Algorithmic Bilinguals

Prasanna B. Tambe*
Wharton School, U. Pennsylvania

April 17, 2024

Draft copy. Results are preliminary and subject to change.

Abstract

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

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

1 Introduction

The impact of algorithmic decision-making on organizations is a topic of growing interest (Rock, 2019; Wu et al., 2019; Agrawal et al., 2018; Zolas et al., 2021). Research in this area has focused on the labor reallocation effects of AI and automation technologies (Acemoğlu and Restrepo, 2016; Autor and Salomons, 2018; Brynjolfsson et al., 2018; Eloundou et al., 2023), as well as demonstrating that these technologies are not simply labor displacing (Agrawal et al., 2019; Gregory et al., 2022). Instead, they are also likely to generate new jobs and new types of jobs (Bessen, 2019; Autor et al.,

*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.

2022), and a key theme of the recent literature in this area is how humans can be most effective when working alongside algorithms (Faraj et al., 2018; Cowgill, 2018; Beane, 2019; Agrawal et al., 2019; Lebovitz et al., 2022; Babina et al., 2022).

This paper develops new theory and evidence which argues that a complement to the effective organizational use of algorithms is "algorithmic literacy" being embedded in jobs requiring domain expertise. It focuses on two types of human capital. *Algorithmic literacy* refers to skills related to the use of algorithmic tools, such as data science, machine learning, or other AI tools, which enable firms to convert data into strategic decisions in the pursuit of business goals.[1] *Domain expertise* refers to the expertise needed to work in a specialized field such as nursing, sales, marketing, or accounting. Prior work suggests that both domain expertise and "interactional" technical expertise are important when using algorithms (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).

Requiring algorithmic literacy from domain experts contrasts with an organizational design in which technical expertise is localized in specialized technology (IT) workers, and it relates to the observation that much of the occupational adaptation to technological change is occurring at the sub-occupational level (Spitz-Oener, 2006). It is also implied by the argument that as a general purpose technology, AI will increasingly become a central component of all occupations. To explain these changes, this paper develops new theory that builds on the literature on the economics of job design and considers how algorithms differ from other business information technologies (Smith, 1776; Becker and Murphy, 1992; Dessein and Santos, 2006; Teodoridis, 2017; Lindbeck and Snower, 2000; Postrel, 2002). It generates hypotheses related to i) how firms adapt jobs when using algorithms for decision-making and ii) how these adjustments impact organizational performance.

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

The analysis produces four findings. First, using the job listings data, I show that markers of algorithmic literacy gradually spread across listings from 2013 to 2016 in a pattern that more closely resembles general-purpose office software (e.g. word processing tools) than more specialized technical skills like database administration. By 2016, only one-third of these technologies were embedded in listings for IT workers. In these data, these markers are particularly likely to be embedded in

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

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

listings for occupations requiring domain expertise.

These findings reflect changes in employer preferences, but not whether the market can meet these changes in demand. Second, I show that these job listing trends are consistent with changes in the corporate employment data over an overlapping seven year period (2015-2022). In public firms, markers of algorithmic literacy increasingly spread to domain experts – especially business, management, and financial occupations – which is consistent with the job listing data and in contrast to other business technologies which have not experienced similar changes during the same period. A third finding, which moves beyond descriptive evidence, shows that technological innovations in software and tools that lower the cost of working with algorithms, such as the spread of no-code tools in the marketplace, increases the likelihood that employers make these changes to jobs.

As a fourth finding, I present correlations suggesting that these workforce adjustments are generating productive, intangible assets for public firms. The spread of algorithmic literacy among a firm's domain experts is most visible in high market value firms. Financial markets assign higher value to algorithmic investments when a firm's domain experts have this literacy. Similar correlations are not evident for other types of IT investment or with other forms of employee expertise.

This study contributes to two streams of academic literature. First, with its focus on employers, it contributes to a literature identifying management complements to investment in new information technology (Bresnahan et al., 2002; Black and Lynch, 2001; Caroli and Van Reenen, 2001; Bartel et al., 2007; Bloom et al., 2012). These analyses have principally been rooted in a view of IT as a technology that can automate "routine" tasks, but the application of algorithmic technologies to contexts where decision rules are not easily mapped to software has reopened the discussion on how IT 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 use of algorithms 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 focuses on IT workers (Ang et al., 2002; Levina and Xin, 2007; Mithas and Krishnan, 2008; Wiesche et al., 2019; Tambe et al., 2020), but there has been limited work on the implications of technical skills for broader workforce outcomes (Atasoy et al. (2016); Deming and Noray (2020) are exceptions). The absence of work in this area is notable given the growing demand from students and workers from all backgrounds for "coding" and other technical skills. Contemporaneously, the costs of doing technical work continue to fall, as new advances (e.g. no-code tools for data manipulation or natural language tools like Code Interpreter/ChatGPT from OpenAI) make it cheaper and faster to do data science with little or no investment in new skills. These findings contribute to our understanding of how the human-algorithm connection will shape the demand for skills as employers increasingly embrace these technologies.

2 Theory and Hypothesis Development

The literature on job design argues that tasks are organized into jobs according to three factors: specialization, coordination, and adaptation. How AI affects job roles may depend on these three factors. Specialization allows for productivity gains, as in the medical field where AI algorithms analyze medical images, which frees radiologists to concentrate on complex cases (Smith, 1776). For coordination, AI can enhance the synchronization of interdependent tasks, exemplified by supply chain management where predictive models can efficiently align inventory, supply, and delivery schedules (Becker and Murphy, 1992). A third factor is adaptation, with AI excelling in tailoring tasks to local information (Dessein and Santos, 2006). Marketing professionals, for instance, might leverage AI to create personalized advertising based on consumer behavior data. These adaptive capabilities are particularly valuable where the application of local knowledge is critical.[3]

Specialization has been a central theme of the literature on IT and jobs. New technologies incentize employers to adjust the mix of skills *within* occupations (Spitz-Oener, 2006). Lindbeck and Snower (2000) argue that task complementarities in knowledge-rich jobs have shifted work away from specialization towards "holistic" work in which workers handle a diversity of tasks. Multi-task work raises productivity when there are informational complementarities among tasks because productivity in one task can be interdependent with levels of activity in the others (Postrel, 2002).[4] For instance, 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).

Effective synthesis of domain and technical expertise is not a new challenge. Many organizations that rely heavily on data-mining processes have been faced with the problem of how to inject domain expertise into the data modeling process. For example, a common and standardized process used to balance data modeling decisions with business objectives is "CRISP-DM" (Cross Industry Standard Process for Data Mining) (Wirth and Hipp, 2000; Chapman et al., 2000).[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. In this model, domain expertise is viewed as separate from technical expertise and conceptualized as 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

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

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

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

costs across stakeholders as a weakness of this paradigm (Saltz, 2021).

These challenges could be amplified for modern data science technologies because they can make decisions that are not viewed as routine (Brynjolfsson et al., 2018; Agrawal et al., 2018). For non-routine decisions, there can be additional costs to separating technical and domain expertise. Reflecting this tension, a recent literature confronts the challenge of injecting domain expertise into the data science process (Mao et al., 2019; Choudhury et al., 2020; Park et al., 2021). The iterative nature of data exploration, experimentation, and learning required for data science may favor generalists, who have a diversity of skills, rather than specialists (Colson, 2019). One example of this tension is within "data scientists" themselves, who by definition of the title, combine technical and statistical skills with domain expertise (Davenport and Patil, 2012; Provost and Fawcett, 2013). The importance of domain expertise for data science has been discussed online[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 other algorithmic decision-making contexts (Jha and Topol, 2016).[8] Users of machine learning tools in high-stakes contexts must evaluate the trade offs when choosing which data to include in a model, how to construct model features, or how to assign value to the costs of prediction errors (Kleinberg et al., 2018; Cowgill, 2018; Cowgill et al., 2020). Research situated in pharmaceutical industries has indicated the importance of embedding the relevant human capital in downstream occupations (Wu et al., 2019) and in healthcare, Jha and Topol (2016) and Lebovitz et al. (2022) describe the challenges healthcare workers face when interpreting the accuracy of machine learning tools and output. The first hypothesis tested in this paper is:

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

This hypothesis, based in specialization, makes an assertion about whether these skills are likely to be bundled, but does not consider the question of which occupations within the organization will receive this bundle. Adaptation provides a context in which to theorize about the control of task bundles in work environments. An instructive parallel is typing pools, through which workers once produced typing services. The typing task eventually became part of the knowledge worker's job because local adaptation is important when creating documents.[9] Similarly, if domain expertise helps with local adaptation, organizations may prefer that domain experts – like those in finance

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

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

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

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

and human resources – receive these skills as the costs of acquiring them falls. Conversely, AI introduces the potential for decentralized control. AI can substitute for some forms of domain expertise, reducing its need in areas such as foreign language proficiency, due to AI-powered translation tools. Here, the technology itself can provided the services once provided by domain experts. Therefore, who gets the bundle is an empirical question.

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

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

> ***H3****: Algorithmic skills are more likely to diffuse into domain expert jobs as the cost of acquiring the skills needed to use algorithmic tools falls.*

Prior work has shown that 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 technologies that can perform routine tasks, allocating decision authority to front-line decision makers has been shown to yield higher productivity levels (Bresnahan et al., 2002), with IT investments being particularly valuable in turbulent environments where the value of decisions depends on rapidly changing external conditions (Mendelson and Pillai, 1998; Pavlou and El Sawy, 2006; Tambe et al., 2012; Black and Lynch, 2001; Bresnahan et al., 2002).

As argued above, firms that invest in algorithmic technologies may realize greater value if they colocate algorithmic skills and domain expertise. These adjustments may be costly, in the form of employers having to navigate more competitive labor markets to hire workers with these skills, but higher values 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 technologies when the complementary skills are dispersed among domain experts.*

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

3 Data sources and key measure construction

3.1 Key data sources

To empirically test the relationships described in the previous section, I use data sources which provide information on a) how employers are designing jobs around algorithms and b) how the skill composition of the workforce is changing in response. I supplement these with financial data for public firms in the sample to assess how these technological investments and workforce changes are connected to the value that investors assign to the firm.

3.1.1 Job listings database

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

Job listings data have been used in several papers on changing workforce skill requirements (Todd et al., 1995; Slaughter and Ang, 1996; Gallivan et al., 2002; Lee and Han, 2008). This study uses data from Lightcast, a labor market analytics firm that 1) uses software to crawl a "near-universe" of online job postings and 2) uses natural language processing to parse skills and other job information.[10] This data provider uses proprietary software to collect and standardize data from over 17,000 job boards and corporate web sites, and these data are processed to ensure that a job listing is not counted multiple times if an employer posts it several places on the web. The processed data include posting date, job location (metropolitan area), employer name, job title, educational requirements, certifications required for the position, and skill expectations for each job. Several studies have used this data source to study labor markets (Hershbein and Kahn, 2018; Deming and Kahn, 2018; Modestino et al., 2019), including the question of 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 the employer in a listing is tagged with a North American Industry Classification Systems (NAICS) industry. Job openings list skills, such as *Python*, *Random Forest*, *Chemistry*, *Supply Chain*, *Accounting*, *Data Science*, *Teamwork*, or *Communication* which are standardized using a skill dictionary maintained by Lightcast. These skill data are not "requirements". Employers can omit skills from listings, some skills may be assumed rather than listed, and successful candidates may not need all of the skills in a listing. Nonetheless, employers are likely to be thoughtful about the skills they put in listings because

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

including or omitting a skill can attract or repel the wrong type of applicant.

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

3.1.2 Corporate employment database

The corporate employment data were collected through a partnership with the workforce intelligence company Revelio Labs.[12] Their databases are constructed from a variety of data sources including online career profiles and federal databases.[13] These data are similar in their informational content to that posted on online professional networks such as LinkedIn and they cover a large fraction of white-collar work in the US. The data cover both public and private US firms but the sample used in this study is limited to public firms so that they can be connected with financial market data. This data source has been less widely used in the literature than the job listings data, so in Appendix A, I report comparisons of these data with administratively sampled workforce data from the Bureau of Labor Statistics. We can see from these comparisons that like the Lightcast data, these data are over-sampled in management, business, and technology occupations and under-sampled in areas such as agriculture and manufacturing which is consistent with the greater use of online professional platforms in knowledge-intensive occupations.[14]

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

3.1.3 Supplementary data sources

To create job expertise measures, the O*NET codes in the job listing data are connected with the Occupational Information Network (O*NET) content model published by the Bureau of Labor

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

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

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

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

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

Statistics.[16] The O*NET database has been very widely used in academic research,[17] is government administered, collected by surveying occupational experts, and provides information on employment, wages, and the work content of US jobs. The O*NET taxonomy reports work requirements including the knowledge required for different occupations.[18] Finally, some analyses also use firms' financial data from Compustat-Capital IQ, which was collected through the WRDS data service.

3.2 Construction of key measures

3.2.1 Algorithmic literacy and other technological expertise

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

Like with this prior work, I group skills into categories for technology measurement. I principally rely on categorizations generated by the data providers themselves, who use data-mining and clustering methods to group skills together into different domains like "data science", "AI", or "Big data". Appendix C delineates the specific skills that fall into each of the technological categories used in this analysis.

The main comparisons in this analysis are between algorithmic technologies and other business information technologies. At the individual level, a worker is denoted as having algorithmic literacy if they report having at least one skill from the algorithms category. Organizational measures of algorithmic literacy are then constructed by first computing the fraction of employees in a firm-year that have algorithmic literacy according to this definition ($\%ALG$). Then, for organization i in year t, a standardized measure of algorithmic literacy (ϕ^{ALG}) is defined as:

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

Similar measures are constructed using other technological skill categories (e.g., $DATA$ and NET).

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

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

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

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

3.2.2 Domain expertise

Jobs are also encoded according to whether they require domain expertise, defined as "knowledge of a specific, specialised discipline or field". The measurement of domain expertise in jobs is constructed to be consistent with the measure of algorithmic expertise as defined above. Workers are coded as having domain expertise if a domain-related skill appears in a record. The list of potential domain-related skills jobs can require is extracted from the O*NET database, which curates a comprehensive list of all of the possible domains with which US-based jobs may require workers to know.[20] These potential knowledge domains are extracted from the "Knowledge" table in O*NET which delineates "organized sets of principles and facts applying in general domains."[21] From the full list, *Computers and Electronics*, *Engineering and Technology*, *Telecommunications*, and *Mathematics* were removed because they overlap with measures of algorithmic literacy.[22]

3.2.3 Additional job characteristics

Beyond algorithmic and domain expertise, the empirical analysis uses indicators of skills related to *cognitive*, *social*, *character*, and *management* job attributes. The construction of these job attributes was based on prior work that uses the same Lightcast data source to construct these measures (Deming and Kahn, 2018). As with measures of algorithmic and domain expertise, records are coded as requiring these job attributes if the listing contains at least one related skill.[23]

3.2.4 Employers' technology investments

Obtaining consistent, firm-level measures of IT investment that span multiple years has been a persistent challenge in the academic literature (Tambe and Hitt, 2012). IT investments are not consistently recorded on balance sheets, so scholars have leveraged alternative data sources to create proxy measures, such as hardware investment measures collected by marketing surveys, IT keywords referenced in legal filings, and investments into complementary skills. The rationale behind the last

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

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

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

[23] 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.

approach is that 1) human capital is the largest component of a firm's technology investment and 2) it has become even more important for AI and data science investment because much of the relevant software is open-sourced, leaving no investment trail, and much of the hardware is cloud-based and therefore poorly measured by instruments that record the firm's owned servers and PCs.[24] On the other hand, most frontier software still requires technical expertise to install and maintain, so quantity measures of complementary, technical human capital may be the most accurate available proxy measure of a firms' technology investments. Studies using wages or worker quantities as a proxy measure of the firms' IT investment have used aggregate numbers of IT workers employed by the firm or, when measuring investment in specific technologies like machine learning, workers that possess a related skill (Lichtenberg, 1995; Brynjolfsson and Hitt, 1996; Tambe, 2014). This paper takes this approach. It follows methods documented in Tambe and Hitt (2012) where proxy measures of investment variables are constructed as quantities of workers at a firm that have skills relevant to the technological domain.[25]

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

4 Results

4.1 Model-Free Evidence

4.1.1 Job listings

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

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

[25] Like most firm-level measures, this approach records investments with measurement error. See Appendix C.

Figure 2a shows the extent to which specific technology skills are bundled with domain expertise in one month of the job listing data (January 2016). Skills with higher values (lines that reach further to right) tend to appear in jobs that also often list domain expertise in their requirements. Skills colored dark blue are classified as algorithms. Algorithmic skills in this diagram are more commonly bundled with domain expertise and appear to have more in common with skills like Excel and ERP systems commonly used by business-facing occupations. In Figure 2b, a higher value indicates skills that are more likely to appear in a broader array of occupations. Algorithmic skills, which are those colored in dark blue, are more diversified across occupations than other technical skills. Skills related to predictive analytics, data science, and data analysis are particularly dispersed and only slightly less so than skills related to the Microsoft Office Suite. This finding is consistent with the claim that employers are bundling algorithmic skills in occupations with domain expertise.

4.1.2 Changes in workforce composition

Analysis of job openings is valuable because it (i) indicates employer preferences and (ii) reflects immediate adjustments by employers. In this sense, it serves as a leading indicator of labor market changes. However, these data cannot tell us whether the listings indicate hard requirements or a "wish list" from employers, or whether the vacancies requiring these skills are ultimately filled. I turn to corporate workforce data to investigate whether changing employer preferences were met by market adjustments. The results from these analyses are shown in the four quadrants of Figure 3.

Figure 3a illustrates how the dispersion of different categories of technical skill in business-facing occupations changes over time in these firms. The y-axis is the intensity with which a skill appears in these occupations and levels are depicted relative to their 2008 base rates. The trend line for algorithmic skills, depicted in blue, indicates steady growth in the rate at which AI and data science skills have penetrated these occupations. By 2021, these skills appear in about 10% more occupations than they did in 2010. In contrast, technologies like those related to networks and the cloud became increasingly specialized. Fewer workers in business occupations needed skills related to these technologies. The incidence of mobile skills in this sample remained flat.

Figure 3b shows how the measures of algorithmic skills in these occupations (referred to as $ORGALG$) varies across different industries. Perhaps unsurprisingly, it is highest in the Information and Professional Services industries, which include technology and finance firms. This is consistent with evidence on the prevalence of these technologies in these industries (Lohr, 2024). Retail has climbed rapidly, reflecting the growing use of consumer data for prediction. Levels are lower in the Arts and Health industries although they have been climbing in Healthcare reflecting the growing use of AI in healthcare domains.

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

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

4.2 Regression tests of adjustments in firms' hiring patterns

Figure 4a tests whether algorithmic skills are more likely to appear in jobs requiring domain expertise. The unit of observation i is the job listing, and the dependent variable is a binary indicator of whether a listing i requires the applicant to have job attributes related to i) domain expertise and for comparison: ii) cognitive attributes, iii) social skills, iv) character, and v) management abilities. The right-hand side includes measures of whether the listing reflects a need for algorithmic skills as well as, for comparison, skills in two different comparison technologies, databases and networks.

$$ATTR_i = \beta_A ALG_i + \beta_D DAT_i + \beta_N NET_i + \gamma_i + \epsilon_i \tag{2}$$

Equation B also includes a vector of control variables (γ) that includes job title, industry, and a measure of the logged number of skills in the listing and i indexes the listing. Figure 4a depicts the estimates of β_A from Equation B. We observe positive correlations between algorithmic skills and domain expertise. Because these tests include job-title fixed effects, positive correlations mean that algorithmic literacy tends to be bundled in jobs requiring domain expertise (**Hypothesis 1**). The full form of these regressions, which can be found in Appendix B, also indicate that correlations with database tasks are negative, which is consistent with that class of skills being more specialized within IT work. There are also positive correlations between the use of algorithmic technologies and cognitive job attributes and negative correlations with management-related job attributes. This negative relationship suggest that employers are not bundling people leadership (character and management) with algorithmic skills.

Next, I investigate where in the data decision-making pipeline this specific combination of skills is important by testing if domain expertise is more likely to accompany algorithmic literacy in job listings for certain tasks in the data-driven decision-making pipeline. For a single month of the Lightcast data, the logistic regression used to evaluate this relationship is:

$$TASK_i = \beta_{DA}(DOM_i \times ALG_i) + \beta_D DOM_i + \beta_A ALG_i + Log(No.\ Skills)_i + \epsilon_i \tag{3}$$

Figure 4b illustrates the estimates (β_{DA}) from this regression. The estimates on *Presentation* and *Decision-making* are consistent with the statement that algorithmic literacy and domain expertise are complements when using data to make decisions. By contrast, *Data management* and *Data modeling* are negatively correlated with this combination which suggests these tasks are more often localized to specialist technical jobs in the vacancies sample. Within the job listing data, these results suggest that employers are searching for both domain expertise and algorithmic literacy from their decision makers.

4.3 Algorithmic literacy and domain expertise as decision complements

We can now turn towards the question of whether these demand changes are met by supply-side adjustments in the labor market. Turning to the data on changes in workforce composition, Table 3 reports tests of how markers of algorithmic literacy have diffused through firms and occupations. Columns (1) through (4) report correlations between decision-making, as indicated by the O*NET measure of decision-making importance for an occupation, and algorithmic literacy. All regressions are conducted at the firm-occupation-year level for a panel of firms spanning the years 2015 to 2021. Column (1) indicates that workers in occupations requiring higher levels of local decision-making are more likely to have algorithmic literacy (t=3.50) (**Hypothesis 2**). Columns (1) and (2) indicate that this is true with and without firm fixed-effects which accounts for the potential confounder that firms with greater algorithmic investment have more workers in decision-making occupations.

Although the relationships discussed to this point are correlational, we can test the idea that a shift in the cost of using these tools accelerates the diffusion of these skills across the firm. In columns (3) and (4), we separate algorithmic skills according to whether the skill corresponds to a "no-code" tool. These tools are a minority of all algorithmic skills, so the main-effect in column (3) is negative (t=-28.88). The estimates in column (4), however, indicate that lower costs for using these tools accelerates the diffusion of these technologies into occupations that require decision-making (t=4.29) (**Hypothesis 3**). Although not conclusive, this evidence is supportive of a causal relationship between the costs of human capital acquisition and the bundling of these skills with domain expertise.

Finally, columns (5) and (6) aggregate by occupation and test these relationships at the firm level. In these firm-year regressions, the dependent variable is the standardized algorithmic skill measure ($ORGALG$) for the organization. A higher measure indicates that a greater fraction of domain experts in the firm has algorithmic literacy. The estimates show robust correlations between investment in algorithms and the dispersion of algorithmic skills in the organization, after conditioning on other assets. This persists whether firm fixed-effects are omitted (t=4.62) or included (t=2.34), which means that firms that rely on algorithms have a greater degree of algorithmic literacy among its domain experts, even after accounting for other static sources of firm differences. This is consistent with the assertion that employers see value in contemporaneously investing in these two factors. The next section investigates whether the financial market also places value on this investment pattern.

4.4 Algorithms, workforce composition, and market value

Finally, I test if the market assigns higher value to firms that have brought these factors together. To do so, this section discusses results from a regression of a firm's market value on its various assets.

Although not causal in its interpretation, this hedonic framework has been used in prior work to decompose a firm's value into its components and test whether firms are producing intangible assets that can raise the returns to their IT investments (Brynjolfsson et al., 2002). Market value is a particularly meaningful dependent variable for an analysis of AI and data science returns because firms need time to adjust these new technologies to their workflow and the literature suggests that firms are not yet consistently realizing value from these investments. Market value figures, however, have the benefit that investors assign value based on the future stream of benefits they will produce. The regression tested in this table is:

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

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

Table 4 reports results from estimating Equation 4 on a seven year panel of public firm investments (2015-2021). Columns (1) through (3) in Table 4 have year and 3-digit industry fixed-effects. Column (1) reports results of logged market value only on logged measures of IT investment, assets, capital (PPE), and employment. Like prior work, these estimates suggest that the market assigns value to investment in digital technologies (t=3.79). The next column adds a measure of investment in algorithms. After adding this measure, the IT coefficient falls to zero, suggesting that most of the market returns to IT investment are from firms that invest at the frontier, in as much as this frontier in this panel is represented by AI and data science investments (t=3.93).

Column (3) reports estimates from the full form of Equation 4, which includes the interaction terms between the algorithmic and organizational skill measures. The main effect of investment in algorithmic technologies is similar to column (2) (t=3.80) but the interpretation of the interaction term is that these investments are valued an additional 15% higher in firms that are one standard deviation higher in the skill measure (t=2.67) (**Hypothesis 4**). These estimates suggest that firms that invest in algorithms and that concurrently employ domain experts with algorithmic skills are building valuable intangibles that will be useful for producing a stream of AI goods and services in the future. Column (4) uses the same specification, except that it includes industry fixed-effects at the four-digit level. The estimate on the interaction term is essentially unchanged (t=2.67). Finally, including firm fixed-effects in column (5) renders the interaction coefficient insignificant.

The analysis to this point has grouped data science and AI investments into a single category. Table 5 separates these investments into AI and data science and further separates the panel into different time periods comprised of the years before and after 2018. The first two columns suggest there is a stronger relationship between data science investment and the firm's market value in the earlier period (t=5.21). In the later period, the value associated these two investments begins to even out. The estimates on both AI (t=3.19) and Data Science (t=2.95) are positive and significant. The last two columns introduce the interaction terms with the organizational skill measures, where a different organizational skill measure is constructed for AI and data science skills. In the early period, neither coefficient on the interaction terms are significantly different than zero. In the latter period, however, the coefficient on the AI interaction term suggests that a one standard deviation higher organizational skill measure raises the value of AI investment by about one-third (t=3.18).

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

4.5 Robustness tests

In a final set of comparisons, Figure 5 reports the estimate on the main interaction term from Equation 4 where the sample is split into size categories according to employment. This figure indicates that the market rewards that firms receive from concurrently investing in these two factors are highest for the largest firms in the sample.

The evidence suggests investments in data science complement algorithmic skills in business and management occupations. I also conduct robustness tests that evaluate whether we find similar results when substituting measures based in different technologies or in different skills. Figure 6a shows the coefficient estimates on the interaction term on ($TECH_{it} \times ORGALG_{it}$) from Equation 4 when using $TECH$ measures based in AI (which is the same as in our main regression), which we compare with measures based in infrastructure technologies like networks and databases. If workforce changes in business-facing occupations are particularly important for AI technologies, we should only observe meaningful correlations for this skill-based interaction term with AI investment, not with investments that firms make in other technologies. We do observe such a pattern in Figure 6a. Neither of the coefficients on the other interaction terms are significantly different than zero. This pattern of estimates supports the argument that a correlation between market value and the interaction term between AI skill and AI investment is not simply reflecting the type of heterogeneity that would be picked up by other measures of technology investment. In other words, the correlations we observe between value and the interaction between skills and AI investment are specific to AI, not general technology investment.

Figure 6b performs a similar comparison, again using the specification in Equation 4, but instead of altering the technology measure, it retains AI as the technology measure for all regressions and varies the skill measure within the organization. It reports the coefficient estimate on the interaction term where the $ORGALG$ measure in business-facing occupations is generated using the prevalence of network and database technology skills in this segment of the workforce, rather than AI skills. Again, we see that the market value correlations are only different from zero when investment in AI technologies is accompanied by AI skills in this layer of the workforce. General tech acumen dispersed within the business layer of the workforce does not appear to raise returns to AI investments.

5 Managerial Implications

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

A corollary is that the costs of using AI and data science technologies are continuously falling. Conversational interfaces driven by generative AI, for instance, represent a shift in how knowledge workers interact with information technologies, making them more accessible and user-friendly than ever before. A fall in the costs of using these tools suggests that employers can accelerate the rate at which data analysis tasks are pushed to domain experts. The implications of this shift for both managers and educators can be significant. From a management perspective, no-code and generative AI tools can democratize technical skill, enabling a more diverse range of employees to contribute to areas that were once the exclusive domain of technical specialists. This can lead to more innovative environments that emphasize the productive combination of human creativity and computational power. Managers, in turn, may need to adapt by focusing less on specific technical skills when hiring and more on general problem-solving abilities and adaptability.

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

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

6 Conclusions

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

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

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

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

References

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* .

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Lohr, S. (2017), 'Where non-techies can get with the programming', *The New York Times, April* **4**.

Lohr, S. (2024), 'Generative A.I.'s biggest impact will be in banking and tech, report says', *The New York Times, Feb* .

Mao, Y., Wang, D., Muller, M., Varshney, K. R., Baldini, I., Dugan, C. and Mojsilović, A. (2019), 'How data 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Tambe, P. 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.

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

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

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

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

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

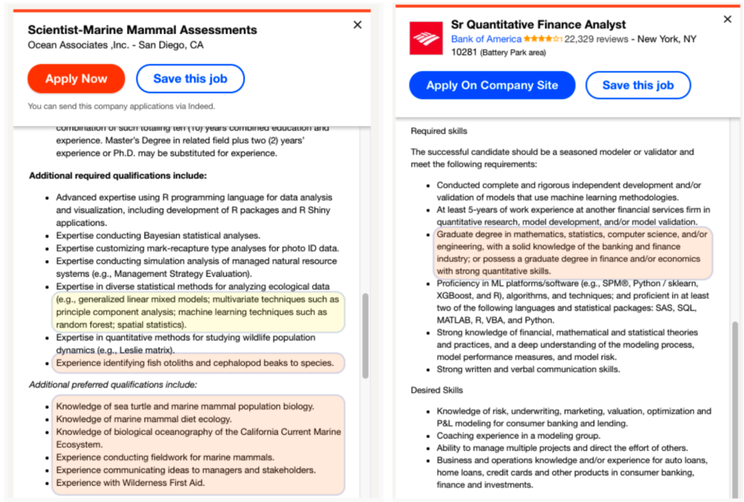
Wirth, R. and Hipp, J. (2000), Crisp-DM: Towards a standard process model for data mining, *in* ‘Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining’, Vol. 1, Manchester, pp. 29–39.

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

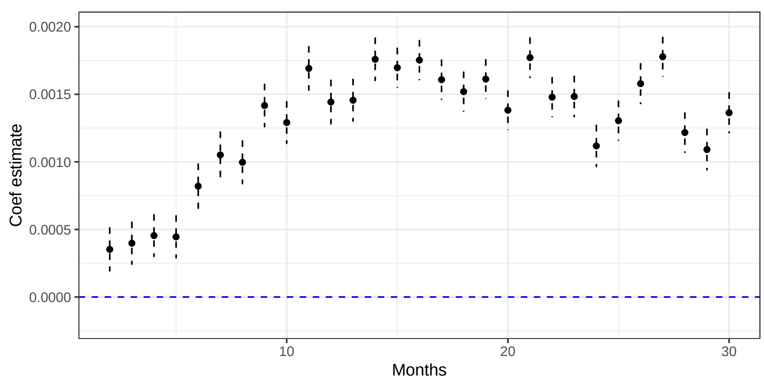
Xue, M., Cao, X., Feng, X., Gu, B. and Zhang, Y. (2022), ‘Is college education less necessary with AI? Evidence from firm-level labor structure changes’, *Journal of Management Information Systems* **39**(3), 865–905.

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

Figure 1: The growth of algorithmic skills in job listings



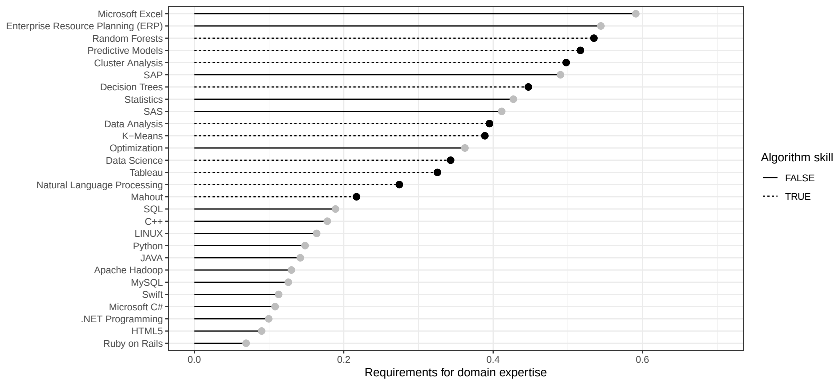
(a) Sample listings with algorithmic and domain expertise



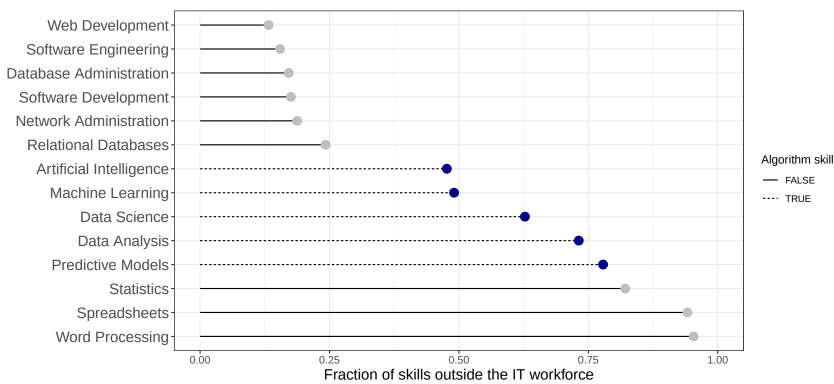
(b) Likelihood an algorithmic skill appears in a job listing

Figure notes: Figure (a) shows two sample listings for jobs requiring familiarity with both algorithmic tools (highlighted in yellow) and domain expertise (highlighted in orange), related in these examples to marine biology and finance. These listings and screenshots were extracted from the website Indeed.com. Figure (b) shows coefficient estimates and standard error bars on the regression $ALG_i = \beta month_i + \epsilon_i$ for the months from 2013 onwards in the Lightcast data (where coefficients reflect differences from the Jan 2013 baseline month), i indexes job listings, and ALG takes the value 1 if a listing contains an algorithmic skill and 0 otherwise. Standard error bars show the 95% confidence interval.

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



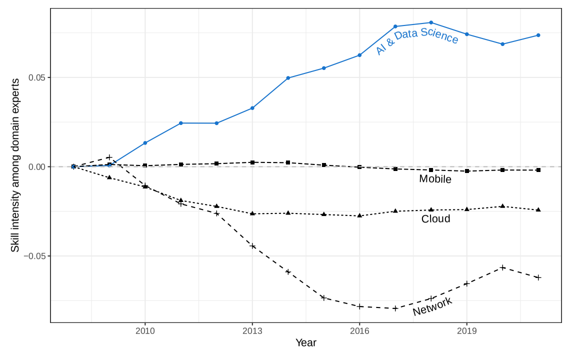
(a) Bundling of technical skills with domain expertise in job listings



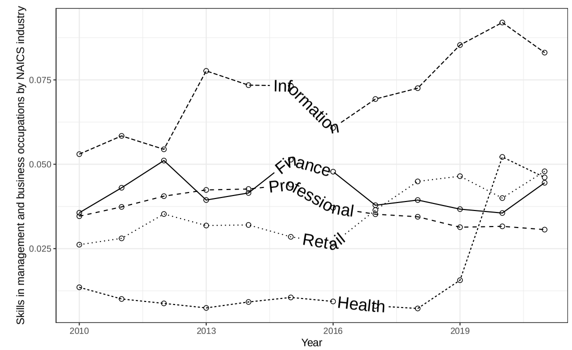
(b) Fraction of skills that appear in job listings for non-IT occupations

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

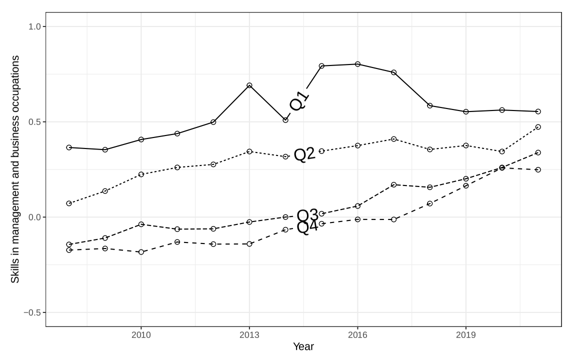
Figure 3: Changes in the locus of technical skills in organizations from 2008-2021



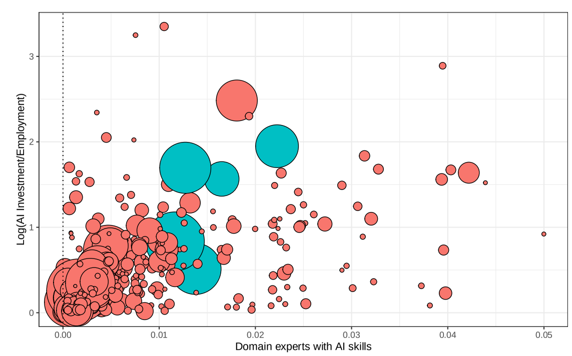
(a) Changes in skills in organizations



(b) Algorithmic skill measure by industry



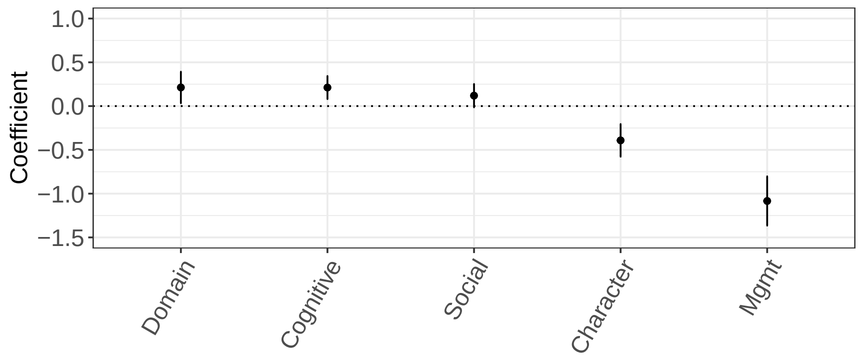
(c) Convergence in skill measure across firm size quartiles



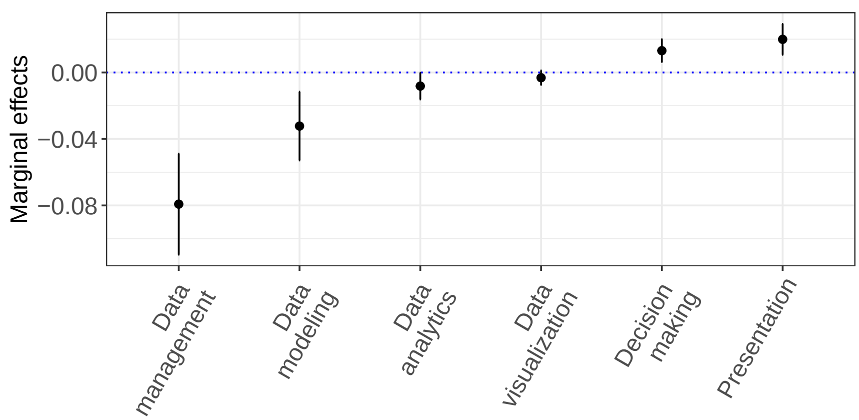
(d) Organizational accumulation of skills (AI)

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

Figure 4: Characteristics of job listings with algorithmic skills



(a) Job characteristics and algorithmic skills



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

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

Table 1: Summary statistics for firms in the workforce panel (2018)

Variable	Units	Mean	Std. Dev.	N
Market value	Millions (USD)	57588.6	213059.0	1179
Assets	Millions (USD)	60851.3	259501.3	1179
Prop., Plant, and Equip. (PPE)	Millions (USD)	4115.9	12358.4	1179
Total Employment	Thousands (Employees)	29.7	87.7	1179
Databases	Tech skill count	1220.4	6217.0	1179
Networks	Tech skill count	320.8	1625.6	1179
Algorithms	Tech skill count	233.9	1105.3	1179
AI	Tech skill count	72.6	454.2	1179
Data science	Tech skill count	175.7	741.2	1179
IT	Tech skill count	9180.1	41890.8	1179
ϕ^{ALG}	Standardized Value	0.0	0.9	1179

Table notes: This table reports summary statistics for firms in the 2018 cross-section of the regression panel. The year 2018 was chosen because it is 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 [*Data capital, Network capital, Alg capital, IT capital, ORGALG*] 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	38
22	Utilities	36
23	Construction	12
31-33	Manufacturing	432
42	Wholesale Trade	42
44-45	Retail Trade	68
48-49	Transportation and Warehousing	39
51	Information	309
52	Finance and Insurance	327
53	Real Estate and Rental and Leasing	31
54	Professional, Scientific, and Technical Services	67
56	Administrative and Support and Waste Management Services	31
61	Educational Services	9
62	Health Care and Social Assistance	29
71	Arts, Entertainment, and Recreation	5
72	Accommodation and Food Services	25
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.

Table 3: Conditional correlations between job attributes and algorithmic skills

	ALG				ϕ^{ALG}	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Decision-making × *NOCODE*				0.120***		
				(0.028)		
Decision-making	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***	−0.038*	0.172
	(0.012)	(0.021)	(0.011)	(0.011)	(0.022)	(0.108)
Log(Alg)					0.039***	0.089***
					(0.006)	(0.028)
Log(Assets)					0.017	0.069
					(0.022)	(0.052)
Log(PPE)					0.006	−0.015
					(0.015)	(0.038)
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	0.095	0.408
Observations	160,614	160,614	321,228	321,228	7,901	7,900

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

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

	Log(Market Value)				
Model:	(1)	(2)	(3)	(4)	(5)
Variables					
Log(Assets)	0.712^{***}	0.706^{***}	0.706^{***}	0.722^{***}	0.564^{***}
	(0.052)	(0.052)	(0.052)	(0.058)	(0.037)
Log(PPE)	0.085^{**}	0.084^{**}	0.084^{**}	0.085^{*}	0.081^{***}
	(0.040)	(0.040)	(0.040)	(0.045)	(0.026)
Log(IT)	0.053^{***}	−0.004	−0.004	0.004	0.034^{***}
	(0.014)	(0.019)	(0.019)	(0.019)	(0.013)
Log(Employment)	0.001	0.005	0.006	0.005	0.227^{***}
	(0.046)	(0.044)	(0.044)	(0.046)	(0.047)
Log(Alg)		0.056^{***}	0.055^{***}	0.045^{***}	−0.007
		(0.015)	(0.015)	(0.016)	(0.007)
ϕ^{ALG}			−0.003	−0.008	−0.002
			(0.010)	(0.010)	(0.007)
Log(Alg) × ϕ^{ALG}			0.006^{*}	0.007^{**}	0.002
			(0.003)	(0.003)	(0.003)
Fixed-effects					
Firm FE					Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Industry FE (NAICS 4)				Yes	
Industry FE (NAICS 3)	Yes	Yes	Yes		
Fit statistics					
R^2	0.869	0.871	0.871	0.890	0.974
Observations	6,985	6,985	6,985	6,985	6,984

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

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

DV	Log(Market Value)			
Years	2015-2017	2018-2021	2015-2017	2018-2021
Model:	(1)	(2)	(3)	(4)
Variables				
Log(Assets)	0.735^{***}	0.693^{***}	0.734^{***}	0.690^{***}
	(0.051)	(0.045)	(0.051)	(0.045)
Log(PPE)	0.079^{*}	0.068^{*}	0.079^{*}	0.059
	(0.042)	(0.039)	(0.042)	(0.038)
Log(AI)	0.040	0.102^{***}	0.042	0.095^{***}
	(0.027)	(0.032)	(0.027)	(0.031)
Log(Data Science)	0.125^{***}	0.115^{***}	0.124^{***}	0.108^{**}
	(0.024)	(0.039)	(0.023)	(0.041)
Log(IT)	-0.107^{***}	-0.111^{**}	-0.107^{***}	-0.101^{**}
	(0.031)	(0.048)	(0.032)	(0.047)
Log(Employment)	0.008	-0.005	0.010	0.008
	(0.046)	(0.046)	(0.046)	(0.042)
ϕ^{AI}			0.022	-0.086^{***}
			(0.016)	(0.026)
ϕ^{DS}			-0.021	0.014
			(0.023)	(0.043)
Log(AI) $\times$ ϕ^{AI}			-0.012^{*}	0.035^{***}
			(0.006)	(0.011)
Log(Data Science) $\times$ ϕ^{DS}			0.005	0.012
			(0.006)	(0.011)
Fixed-effects				
Year FE	Yes	Yes	Yes	Yes
Industry FE (NAICS 3)	Yes	Yes	Yes	Yes
Fit statistics				
R^2	0.906	0.854	0.906	0.858
Observations	3,852	2,340	3,852	2,340

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

Figure 5: Estimates on ϕ^{ALG} by employer size quartile

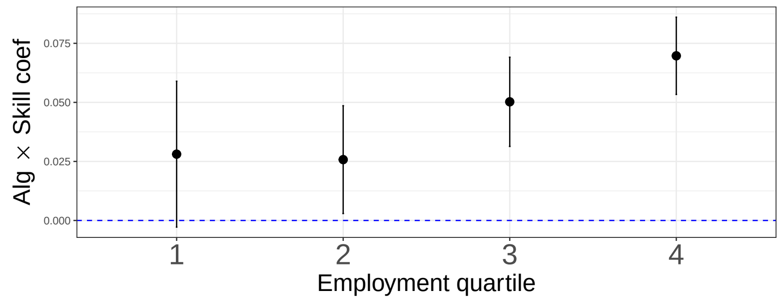
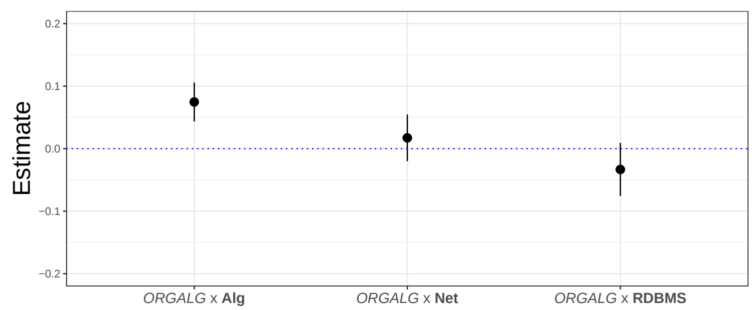
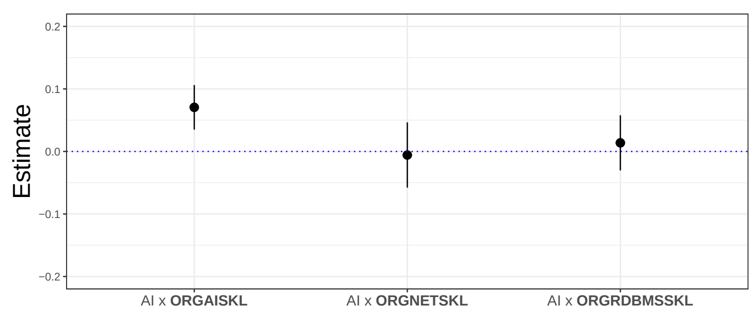


Figure notes: The y-axis indicates the coefficient on the interaction term between algorithm investment and *ORGALG* from the main specification used in column (4) of Table 4 where the sample is divided into quartiles by size. The sample size in each regression is roughly one-quarter the size of that from column 4 of Table 4. The x-axis divides firms by employment quartile where "1" is the smallest firms in the sample and "4" is the largest firms in the sample. Standard error bars indicate the 95% confidence interval.

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



(a) Comparisons with investment in other technological classes



(b) Comparisons with organizational measures of other technical skills

Figure notes: The top facet illustrates placebo tests from interacting $ORGALG$ with investment in other technology classes. It displays interaction terms from the regression in column (4) of Table 4 ($Log(MV)_{it} = Log(Assets)_{it} + TECH_{it} + ORGALG_{it} + (TECH_{it} \times ORGALG_{it}) + \gamma_{it} + \epsilon_{it}$). The marker on the left is for AI investment, the middle one is for networks, and the one towards for the right is relational databases. The bottom facet takes a similar approach but uses measures of organizational skill dispersion in other technologies. These measures are based in algorithms, networks, and relational databases, respectively. The standard error bars in both figures indicate 95% confidence intervals.

A Description of corporate workforce data

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

A.1 Data generating process and sampling frame

Revelio is a workforce intelligence company that federates data across a range of Internet sources including federal databases, professional networking sites, and job posting aggregators. This analysis relies on their workforce, position, and skill databases which contain data on the movements of an extremely large sample of US-based employees across firms, the job titles they hold, and the skills they acquire. Data on employment spells, at scale, are not otherwise collected by government agencies. They are only available through resume banks so these types of data are particularly useful for studying quantities of workers in firms with different skills and the flow of workers of different types between organizations.

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

A.2 BLS-SOC share comparisons

The distribution of Revelio workers across occupations is shown in Figure 7a. Figure 7b presents differences in shares of the major occupational groups as reported by the BLS and represented in the Revelio data, where the assignment of workers to SOC areas in the Revelio data is provided by Revelio. The blue line indicates no (zero) difference in shares such that bars to the right (left) are those occupations where the occupation accounts for a higher (lower) proportion of workers in the BLS data than the Revelio data.

From this comparison, we can see that "white-collar", knowledge-intensive occupations like management and Information Technology work tend to be over represented in the Revelio data set whereas front-line occupations in sectors like manufacturing, production, and transportation are underrepresented. This is not a surprise given that these data are gathered from professional networking sites on which white-collar workers tend to be over represented. The length of each bar is the difference in shares across these data sources. The largest imbalance in occupations in Management. The difference in the share of total workers that managers account for in the Revelio data set (15%) and the BLS (7%) is about 8% percentage points.

A.3 NAICS Industry comparisons

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

A.4 Geographic (state) comparison

A final comparison, shown in the bottom panel (Figures 7e and 7f) is state-level comparisons. This comparison evaluates the reported geographic location of workers in the Revelio data set with the distribution of workers across US states. Unsurprisingly, we can see that states with significant industry representation for finance and technology (such as New York) account for a relatively larger share of workers in the Revelio data. The largest imbalance is in North Dakota, where industries like oil extraction and agriculture play a larger role in the state economy.

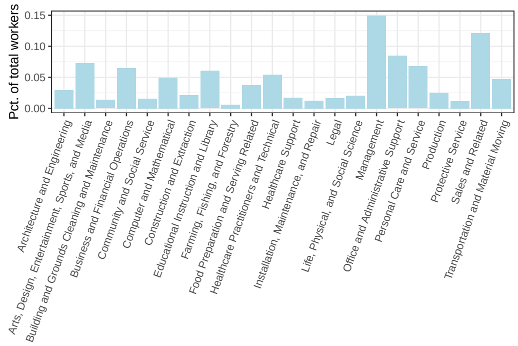
A.5 Discussion

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

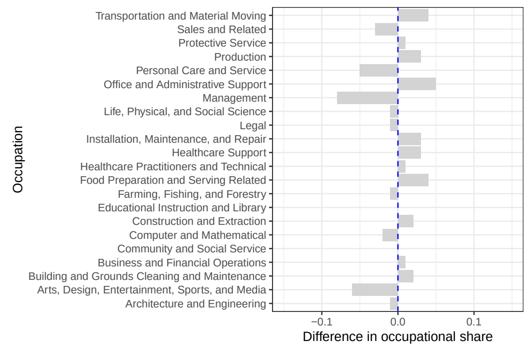
Having greater quantities of workers in this database from some sectors and occupations will affect the precision of the measurement, but this may fall into under normal, random measurement error if those workers who do report their skills are not very different from the ones who do. The number of workers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

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

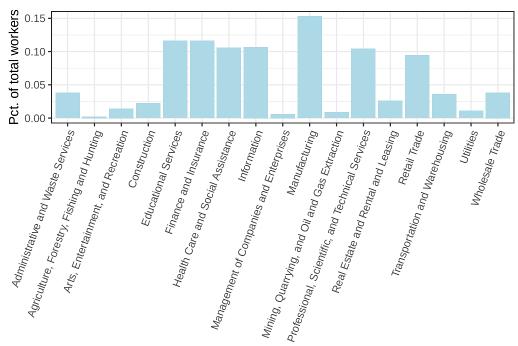
Figure 7: Revelio data distributions



(a) Occupational code (SOC)

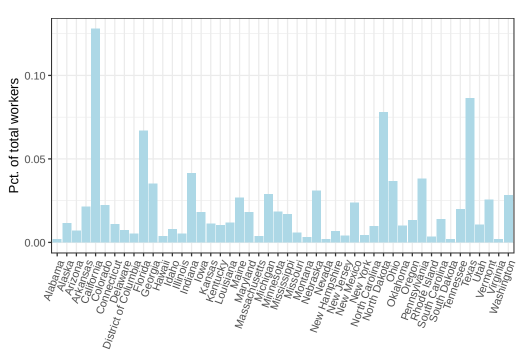


(b) Occupational code (SOC)

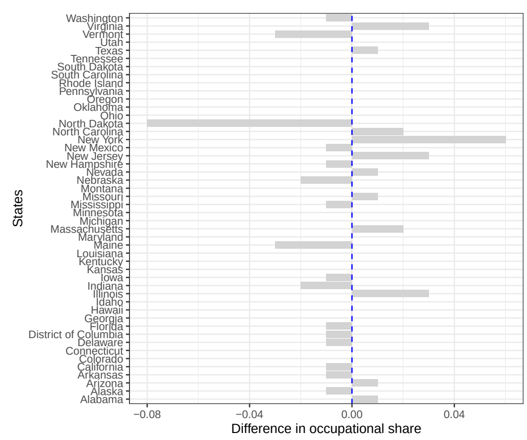


(c) Industry (NAICS)

(d) Industry (NAICS)



(e) States



(f) States

Notes: These three figures illustrate the difference in compositional shares between the Revelio and BLS data sets. The top row is comparison of occupations. The second row is comparison across NAICS industries. The third row is comparisons across states. The length of each bar for plots in the second column is computed as the difference in the share that the worker category accounts for in the Revelio data and in the administrative data. For instance, Management workers comprise 15% of the Revelio data set and 7% of the BLS data set so the length of the bar indicates an 8% difference between the two.

B Additional results

In this section, we present the full form of some shortened tables discussed in the text. Table B.1 corresponds 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 shows that algorithmic skill is correlated with domain expertise and cognitive skill. There are also negative correlations with Character and Management, after conditioning on job title and industry.

This regression also includes measures of skills related to databases and network administration. Database management is negatively correlated with all dependent variables. Network administration exhibits relatively weak correlations with all of these job attributes.

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

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

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

C 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, I principally 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.

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.

C.1 Technology categories created from skills in the job listings data

Algorithmic Technologies. 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;; machine learning, natural language processing, image processing/computer vision, artificial intelligence, tensorflow, pytorch, scikit-learn

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 Technology (IT) and Software Development. 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

C.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 & Network*]. 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