorithmic expertise and key job skills



(b) Marginal effects of algorithmic and domain expertise (joint) on different data tasks

Figure notes: Figure (a) depicts correlations between skills needed on-the-job and algorithmic expertise in the job listings from January 2016. Each vertical bar is a coefficient estimate from a separate regression of the form $SKILL_i = \alpha_{ALG}ALG_i + \alpha_{DATA}DATA_i + \alpha_{NET}NET_i + Log(No.\ Skills)_i + \gamma_i + \phi_i + \epsilon_i$ 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. The estimates from the full form of each of these regressions is shown in Appendix C. Figure (b) reports results from tests of which data tasks require a combination of both algorithmic and domain expertise using the January 2016 job listings. The logistic regression is $DATATASK_i = \beta_{DA}(DOM_i \times ALG_i) + \beta_D DOM_i + \beta_A ALG_i + Log(No.\ Skills)_i + \epsilon_i$ where DOM_i and ALG_i are binary variables indicating that a listing requires domain 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 point estimate that is presented is the marginal effect of the β_{DA} coefficient. Standard error bars show the 95% confidence interval.

Table 3: Correlations between occupational decision-making, ease of technology use, and algorithmic expertise

	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: The observations in the first four columns are at the firm-occupation-year level where occupations are at the 6-digit Standard Occupational Classification (SOC) occupation level. The model is $ALG_{ijt} = DMK_{ijt} + Log(Occ\ Count)_{ijt} + Log(Employment)_{it} + \gamma_{ijt} + \epsilon_{ijt}$ 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. *DMK* is an indicator of the importance of decision-making for the occupation as recorded in O*NET. *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 employer. For columns (5) and (6), observations are at the firm-year level and the regression model is $\phi^{ALG}_{it} = Log(Assets)_{it} + Log(AlgIT)_{it} + Log(PPE)_{it} + Log(Employment)_{it} + \epsilon_{it}$. *Assets*, *Employment*, and *PPE* are firm level measures from the Capital IQ database. *Alg IT* is a measure of investments in algorithms. ***p<.01, **p<.05, *p<.10.

Table 4: OLS regressions of algorithms and the decentralization of expertise 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 firm panel ranging 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 firm-year level. 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 firm fixed-effects instead of industry controls. Column (5) replaces ϕ^{ALG} with a binary indicator of whether ϕ^{ALG} is above or below the mean value for that variable. Standard errors are clustered on employer. ***p<.01, **p<.05, *p<.10.

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

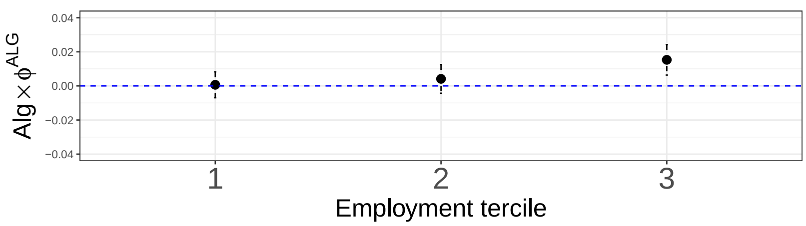


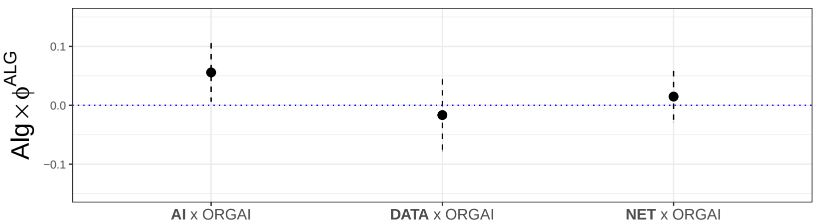
Figure notes: The y-axis indicates the coefficient on the interaction term between algorithm investment and ϕ^{ALG} from the main specification used in column (4) of Table 4 where the sample is divided into terciles by employment size. The sample size in each regression is approximately one-third the sample size used in column 4 of Table 4. On the x-axis, "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 in the market value regressions (2015-2021)

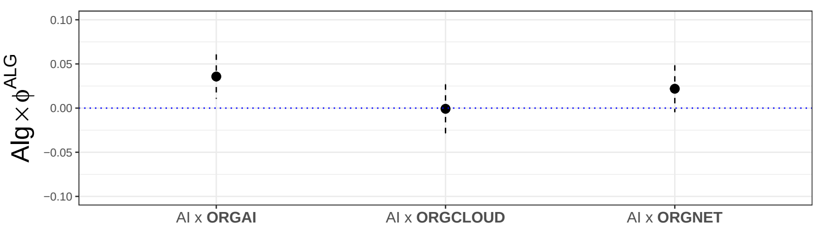
DV	Log(Market Value)			
Years	2015-2017	2018-2021	2015-2017	2018-2021
Model:	(1)	(2)	(3)	(4)
Variables				
Log(Assets)	0.733^{***}	0.696^{***}	0.733^{***}	0.695^{***}
	(0.051)	(0.044)	(0.051)	(0.043)
Log(PPE)	0.086^{**}	0.073^{*}	0.085^{*}	0.064^{*}
	(0.043)	(0.037)	(0.043)	(0.037)
Log(AI)	0.038	0.104^{***}	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 algorithms and expertise measures relate to market value across the earlier and later parts of the panel where algorithms are separately broken into AI and data science investment and skills. 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 on employer. ***p<.01, **p<.05, *p<.10.

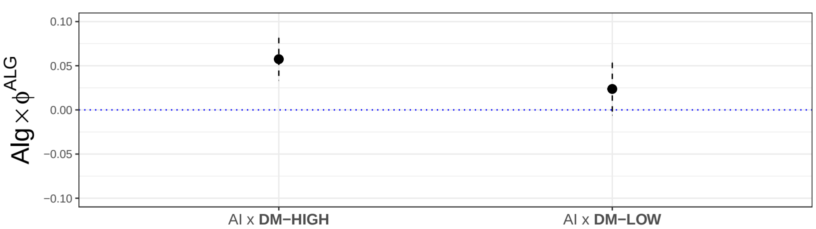
Figure 6: Placebo tests using alternative measure constructions for ϕ^{AI} and AI



(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: This figure illustrates placebo tests for the market value regression for AI investment in the last two years of the sample. In addition to the main interaction measure $AI \times \phi^{AI}$, it also includes interaction terms between ϕ^{AI} and investment in databases (DATA) and networks (NET) (N=1,312). In the top facet, the marker on the left is AI, the middle is databases, and the right is networks. The middle facet takes a similar approach but uses AI investment for all measures and adds interaction terms for ϕ^{CLOUD} and ϕ^{NET} in addition to ϕ^{AI} (N=2,479). The third facet separates the O*NET occupations used to construct ϕ into two separate regressions: (i) those where decision-making is important (left) and (ii) where it is unimportant (right). N=1,312 for both regressions. The standard error bars in all three facets 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 A.1a. Figure A.1b 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 A.1c and A.1d. 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 A.1e and A.1f) 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.

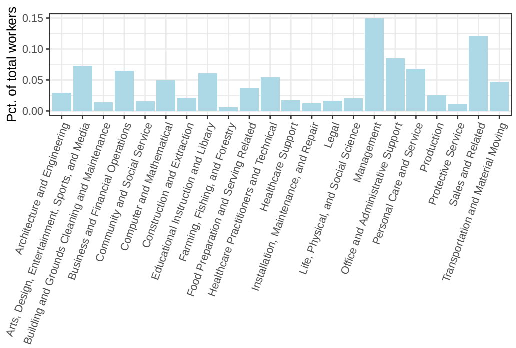
A.5 Discussion

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.

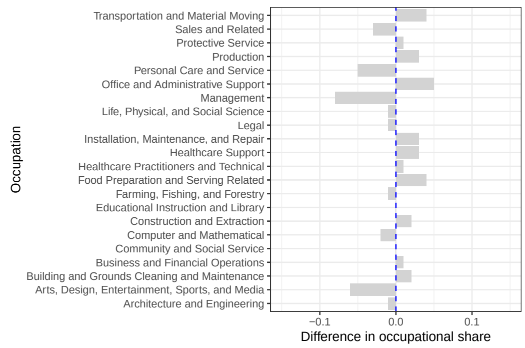
Having greater quantities of workers in this database from some sectors and occupations will affect the precision of the measurement, but this may fall into under normal, random measurement error if those workers who do report their skills are not very different from the ones who do. The number of workers 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.

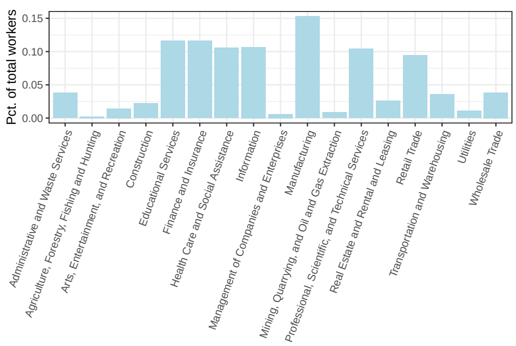
Figure A.1: Revelio data distributions



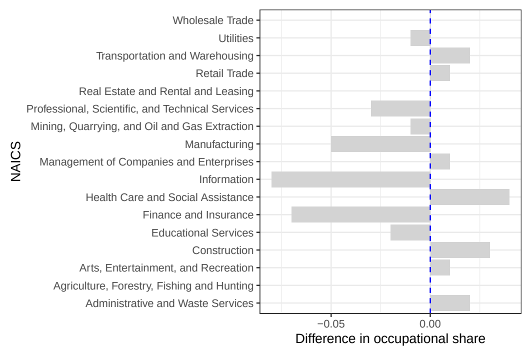
(a) Occupational code (SOC)



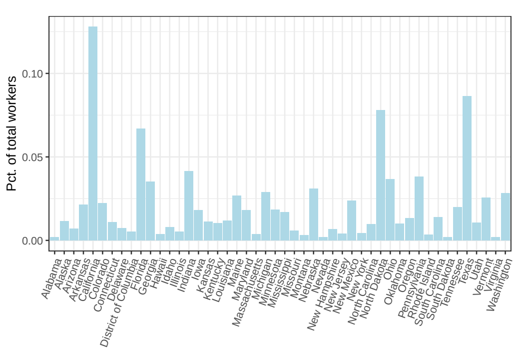
(b) Occupational code (SOC)



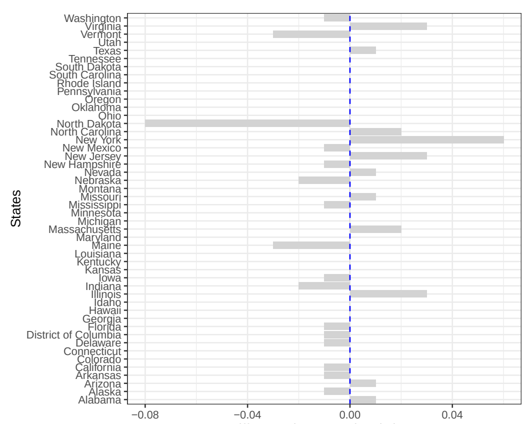
(c) Industry (NAICS)



(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 Full correlation table between technologies and job skills

In this section, we present a fuller discussion for some of table results discussed in an abbreviated manner in the main text. Table C.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 C.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 expertise and other job attributes. $Log(No.of skills)$ 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. ***p<.01, **p<.05, *p<.10.