RAND CORPORATION

Research Report

TOBIAS SYTSMA, ÉDER M. SOUSA

Artificial Intelligence and the Labor Force

A Data-Driven Approach to Identifying Exposed Occupations

KEY FINDINGS

- By 1989, all occupations were exposed to technology patents to some extent, and there are no occupations in the United States that are completely unexposed to general technology patents.
- Work tasks that are typically carried out less frequently have the highest technology patent exposure.
- By 2020, nearly all occupations had been exposed to technology patents involving *artificial intelligence (AI)*, defined as the ability of computers and machines to simulate human intelligence, to some degree, but the level of exposure varies across occupation groups, time frame, and technology categories.
- In 2019, up to 15 percent of workers were employed in occupations that were highly exposed to AI technologies.
- The nature of occupational exposure to technology patents has changed over time. In contrast to earlier decades, occupations that require more education and pay higher wages have become more exposed to technology patents in general.
- Greater exposure to natural language processing, speech recognition, and evolutionary computation technology patents is associated with declines in employment growth occupations that specialize in more-routine tasks.

Artificial intelligence (AI) is a rapidly evolving technology with enormous economic potential. According to a recent McKinsey report, new generative AI tools could contribute trillions of dollars to the global economy by 2040 through productivity growth (Chui et al., 2023). This potential is reflected in the growing interest in AI tools by the private sector, as evidenced by the increase in mentions of "generative AI" in global corporate earnings call transcripts from just five in December 2022 to 390 in June 2023 (FactSet, undated). Recent suggestions that there has been a surge in demand for workers with AI skills indicates that AI is already creating job opportunities (Acemoglu, Autur, et al., 2022).

However, with these opportunities comes concern. Recent reports show that more than 4,000 jobs were cut in the U.S. labor market in May 2023 because of AI (Challenger, Gray, and Christmas, Inc., 2023), which raises questions about the impact of AI on workers and labor markets. Some estimates suggest that a significant portion of work activities could be automated over the next several decades (Chui et al., 2023; Eloundou et al., 2023). Additionally, the distinct features of AI suggest that its effect on the job market might deviate from that of prior waves of automation (Brynjolfsson and McAfee, 2014). Specifically, unlike previous technologies, AI can automate tasks that were previously considered hard to codify, which creates the potential for a broader variety of tasks and occupations to be automated (Manyika et al., 2017; Brynjolfsson, Mitchell, and Rock, 2018; Chui et al., 2023; Eloundou et al., 2023; Webb, 2019). However, exposure to AI does not necessarily lead to labor market displacement; technologies might reduce the cost of hiring and create new job opportunities for some workers. Consequently, there is a growing need for research to understand the implications of AI on workers, firms, and markets. In this report, we focus on the United States, but these results are broadly applicable to other countries.

To address this pressing need, we aim to answer the following research question: What is the relationship between (1) an occupation's exposure to general and specific AI-related technologies and (2) wages, and employment? Using a methodology that identifies the exposure of occupational tasks to relevant technology patents, we provide insights into the potential implications of AI on the labor market and inform policy discussions around this emerging issue. More specifically, we use natural language processing (NLP) to identify semantic similarities between job task descriptions and technology patents awarded between 1976 and 2020.[1] We contribute to the growing literature in this area by evaluating occupation exposure to various AI technology categories, including computer vision, evolutionary computation, AI hardware, knowledge processing, machine learning, NLP, planning and control, and speech recognition.[2] We also provide new insights into dynamics in exposure over time by identifying the types of occupations that have become more (or less) exposed to different forms of technology over the past 40 years. We define *exposure* as using the technology to perform occupational tasks.

Our findings suggest that, in the United States, exposure to all technology patents—as well as to AI-specific patents—is not uniform across occupational groups, over time, or across AI technology categories. For instance, occupations that are the most exposed to speech recognition and NLP technologies generally involve communication, writing, and active listening, while the occupations most exposed to planning and control AI generally involve business and finance. However, several broad patterns emerge. In general, occupations that require more education and cognitive skills have become more exposed, while those that require manual labor have become less exposed. Many of the occupations that are currently highly exposed to AI technologies are the same occupations that are expected to grow over the coming decade. Moreover, the relationship between wage distribution and exposure has changed over time, with the highest exposure being associated with occupations in the upper end of the wage distribution, as of this writing. Overall, we estimate that up to 15 percent of workers were highly exposed to AI technology patents by 2019.

We also estimate a series of regressions aimed at better understanding the correlation between technology exposure (to general and AI patents) and employment growth. Although we do not find evidence of a statistically meaningful correlation between general technology patent exposure and employment growth, we do find that exposure has a positive correlation with employment growth for some AI technology categories. However, we also find that the correlation with employment growth can depend on the relative importance of routine tasks compared with nonroutine tasks (routine intensity) within the occupation. For example, for less routine-intensive occupations, exposure to NLP, speech recognition, and evolutionary computation has a positive correlation with employment growth since 1990. However, for more routine-intensive occupations, the correlation between exposure to these technologies

LIMITATIONS

- The matching was done using one large language model (LLM), and the results could change if a different model were used.
- We do not assess changes to the task structure within occupations, which could also respond to exposure to technology.
- Our analysis does not extend past 2020, and, as a result, we do not account for patents granted over the past several years.
- Some new technologies, such as ChatGPT, do not have a patent. Instead, our findings are based on an academic research report. As a result, our analysis does not include all forms of modern AI technology platforms.
- Our analysis does not assess technology adoption because patents might not translate into actual technology development and adoption by employers.

and employment growth is negative. This finding suggests that technology can either substitute for or complement worker tasks. Specifically, technology might increase the productivity of some workers and allow firms to expand hiring in some areas while simultaneously replacing other workers that tend to specialize in more-routine tasks. Additionally, this finding differs from the presumption that occupations that specialize in more-complex cognitive tasks might be at greater risk of disruption from AI because the ultimate employment impact depends on whether the AI technology is complementary with or substitutable for the worker's tasks. Although we have not yet witnessed the full labor market effect of AI, so far, the occupations that have seen job loss tend to be more routine in nature.

Although this report sheds light on the relationship between technological progress and the labor market, our analysis has several important limitations. Primarily, we do not assess changes to tasks within occupations; instead, we focus our analysis on changes in the exposure of various occupations' existing task structures, as of the time of this writing. This introduces a potential source of endogeneity because technology might change the task content of occupations, which would make some tasks more or less important relative to others. However, our results are consistent with those in the prior literature on the effect of technology in the labor market. Overall, the results of the analysis should be considered correlative rather than causal.

This study contributes to several lines of existing research. There are a handful of studies that use patent-task matching methods to measure technology exposure. These studies use a variety of methods to evaluate the similarity between U.S. patents and job task descriptions. For example, Webb (2019) uses verb-noun pair matches, Montobbio et al. (2021) uses a "bag of words" approach, and Kogan et al. (2021) uses word embeddings. We use a version of the Bidirectional Encoder Representations from Transformer (BERT) developed by Google that has been fine-tuned specifically for patent analysis. Using BERT for Patents allows us to achieve better text similarity matches than such methods as verb-noun pair matching because BERT attends to the whole input sequence instead of using the previous sequence of words to predict the next word.

We also contribute to the growing literature on the labor market implications of AI. Over the past several years, several studies have estimated that many occupations are exposed to AI technology. Early research used expert elicitation to identify occupations exposed to emerging technology and found that 47 percent of employment is at risk of displacement because of computerization (Frey and Osborne, 2017). More recently, Eloundou et al. (2023) used expert elicitation and Generative Pretrained Transformer 4 (GPT-4) to identify exposure to LLMs.[3] The authors of that report estimate that 80 percent of the U.S. workforce could see 10 percent of their work tasks automated by LLMs. Webb (2019) used patent-task matching to identify occupations exposed to AI, broadly defined, and found that higher-wage occupations were more exposed. We contribute to this literature by assessing exposure to several different forms of AI technologies, tracing out exposure over the past several decades, and estimating how exposure correlated with employment growth.

Finally, this report contributes to the large literature on the role of technology in the labor market. Prior research has found evidence that more-routine occupations are typically at greater risk of displacement because of technology. For instance, Autor, Levy, and Murnane (2003) found that computerization is associated with declines in the routine and manual tasks of labor but an increase in nonroutine cognitive tasks. Other studies found a similar relationship between technology and the composition of the labor force in which technological growth is associated with job growth among nonroutine cognitive tasks and job loss among routine tasks (Autor, 2015). Lastly, recent research used administrative data to show that larger firms typically adopt automation technologies and, consistent with prior work, found that adopting firms have higher productivity but lower labor shares (Acemoglu, Anderson, et al., 2022).

The organization of this report is as follows. First, we discuss our process of matching patent titles to job task descriptions. Then, we discuss how we aggregate the patent-task matches to occupation-level exposure scores and use these scores to evaluate how

exposure has evolved over time and across occupations. Lastly, we use employment and wage data to assess the correlations between exposure and labor market outcomes.

Datasets Used

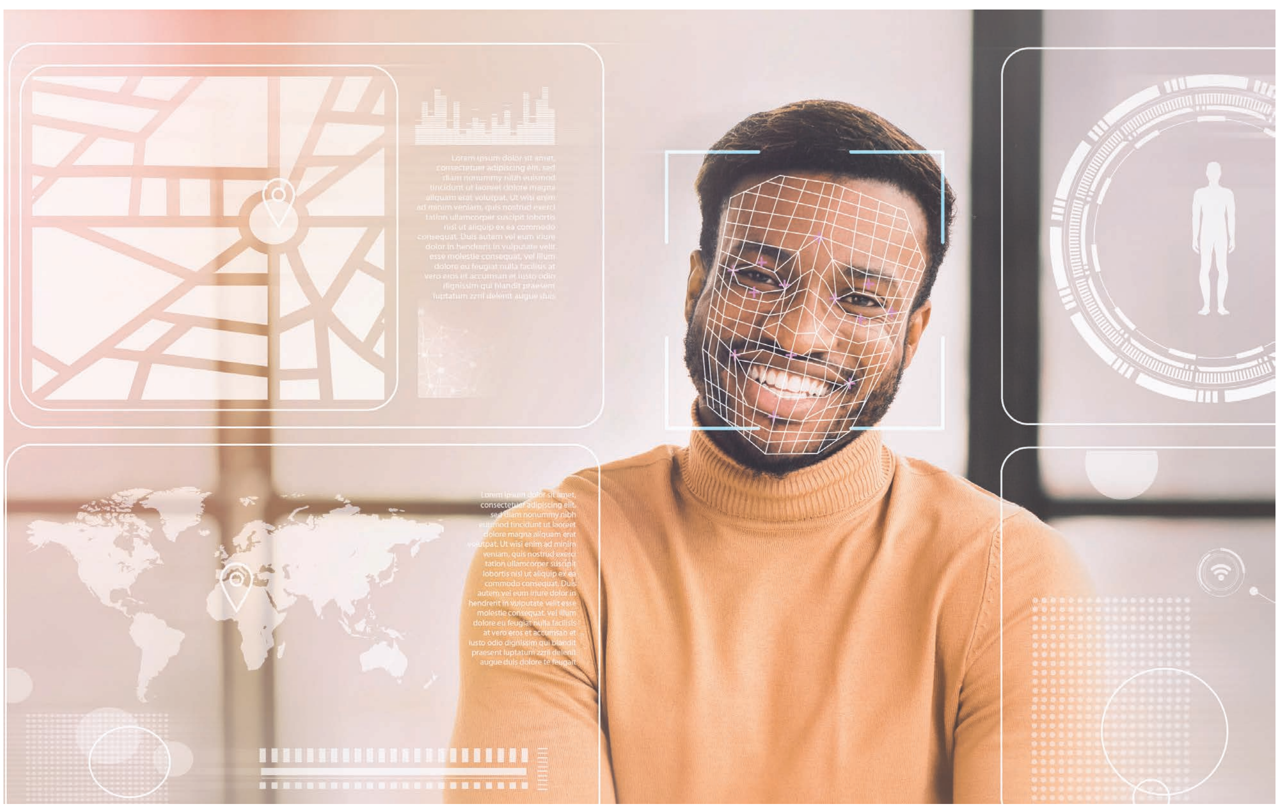
In this section, we discuss the datasets used to construct patent-task matches. The Occupational Information Network, or O*NET, managed by the U.S. Department of Labor, provides detailed information on job duties for different occupations. The second dataset is the U.S. Patent and Trademark Office (USPTO) dataset, hosted on PatentsView, which contains approximately 8 million patents awarded from 1976 to 2021. We also use the AI Patent Dataset (AIPD), which categorizes AI patents into eight different types using machine-learning algorithms validated by patent examiners, and data from the U.S. Census and the American Community Survey (ACS). We collect data on wages and employment for the years 1980, 1990, 2000, 2010, 2019, and 2020 from these sources.

O*NET Dataset

Each occupation requires a specific set of tasks. Each occupation's characteristics are described in O*NET, which includes information on job duties (e.g., tasks), skills, education and training requirements, earnings, and job outlook. The data are regularly updated, and we use the version of the O*NET data from August 2022 for the following analysis. O*NET was the logical choice as a source for task-level information for each occupation because it describes tasks that are core, more frequent, more relevant, and most important in each occupation.

O*NET has updated the work tasks database semi-regularly since 2003 but does not update all occupations at the same time. For instance, in the August 2022 release, some occupations' tasks were last updated in 2004, while others were updated 2022. In total, 50 percent of occupation tasks were last updated in 2018 or earlier.

The use of the August 2022 O*NET database has important implications for our analysis, which focuses on the exposure of current tasks. That is, we do not assess whether patent exposure is associated with changes to the task structure of occupations.



Linking O*NET data over time is challenging. The survey solicits responses from both occupational experts and workers, resulting in irregular updates to each occupation's information and variations in the types of responses received (Lopez Garcia, Maestas, and Mullen, 2020). Additionally, given that new technologies can take decades to affect the labor market (Webb, 2019; Kogan et al., 2020; Meindl, Frank, and Mendonça, 2021), it is not clear whether earlier versions of the O*NET would be better suited for our analysis.

Patent Datasets

We use two separate USPTO datasets for our analysis. The first consists of the full set of all patents (whether or not they are AI-related) awarded by the USPTO from 1976 to 2021. The dataset contains approximately 8 million patents, which are hosted by PatentsView, an open data platform that provides tools and data on patents to foster research and insights into invention and innovation. PatentsView offers patent visualizations, a community forum, an Application Programming Interface, a data query builder, and bulk data downloads. The platform was developed through a collaboration between the USPTO, research organizations, and private companies and is supported by the Office of the Chief Economist at the USPTO.

The other patent dataset used is the AIPD. In Giczy, Pairolero, and Toole (2022), the authors describe this dataset by using a machine learning algorithm to identify and bin USPTO patents into eight AI categories:

- **computer vision—**methods to understand images and videos
- **evolutionary computation—**methods mimicking evolution to solve problems
- **AI hardware—**physical hardware designed specifically to implement AI software
- **knowledge processing—**methods to represent and derive new facts from knowledge bases
- **machine learning—**algorithms that learn from data
- **NLP—**methods to understand and generate human language
- **planning and control—**methods to determine and execute plans to achieve goals
- **speech recognition—**methods to understand speech and generate responses.

The model classification was validated by comparing its results with the manual annotations from patent examiners with AI expertise, and it achieved high performance when compared with existing studies. In the following analysis, we use both the full USPTO database and the subset of AI technologies in the AIPD.

Patent–O*NET Matching Methodology

Our goal is to match the approximately 8 million patents in the USPTO dataset with the almost 18,000 unique tasks in the O*NET database. Because we use text data, we take advantage of NLP algorithms. To implement any NLP application, texts first need to be converted into a numeric representation to run mathematical operations (Goodfellow, Bengio, and Courville, 2016). When words are converted into vectors of numbers, they are called *embeddings*. The embedding vectors are then used in different NLP tasks, such as text classification, sentiment analysis, and question answering. Embeddings have been shown to be very effective for a variety of NLP applications and are one of the most important innovations in NLP in recent years (Mikolov et al., 2014).

There are different ways of generating embeddings. Many embedding models have been used to address the problem of matching patent data with the O*NET dataset (Wang, Zhao, and Jiang, 2020). We have chosen to use embeddings that take advantage of transformers architectures,[4] which in turn takes advantage of the attention concept, developed by Vaswani et al. (2017). *Attention* allows the model to learn long-range dependencies between different parts of the input sequence that other neural networks struggle with (e.g., recurrent neural networks).

More specifically, attention allows a language model to understand the meaning of a word based on the words around it. It works by examining all possible pairs of words in a sentence to determine how each word is related to all the other words, including itself (Rothman, 2021). This helps the model understand the context of words and how they relate to each other. For instance, the word "bank" can have different meanings depending on the context in which it is used. Attention helps the model understand the context by associating each word with other words that regularly appear together. This allows the model to differentiate between different meanings of the same word, such as "financial bank," "bank of a river," and "banking airplane." By understanding the relationships between words, the model can better understand the meaning of a sentence and perform more-accurate NLP tasks.

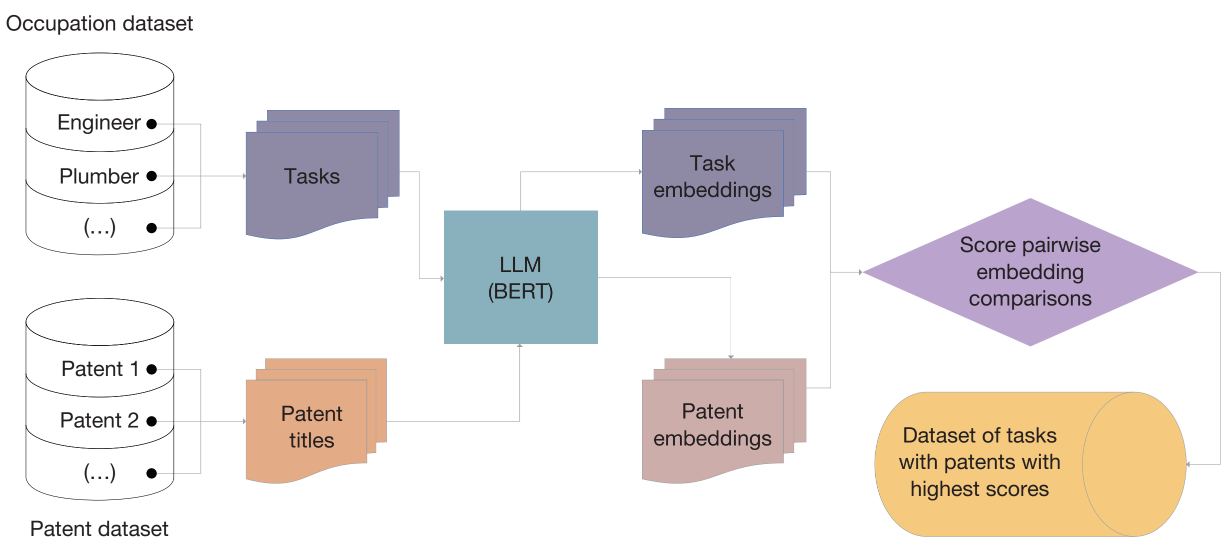
Transformers have had a significant impact on the field of NLP. They have been used to achieve state-of-the-art results for a wide variety of tasks, including machine translation, text summarization, sentence similarity, and question answering, and they are fundamental to the success of LLMs, such as ChatGPT (OpenAI, undated), Bard (Google, undated), and Claude (Anthropic, undated).

To match patents with O*NET tasks, we employ the NLP concept of sentence similarity, which allows us to measure how similar two sentences or phrases are.

Our algorithm, which is illustrated in Figure 1, is composed of the following steps:

1. From the occupation dataset, compile a list of tasks for every occupation and a list of all patent titles from the patent dataset.
2. Select the embedding model:
    a. To generate embeddings, we have chosen the BERT, which is a pretrained language model developed by Google that attends to the whole input sequence instead of using the previous sequence of words to predict the next word. BERT is based on the transformer architecture introduced in Vaswani et al. (2017).
    b. LLMs can be fine-tuned for a specific task or topic, which can lead to significant improvements in performance on the task or topic downstream. With that in mind, we searched for and found a BERT model that was fine-tuned on patents. BERT for Patents is fine-tuned on more than 100 million patents and the multiple texts pertaining to the patent, such as the abstract, claims, and description, which ensures that

FIGURE 1
Patent and Occupations Matching Algorithm



NOTE: Both the patent titles and the occupation tasks are embedded using the same model, and we generate a task embedding list and patents embedding list before we pairwise compare every task and patent. The comparison is done efficiently using a nearest neighbor algorithm.

it will have a better matching performance than the base BERT model.

3. Generate the embedding for both the task and the patent lists: Use the selected model to generate a numerical representation of the patent titles and the O*NET task descriptions.[5]
4. Compare the embeddings using a preselected metric and give the pairwise matches a score: The metric that we have selected is the cosine similarity, which is most used to compare two embeddings.

Cosine Similarity Metric

We use the cosine similarity metric to measure the semantic closeness of a match between task and patent. The process of generating embeddings for the tasks and patents results in multidimensional vectors; therefore, we can perform linear algebra operations on them. Mathematically, cosine similarity is just the process of calculating the cosine between any two vectors, which varies from –1 to 1, where a score of 1 would imply a complete match, a score of –1 would imply a completely opposing meaning, and a score of zero would imply that the embeddings are not related at all. Cosine similarity is a popular measure of similarity in NLP because it is computationally efficient, easy to interpret, and invariant to the length of the vectors.

Datasets Matching Results

We present the results of the matching between patents and occupational tasks in Table 1. Using the methodology described previously, the results of the matching indicate, in most cases, significant alignment between the task and patent. The table presents examples of several selected tasks for particular occupations and the top four patent title matches. (The tasks presented in this table were chosen at random.) It is important to note that the matchings are done on a *contextual basis*, which means that the LLM associates words that frequently appear together. For example, if we look at the first task in Table 1, we see that the first task is related to lifting, slings, and hooks, and therefore the LLM and the algorithm will match technology patents that involve slings, lifting, and hooks irrespective of the true match quality.

Measuring Occupation Exposure

In this section, we explore several dimensions of how exposure to technology patents varies across occupations and over time. Understanding differences in exposure is critical for assessing the potential implications of AI and automation for the labor market. We begin by evaluating how patent-task matches differ by the similarity threshold used, as the chosen threshold affects the number and quality of matches. Next, we examine whether certain types of tasks tend to attract more patent matches. We then aggregate the patent-task matches to the occupation level and analyze exposure differences across occupations, skills, education requirements, and routine task intensity. Evaluating exposure by routine task intensity allows us to connect our analysis to prior research on how routine versus nonroutine tasks have been affected by technology. Tracing out exposure over time provides insights into how exposure has evolved across occupations. Overall, analyzing the different dimensions of exposure is essential for understanding where and how new technologies might affect the workforce.

Differences in Patent-Task Matches by Cosine Similarity Threshold

We begin by assessing which tasks are the most exposed to general technology patents. Specifically, we calculate the total number of patent matches per task from 1976 to 2020 for two cosine similarity thresholds, 0.75 and 0.80.[6] Task-patent matches with cosine similarity values below these thresholds are not counted as a match, while matches above them are. The rank correlation between the total number of patent-task matches between 1976 and 2020 across the two thresholds is 0.83.

Figure 2 shows the distribution of patent-task matches for each threshold. For both thresholds, there are a significant number of tasks with zero patent matches. Some tasks with zero patent matches over the sample time frame include more-abstract tasks, such as adhering to local, state, and federal

TABLE 1

Top Four Patent Matches for a Few Selected Occupational Tasks

Occupation	Task	Title of Top Patent Matches (Cosine Similarity)
Fishing and hunting workers	Attach nets, slings, hooks, blades, or lifting devices to cables, booms, hoists, or dredges.	• Rope and a mooring device, particularly for clamping goods, mooring ships, and anchoring floating landing stages, buoys, navigation marks, etc. (0.78). • Method and device for attaching and removing an additional device to the main boom of a mobile crane (0.77). • System for lifting, moving, and transporting a vehicle via multiple slings connected to a common lifting vertex, and method of retrofitting a vehicle to facilitate lifting (0.76). • Method of retrieving and securing anchors, fish traps, and lobster pots (0.76).
Maintenance and repair workers, general	Assemble, install, or repair wiring, electrical or electronic components, pipe systems, plumbing, machinery, or equipment.	• Method for joining piping systems and piping to equipment, fixtures, devices, structures, and appliances (0.82). • Hole protector device for mechanical, plumbing, and electrical systems (0.77). • Method and machine for installing electrical box, wiring, and receptacle, or switch simultaneously (0.76). • Device, system, and method for the location and identification of as-built plants of pipes, conduits, cables, or hidden objects (0.76).
Gambling and sports book writers and runners	Deliver tickets, cards, and money to bingo callers.	• Method for using a camera phone to acquire, store, manage, and redeem discount coupons (0.77). • Method and apparatus for using greeting cards distributed with electronic commerce transactions as pick tickets (0.77). • System to offer coupons to fans along routes to game (0.76).
Photonics engineers	Design or develop new crystals for photonics applications.	• Design and synthesis of advanced nonlinear optics materials for electro-optic applications (0.88). • Methods for synthesizing semiconductor quality chalcopyrite crystals for nonlinear optical and radiation detection applications, etc. (0.86). • High-quality-factor photonic crystal nanobeam cavity and method of designing and making same (0.85). • Method and structure for stub tunable resonant cavity for photonic crystals (0.85).
Materials scientists	Research methods of processing, forming, and firing materials to develop such products as ceramic dental fillings, unbreakable dinner plates, and telescope lenses.	• Method and compositions for producing lifelike dental porcelain restorations and dental porcelain restorations so produced (0.81). • Method of fabricating high-light-transmission zirconia blanks for milling dental appliances into a natural appearance (0.81). • Method for producing pieces that are in high mechanical demand, especially tools made of low-cost ceramics or polymers (0.81). • Methods for enhancing optical and strength properties in ceramic bodies that have applications in dental restorations (0.81).

SOURCE: Occupations and task descriptions come from the O*NET dataset.

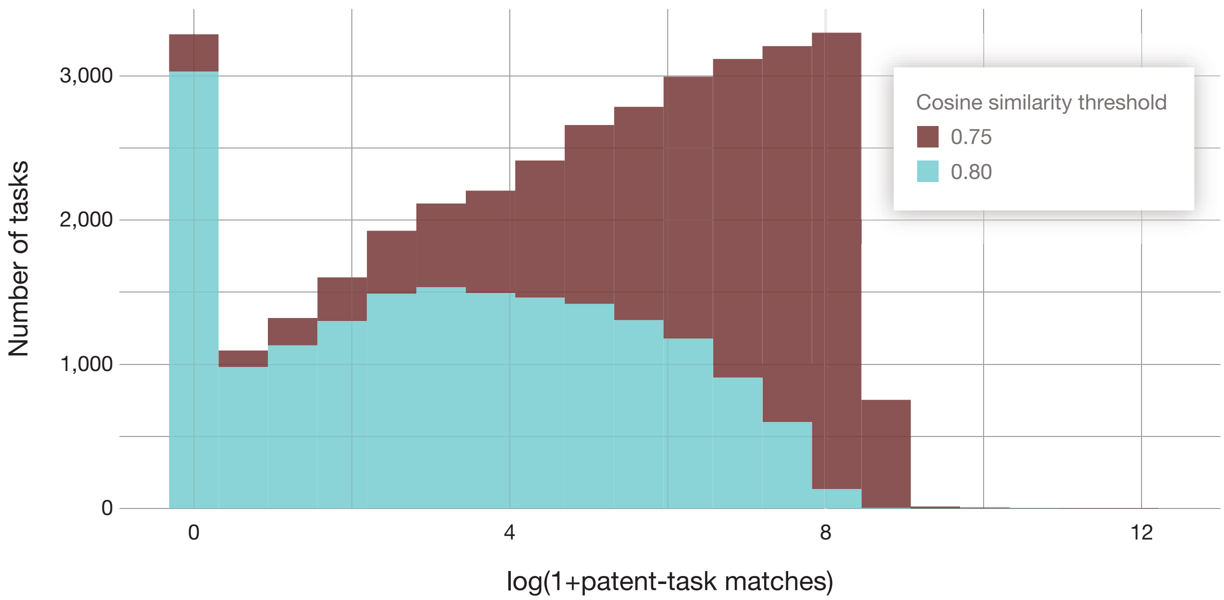
NOTE: Results from O*NET task and all the patent titles were matched using the algorithm developed by the authors, which uses NLP.

laws, and asking speakers to clarify topics. Apart from the mass at zero, the distribution of task matches differs between the two thresholds. Specifically, the 0.75 threshold distribution is truncated at the upper end, while the 0.80 threshold has a more normal distribution. This truncation is because of a 200-match limit per task that we set in the matching algorithm, and, given the lower similarity threshold, this limit is more likely to be met with the 0.75 threshold. Overall, there are many more patent-task matches under the 0.75 threshold. To be conservative, we use the 0.80 similarity threshold in the following analysis even though the results are all qualitatively similar using the 0.75 threshold.

Patent Matches by Task Frequency

To explore which types of tasks are the most exposed to patents, we use data from O*NET on the frequency with which the task is commonly performed. For each task, O*NET reports whether it is typically performed yearly or less frequently, more than yearly, more than monthly, more than weekly, daily, several times per day, or hourly or more. These frequency categories are mutually exclusive. We count the total number of patents per task in each frequency

FIGURE 2
Distribution of Patent-Task Matches by Similarity Threshold



SOURCES: Author calculations using O*NET and USPTO data.
NOTE: We add one and take the log of total patent matches between 1976 and 2020. The *x*-axis is measured in natural log scale, and one is added to the patent task matches because the natural log of zero is undefined.

category (the results are plotted in Figure 3). The figure shows that the distribution of patents per task is skewed toward less frequent tasks. Specifically, tasks that are performed more than yearly and more than monthly have the highest exposure, while those typically performed at higher frequencies have lower exposure. There are over 5,000 tasks that fall into the frequency categories of more than yearly and more than monthly, and these tasks span a wide variety of activities. In total, approximately 30 percent of tasks fall into these two frequency categories.

Aggregation of Patent-Task Matches to Occupation Level

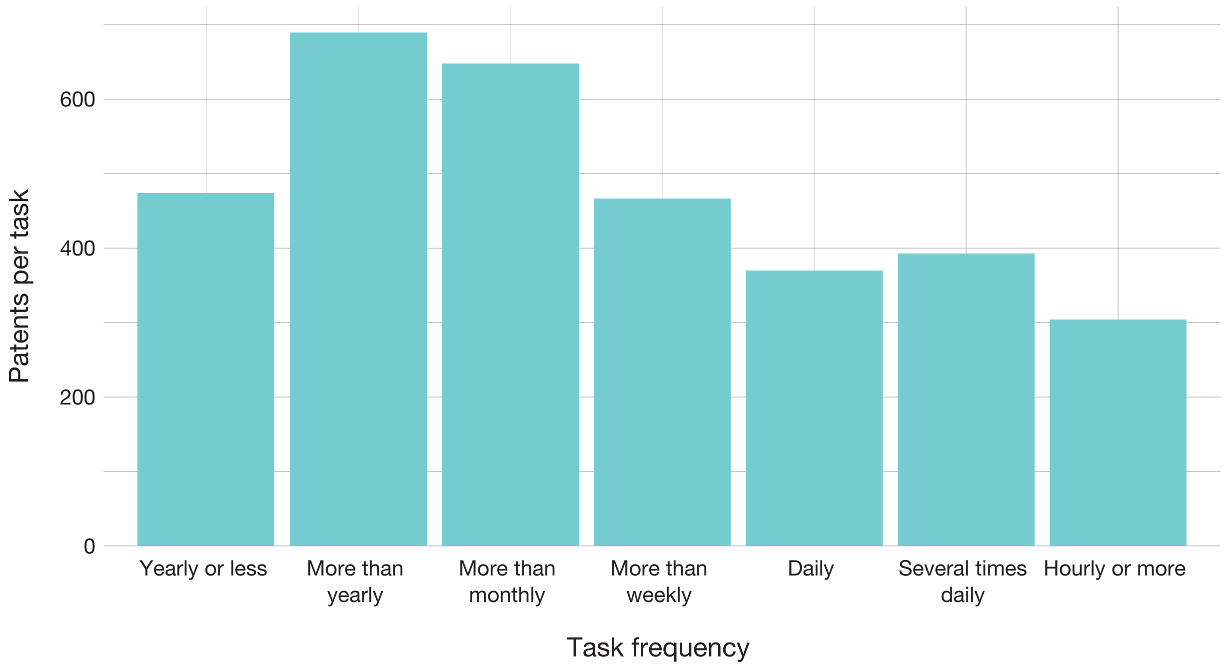
Next, we aggregate the patent-task matches to the occupation level using the task importance measures provided by O*NET. Specifically, for each occupation o in year t, we calculate:

$$exposure_{ot} = \sum_{i \in O} 100 * \left(\frac{patents_{it}}{\sum_i patents_{it}} \right) * importance_{io},$$

where $patents_{it}$ is the number of patent-task matches for task I in year t, and $importance_{io}$ is the importance value assigned to task i for occupation o in the O*NET-26 database.[7] Importance values range from one to five and are based on ratings from workers in the occupation or occupational experts. We scale the number of patent-task matches for task i in year t by the total number of patent-task matches in year t to account for the fact that the number of patents grows over time. We sum this measure over all tasks in occupation o to create an occupation-level measure of exposure. O*NET provides additional measures of task importance, like the relevance of the task to the occupation and the frequency with which the task is performed in the occupation. We test the differences in exposure based on these occupation-task measures and find that they are all highly correlated. For instance, the correlation between exposure using O*NET's measure of task importance and using the measure of task relevance is 97 percent.

FIGURE 3

Patents per Task Frequency



SOURCES: Author calculations using O*NET and USPTO data. Task frequency categories come from the O*NET database.

Exposure to Technology over Time

As of 2022, we find that 87 percent of occupations have some exposure to technology patents in 1976 and that all occupations have some exposure to technology patents by 1989. However, the most-exposed occupations vary over time. Figure 4 displays the top ten most exposed occupations in 1980, 2000, and 2020. In the figure, we measure exposure as the cumulative exposure up to the years listed. For instance, cumulative exposure in 2000 is calculated as:

$$Cumulative\ Exposure_{o,2000} = \sum_{t=1976}^{T=2000} exposure_{ot}.$$

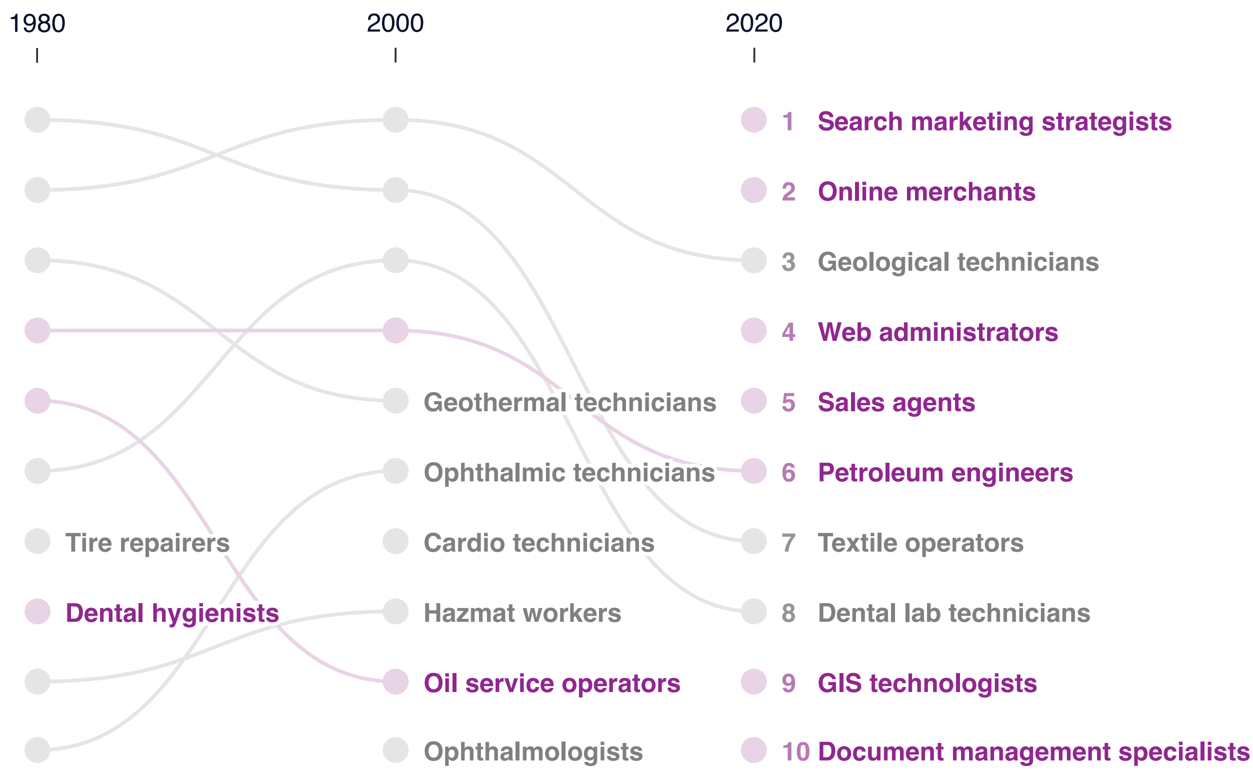
The figure highlights how some highly exposed occupations in the 1980s become relatively less exposed by 2020. For instance, textile operators were the most exposed occupation in 1980, the second most exposed in 2000, and the seventh most exposed in 2020, which indicates that the patents related to the tasks in this occupation did not grow as much as patents related to tasks in other occupations. Of the top ten most exposed occupations in 1980, only three remained in the top ten most exposed by 2020.

Figure 4 also shows which occupations are currently bright-outlook occupations, as classified by the 2022 version of the O*NET (denoted by the violet shading). *Bright-outlook occupations* are projected to grow faster than the average occupation in terms of employment over the next decade or to have at least 100,000 job openings over the next decade. There are relatively few occupations that were highly exposed in 1980 that have a bright outlook as of 2022. However, the majority of the most-exposed jobs in 2020 have a bright outlook.

Skill Exposure

Next, we merge our occupation exposure measure with O*NET's skills data to calculate the change in exposure of different work skills over time. For each of the 35 skills listed in the O*NET, we calculate skill exposure as follows:

FIGURE 4
Top Ten Most-Exposed Occupations in Select Years



SOURCES: Author calculation from O*NET and USPTO data.

NOTE: *Exposure* is measured as the cumulative exposure to all patents up to the given year. Occupation names have been abbreviated from the original names in the O*NET database. Violet coloring indicates that the occupation has a bright outlook according to O*NET.

$$Skill\ Exposure_j = \sum_o \left(Importance_{jo} * \frac{Cumulative\ Exposure_o}{\sum_o Cumulative\ Exposure_o} \right).$$

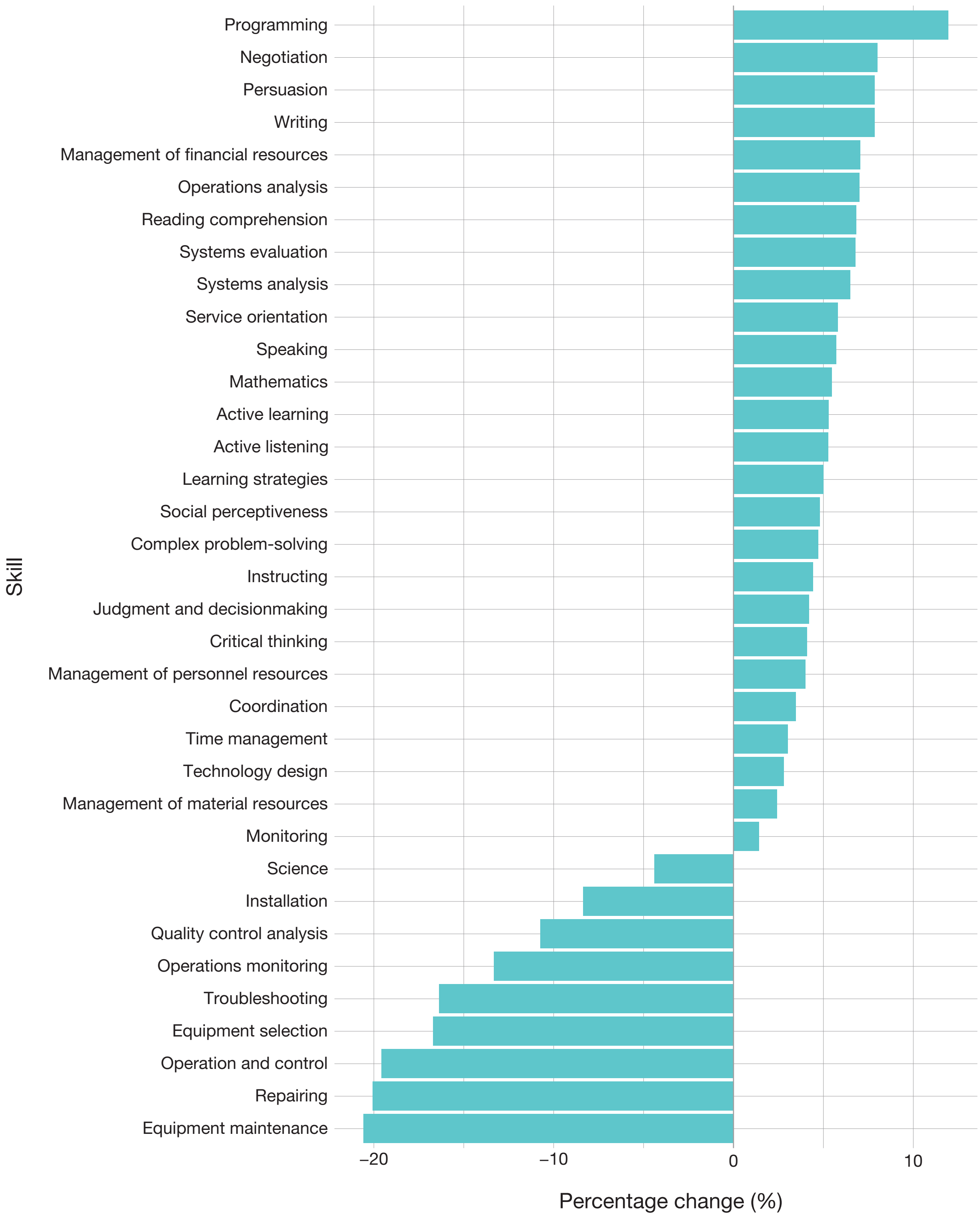
where $Importance_{jo}$ is the importance score for skill j for occupation o, and $cumulative\ exposure_o$ is the cumulative exposure of occupation o at a given time.[8] The resulting measure expresses the weighted average exposure of skill j to technology patents at a given time.

Figure 5 displays the percentage change in skill exposure between 1980 and 2020. The figure highlights that growth in exposure differs across skills. For example, exposure of programming skills increased by over 10 percent. In contrast, the exposure of equipment maintenance skills fell by over 20 percent over this period. In general, the skills with negative exposure growth involve equipment maintenance, operation, repairing, and selection, while those that saw the positive exposure growth involve programming and soft skills, such as negotiation and persuasion. Other research has noted the growing importance of social skills in the labor market (Deming, 2017).

Exposure to Artificial Intelligence Technology

Using the same matching procedure and exposure measures discussed above, we present the top ten most-exposed occupations in 2020 for each AI technology in Table 2. While there is overlap across columns, common themes emerge among the most-exposed occupations. For instance, occupations that

FIGURE 5
Change in Skill Exposure Between 1980 and 2020



SOURCES: Author calculations from O*NET and USPTO data.
NOTE: Skill exposure is measured using the weighted average across occupations that use a given skill in their work tasks.

TABLE 2

Top Ten Occupations Most Exposed to AI Technologies

Computer Vision	Evolutionary Computation	AI Hardware	Knowledge Processing	Machine Learning	NLP	Planning and Control	Speech Recognition
GIS technologists and technicians	Nondestructive testing specialists	Search marketing strategists	Search marketing strategists	Audiologists	Captioners	Search marketing strategists	Captioners
Search marketing strategists	Machinists	Information security analysts	Statisticians	Statisticians	Marketing strategists	Online merchants	Special education teachers, secondary
Captioners	Cytotechnologists	Statisticians	Geological technicians	Critical care nurses	Special education teachers, secondary	Sales agents	Interpreters and translators
Statisticians	Ophthalmic technologists	Document management specialists	Audiologists	Search marketing strategists	Speech and language pathology assistants	Clinical nurse specialists	Speech and language pathology assistants
Special education teachers, secondary	Search marketing strategists	Web administrators	Clinical nurse specialists	Geneticists	Document management specialists	Treasurers and controllers	Search marketing strategists
Radiologic technicians	Statisticians	Data warehousing specialists	Online merchants	Speech and language pathology assistants	Interpreters and translators	Advanced practice psychiatric nurses	Speech and language pathologists
Physicians, pathologists	Geological technicians	Special education teachers, middle school	GIS technologists and technicians	Special education teachers, secondary	English teachers, postsecondary	Bookkeeping clerks	Hearing aid specialists
Document management specialists	Ophthalmic technicians	Telecom engineering specialists	Advanced practice psychiatric nurses	Special education teachers, middle school	GIS technologists and technicians	Web administrators	Music directors and composers
Special effects artists and animators	Astronomers	Special education teachers, secondary	Web administrators	Clinical nurse specialists	Speech and language pathologists	Claims adjusters	English teachers, postsecondary
Speech and language pathologists	Substance abuse counselors	GIS technologists and technicians	Special education teachers, middle school	Captioners	Telecom engineering specialists	Critical care nurses	Statisticians

SOURCES: Author calculations from O*NET and AIPD data.

NOTE: GIS = geographic information systems. Exposure is measured as the cumulative exposure by 2020. Occupation names have been shortened from their original length in the O*NET database. Green cells indicate that the occupation has a bright outlook according to O*NET.

are the most exposed to speech recognition and NLP technologies generally involve communication, writing, and active listening while the occupations most exposed to planning and control generally involve business and finance.

In Table 2, we highlight bright outlook occupations in green. There is variation in the share of bright-outlook occupations across technology categories. For example, nearly all of the occupations most exposed to planning and control and knowledge processing technologies have bright outlooks while only four of the occupations most exposed to evolutionary computation have bright outlooks, according to O*NET data.

Exposure by Required Level of Education

Using O*NET's educational requirements data, which provide information on the typical level of education required per occupation, we calculate the average exposure by educational requirement groups. We group occupations into those that require less than a high school diploma, more than high school completion but less than a college degree, a bachelor's degree, or an advanced degree and then calculate the average exposure in each year for each group.

Figure 6 displays the locally estimated scatterplot smoothing (LOESS) plot of exposure over time for each occupation group and technology. The first subgraph ("All Patents") shows the results for all patents, which includes AI and non-AI patents. We find that in the early years of our sample, occupations that require less than a high school degree were the most exposed and those that require a college degree were the least exposed. However, by 2020, these trends are reversed, and occupations that require a college degree are the most exposed group. The exposure of occupations that require at least a college degree appears to have reached a maximum around 2010, while the exposure of occupations requiring an advanced degree still appears to be on an upward trajectory.

The trends for AI technologies, shown in the other subgraphs of Figure 6, are all relatively similar. Occupations requiring less education are generally less exposed throughout the sample period, while

FIGURE 6

Exposure by Technology and Required Education Level

SOURCES: Author calculations using O*NET, AIPD, and USPTO data.

NOTE: This figure shows the results of an LOESS regression of exposure over time for different technologies and different occupation education requirements. The *y*-axis shows the estimated exposure of an occupation, with higher values indicating greater exposure to technology patents. The LOESS uses a 0.80 bandwidth, as was used in Webb (2019). *Less than HS* refers to occupations that typically require less than a high school diploma; *less than college* refers to occupations that require more than high school completion but less than a bachelor's degree; *college degree* refers to occupations that require a bachelor's degree; and *advanced degree* refers to occupations that require more than a bachelor's degree. The shaded area shows the 95 percent confidence interval.

those requiring a college degree or advanced degree have become more exposed since the late 1990s. In all cases, occupations that require a college degree are the most exposed by 2020. In several instances, occupations that require an advanced degree were the most exposed group in the early part of the sample and have fallen in exposure since. This is particularly clear in knowledge processing and machine learning technologies, but also prevalent in planning and control and speech recognition.

Task Inputs

Next, we explore the relationship between occupation exposure and task inputs. *Task inputs* are a characterization of the types of tasks required to be carried out by workers in the occupation. A relatively large body of research has linked the labor displacing effect of automation in the 1980s and 1990s to the routineness of the tasks in the occupation because routine tasks were historically easier to automate than those that require abstract thinking, creativity, or those that require manual tasks that depend on dexterity and manipulation of objects (see Acemoglu and Autor [2011] for a review of this literature). For instance, Autor, Levy, and Murnane (2003) find that automation technology reduces demand for occupations with routine tasks and increases demand for those with more-abstract, problem-solving, and interpersonal-related tasks.

We use the methods described in Autor and Dorn (2013) to create task input indexes. Using data from the O*NET-26, we update the code provided by Autor and Dorn (2013) to construct task input ratings for nonroutine cognitive, nonroutine manual, routine cognitive, and routine manual tasks based on work activities, skills, work context, and abilities datasets provided by O*NET. We also construct measures of the offshoring potential of the occupation (how "offshorable" it is) from Autor and Dorn (2013). Offshorable occupations are those that require limited face-to-face contact and on-site work and could theoretically be performed in other countries without a significant loss of productivity (Firpo, Fortin, and Lemieux, 2011). We operationalize the offshorable measure by taking the average of the importance of face-to-face contact and on-site job requirements measures derived from the O*NET. See Autor and Dorn (2013) for a complete description of these measures and their calculation.

With this data, we estimate the following regression separately for the years 1980, 1990, 2000, 2010, and 2020 and for each technology group:

$$\begin{aligned}&\mathit{Cumulative\ Exposure}_o = \\ &\quad \beta_0 + \beta_1 \mathit{nonroutine\ cognitive}_o + \\ &\quad \beta_2 \mathit{nonroutine\ manual}_o + \beta_3 \mathit{routine\ cognitive}_o + \\ &\quad \beta_4 \mathit{routine\ manual}_o + \beta_5 \mathit{offshorable}_o + \epsilon_o\end{aligned}$$

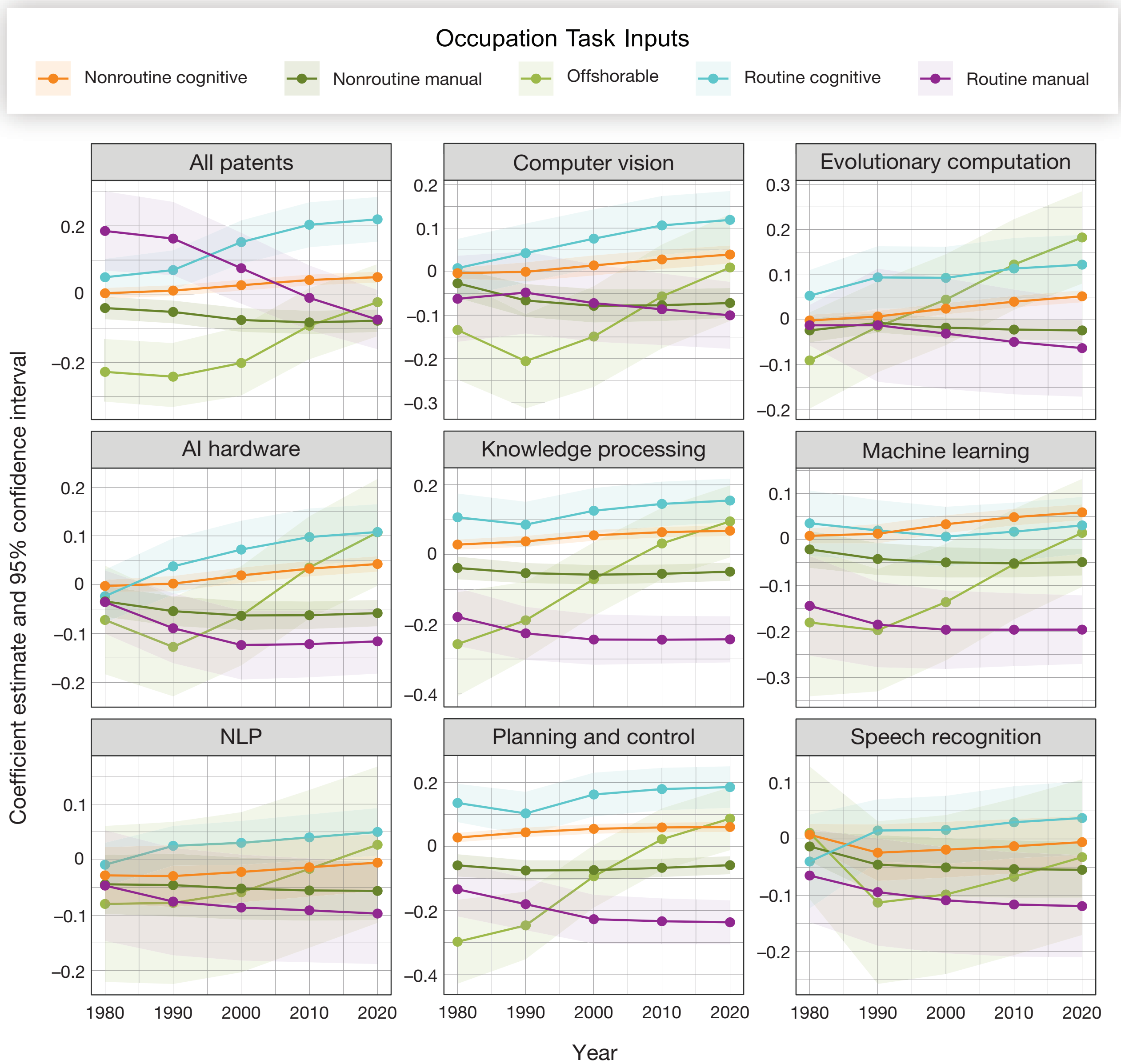
where $\mathit{Cumulative\ Exposure}_o$ is the standardized (z-score) exposure of occupation o, and the task variables measure the task input indexes of occupation o, which are also measured in z-scores. The coefficients can be interpreted as the correlation between the task input indexes and the exposure of the occupation. For instance, β_1 measures the relationship between how nonroutine cognitive occupation o's tasks are and the exposure of occupation o to a given technology, controlling for the occupation's other task inputs. The lack of time subscripts on variables in the equation is because of the separate estimations for each decade.

Figure 7 displays the results. The x-axis displays the year, and the y-axis shows the estimated coefficients and 95 percent confidence intervals. The results, which use all patent data, are shown in the top left subgraph. Here, the results suggest that there is a positive correlation between routine cognitive task inputs and patent exposure, and that the correlation has grown in magnitude since the 1980s. On the other hand, the correlation between exposure and routine manual task inputs has fallen since the 1980s. The growth in the relationship between routine cognitive tasks and exposure aligns with prior research, which has found that occupations with routine cognitive tasks tend to be the easiest to automate with software and robotics (Autor, Levy, and Murnane, 2003; Webb, 2019).

For AI technologies, routine cognitive tasks are the most positively correlated with exposure and have remained the most correlated over time. Examples of occupations with a high degree of routine cognitive tasks include bill and account collectors, telephone operators, proofreaders, and nuclear power reactor operators. Machine learning technologies are an exception, where nonroutine cognitive tasks have become the most positive predictor of exposure since the 2000s. Examples of occupations with a high degree of nonroutine cognitive tasks include education administrators, training and development managers, postsecondary teachers, and emergency management directors. The AI technologies also generally display a pattern of increasing correlation with offshorable tasks. In fact, in some technology categories, offshorable tasks have become one of the strongest predictors of exposure. This suggests that there might be growing application of these technologies to reduce the need for face-to-face interaction and working on-site, thus increasing the correlation with offshorable tasks. Examples of occupations with a high degree of potentially offshorable tasks include

FIGURE 7

Relationship Between Exposure and Task Inputs over Time



SOURCES: Author calculations using ACS, O*NET, Autor and Dorn (2013), AIPD, and USPTO data.

NOTE: This figure shows the results of regressing occupation exposure (measured as cumulative exposure) on occupation task input measure from Autor and Dorn (2013). Exposure is standardized as z-scores. Each point represents the estimates of a separate coefficient, and the shaded region is the 95 percent confidence interval. Regression results are estimated separately for each decade.

creative writers, business intelligence analysts, economists, and biostatisticians.

Percentage of Jobs Exposed to Technology Patents

What percentage of jobs are exposed to technology patents? To answer this question, we define several categories of exposure based on each occupation's cumulative exposure by 2020. For each technology category, we categorize occupations as being *mildly exposed* if the cumulative exposure is between the average and one standard deviation above the average; *highly exposed* occupations have a cumulative exposure of between one and two standard deviations above the average; and *extremely exposed* occupations have a cumulative exposure greater than two standard deviations above the average.[9] Then, we calculate the share of job titles that fall into each exposure category.

Table 3 displays the results. The first column displays the technology type, and the following columns show the share of jobs that are in each category. The last column shows the total share of jobs with some degree of exposure. For example, we find that 38 percent of jobs have some degree of exposure to all patents by 2020. Across technology categories, approximately 25 percent of jobs are mildly exposed to patents, on average. NLP technology has the highest share of mildly exposed occupations across the AI technology categories. Approximately 10 percent or fewer jobs are highly exposed to patents across technology categories, and even fewer are extremely exposed. Five percent of job titles are extremely exposed to all patents and to knowledge processing patents, and even fewer are highly exposed to other technologies. Overall, the results suggest that a significant share of occupations have some exposure to technology patents, though most of the exposure is mild in relative magnitude.

Correlation Between Exposure, Wage, and Employment

In this section, we explore the relationship between an occupation's exposure to technology patents and key labor market outcomes—wages and employment growth. Understanding these correlations is critical for evaluating the potential impacts of such new technologies as AI on workers. We analyze how exposure correlates with an occupation's position in the wage distribution over time. This sheds light on whether low, middle, or high wage earners are most exposed. Next, we estimate regression models to test whether technology exposure correlates with employment

TABLE 3

Share of Job Titles Exposed to Technology in 2020

	Mildly Exposed	Highly Exposed	Extremely Exposed	Total Exposure
All patents	0.25	0.08	0.05	0.38
Computer vision	0.20	0.06	0.04	0.30
Evolutionary computation	0.20	0.06	0.03	0.29
AI hardware	0.21	0.05	0.04	0.31
Knowledge processing	0.22	0.08	0.05	0.36
Machine learning	0.19	0.08	0.04	0.31
NLP	0.25	0.01	0.02	0.28
Planning and control	0.22	0.10	0.04	0.36
Speech recognition	0.19	0.02	0.02	0.22

SOURCES: Author calculations using O*NET, AIPD, and USPTO data.

NOTE: Exposure is measured using the cumulative exposure to patents between 1976 and 2020. *Mildly exposed* are occupations with an exposure between the average and one standard deviation above the average. *Highly exposed* are occupations with an exposure between one and two standard deviations above the average. *Extremely exposed* are occupations with an exposure greater than two standard deviations above the average. *Total exposure* is the sum of mildly, highly, and extremely exposed.

growth and whether this depends on the routine task intensity of an occupation. Connecting exposure to employment growth builds on prior research showing that routine tasks are more susceptible to automation. Overall, analyzing the correlation between exposure and labor market outcomes provides indicative evidence regarding how AI and automation might affect different types of workers and occupations. However, it is important to note these relationships are correlational and not necessarily causal.

Using data from the decennial census and the ACS, we estimate the relationship between exposure and wages. We use decennial census data from 1980 and 1990, and the 2000, 2019, and 2020 ACS survey waves. We collect these data from the Integrated Public Use Microdata Series (IPUMS) (Ruggles et al., 2022). Following Webb (2019) and Acemoglu and Autor (2011), we restrict the sample to workers between 18 and 65 years old, and we calculate a labor-supply weight by multiplying the IPUMS survey person weight by the fraction of full-time work for each observation. We aggregate the data to the occupation level, meaning that our final measure is the number of full-time-equivalent employees per occupation. We also use the labor-supply weight to calculate the weighted average wage (in 2010 U.S. dollars) for each occupation.[10]

Wages

Figure 8 displays the relationship between cumulative exposure and wages in different time periods. In each subgraph, the *x*-axis is the occupations wage percentile, and the *y*-axis is the standardized (z-score) cumulative exposure as of the specific year. The figures show the LOESS plot fitted values and 95 percent confidence intervals for 1980, 2000, and 2020. The first subgraph shows the relationship for all patents. In 1980, occupations just below the median wage were the most exposed, and those in lower and upper wage percentiles were relatively less exposed. However, in 2000, exposure of higher wage occupations began growing; by 2020, the highest wage occupations were the most exposed. A similar pattern emerges across AI technologies. By 2020, there is a positive correlation between an occupations wage percentile and its exposure.

Employment

Next, we calculate the share of full-time equivalent employees that are highly exposed to technology. Here, we define a worker as *highly exposed* if the cumulative exposure of their occupation is at least one standard deviation above the average exposure for that technology and year. We then sum the total number of employees in highly exposed occupations and calculate the share of total employment that these exposed employees make up each year.

Table 4 displays the results for 2019. We use the year 2019 to avoid any confounding issues in employment because of the COVID-19 pandemic, though the results are similar if 2020 is used as a comparison year instead. Technologies are shown in the first column and are ordered by the share of workers exposed in 2019. There is a wide range in exposure across technology categories. For instance, 15 percent of workers are highly exposed to planning and control technology patents in 2019, while only 2 percent of workers are highly exposed to evolutionary computation technology patents.

Correlation Between Exposure and Employment Growth

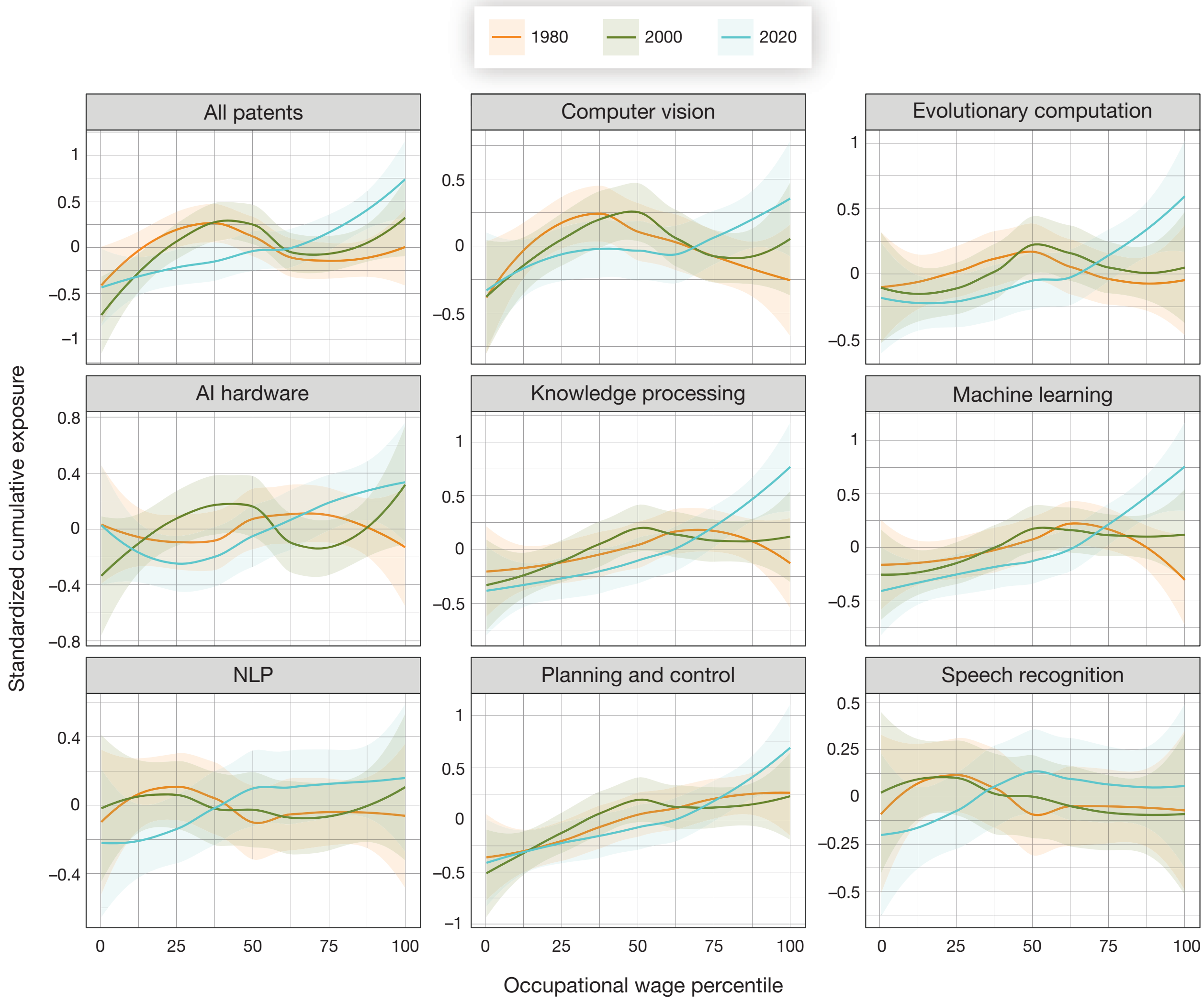
As a final exercise, we estimate a series of regressions aimed at understanding the relationship between exposure and employment growth. Specifically, for each technology category, we estimate the following regression:

$$\begin{aligned}\Delta y_o^{1990-2019} = {} & \beta_1 + \beta_2 exposure_o^{2019} + \beta_3 RI_o + \\ & \beta_4 (exposure_o^{2019} * RI_o) + \epsilon_o,\end{aligned}$$

where $\Delta y_o^{1990-2019}$ is the percentage change in employment in occupation o between 1990 and 2019 and $exposure_o^{2019}$ is the cumulative exposure of occupation *o* in 2019, in z-scores. We also include the term RI_o from Autor and Dorn (2013), which is the routine-intensity index score of occupation *o*. We use 1990 as a base year because several occupations enter the census data between 1980 and 1990, so there are more occupations for which we have employment data in later census years. However, the results are generally consistent if 1980 is used as a base year instead. We calculate the routine intensity as the sum

FIGURE 8

Relationship Between Exposure and Wages



SOURCES: Author calculations using ACS, O*NET, AIPD, and USPTO data.

NOTE: This figure shows the LOESS regression of cumulative exposure on wage percentiles. Cumulative exposure is calculated in 1980, 2000, and 2020 and standardized in z-scores. Occupation wage percentiles are calculated using the employment-weighted mean hourly wage by occupation in the ACS. Wages are in 2010 U.S. dollars.

of the routine cognitive plus routine manual task inputs, minus the sum of nonroutine cognitive and manual task inputs.[11] This routine-intensity variable is also measured in z-scores. Lastly, we include an interaction term $exposure_o^{2019} * RI_o$, which captures the additional response of employment growth to exposure at different levels of routine intensity. We allow the errors to cluster at the occupation level.

Table 5 displays the results from estimating the regression using the change in employment as a dependent variable. Each column represents the results from a separate regression that is estimated using exposure data for a specific technology category.[12]

We find that, in general, increased exposure to technology has a mixed relationship with employment growth over this period, which is consistent with technology sometimes complementing human work and sometimes substituting for it. For example, increased exposure to evolutionary computation, NLP, and speech recognition technology are all

TABLE 4

Share of Full-Time-Equivalent Employees Exposed to Technology in 2019

	Share of Workers Highly Exposed in 2019
All patents	0.15
Computer vision	0.09
Evolutionary computation	0.02
AI hardware	0.07
Knowledge processing	0.14
Machine learning	0.10
NLP	0.04
Planning and control	0.15
Speech recognition	0.05

SOURCES: Author calculations using O*NET, AIPD, USPTO, and ACS data.

NOTE: This table shows the share of full-time equivalent workers between the ages 18 and 65 exposed to technology in 2019. Exposure is measured as the cumulative exposure each year, and we characterize an occupation as *highly exposed* if the exposure is greater than one standard deviation above the average for that technology and year.

positively correlated with employment growth. We do not find a statistically significant relationship for other technologies.

We also find that more routine-intensive occupations saw employment declines over this period. In general, a one standard deviation increase in the routine intensivewness of an occupation is associated with approximately a 25 percent decline in employment growth, holding exposure constant. For context, a one standard deviation increase in the routine intensiveness of an occupation is approximately equivalent to going from the routineness of an industrial machine repairer to that of a cashier.

Finally, the estimate of the coefficient on the interaction term suggests that, in some instances, the correlation between technology exposure and employment growth depends on the routine intensiveness of the occupation. Specifically, exposure to NLP, speech recognition, computer vision, and evolutionary computation technology patents are negatively correlated with the change in employment growth for more routine-intensive occupations. For example, for an occupation that is two standard deviations more routine-intensive than the average occupation, an increase in NLP exposure is associated with a net decline in employment growth of 16 percent.

Conclusion

This project aims to identify occupations exposed to AI technologies. To achieve this, we use an NLP text similarity algorithm between patent titles and job task descriptions. The algorithm matching is based on the contextual basis and uses the BERT LLM that is fine-tuned to patent data. The matching between patents and occupational tasks worked well for most occupation tasks within our predefined cosine similarity threshold of 0.75 and 0.80. The matching was done for both the full patents granted from 1976 to 2021 and an AI patent dataset, which categorizes patents into eight different AI topics.

We evaluate the exposure to multiple categories of AI technologies over the past several decades and estimate how exposure correlates with employment growth. We find that many occupations are exposed to AI but that exposure varies over time, across occupation groups, and between technology categories. Overall, we find that occupations that require more education, pay higher wages, and involve more-routine tasks tend to be more exposed to AI technologies. We also estimate that technology exposure is positively correlated with employment growth in some instances, but that this correlation is negative for more-routine occupations.

The result of our regression analysis that looks at the relationship between exposure and employment growth suggests that AI might play a nuanced role in the labor market as firms continue to adopt new

TABLE 5

Results from Regressing Employment Growth on Exposure and Routine Intensity

	All Patents	Computer Vision	Evolutionary Computation	AI Hardware	Knowledge Processing	Machine Learning	NLP	Planning and Control	Speech Recognition
$exposure_o^{2019}$	0.05	−0.02	0.09**	0.02	0.07	0.09	0.20***	0.07	0.17***
	(0.05)	(0.07)	(0.04)	(0.10)	(0.06)	(0.12)	(0.08)	(0.04)	(0.04)
RI_o	−0.27***	−0.27***	−0.28***	−0.27***	−0.25***	−0.24***	−0.25***	−0.26***	−0.25***
	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
$exposure_o^{2019}*RI_o$	−0.05	−0.12*	−0.24*	−0.02	−0.05	−0.04	−0.18**	−0.05	−0.14***
	(0.04)	(0.07)	(0.13)	(0.07)	(0.05)	(0.08)	(0.08)	(0.04)	(0.04)
Observations	261	261	261	261	261	261	261	261	261
Adjusted R^2	0.11	0.12	0.13	0.11	0.12	0.13	0.14	0.12	0.15

SOURCES: Author calculations using O*NET, USPTO, AIPD, ACS, and Autor and Dorn (2013) data.

NOTE: *Patent exposure* is measured as the cumulative exposure by technology type in 2019 and *routine intensity* is measured as the relative importance of routine tasks to nonroutine tasks in an occupation. The constant term was estimated but not shown in the table. Errors in the table allow for clustering at the occupation level.

*Significant at the 10 percent level, **significant at the 5 percent level, ***significant at the 1 percent level.

AI tools, and its effect on workers might depend on the types of skills and tasks the worker typically performs. For instance, we find that greater exposure to some AI technologies is associated with employment growth for relatively less routine-intensive occupations. At the same time, we find that greater AI exposure is associated with slower employment growth—and even employment declines—in occupations that tend to be relatively more routine-intensive. However, the nature of the relationship between AI exposure and employment might change in the future as new tools are developed that are better able to perform nonroutine tasks.

There are several limitations of our analysis. First, although exposure to technology patents can offer insights into the potential impact of AI on different occupations, it is not a perfect measure of the actual risk of displacement or job loss. For instance, some jobs that are highly exposed to technology patents might require social or emotional intelligence that AI is currently unable to replicate, making those jobs less susceptible to automation. Moreover, exposure to AI technology does not necessarily translate into a risk of displacement. In fact, if AI technologies reduce the cost of acquiring workers with expertise, it might enable employers to expand and create new job opportunities and result in an increase in labor demand. Therefore, although exposure to technology patents is a useful starting point for understanding the potential impact of AI on the labor market, a variety of factors can influence actual labor market outcomes.

Additionally, we do not analyze how the task content within occupations evolves over time in response to technology. Occupations might become less routine as technology automates away some tasks and frees workers to focus on more abstract tasks that cannot be automated. Similarly, new tasks might be added to occupations in response to technology. Additional research is needed to understand how AI might shift the composition of work within occupations. As a result, our analysis does not imply that there is a direct, causal relationship between patent exposure and employment growth, shifts in educational requirements, or changes in the wage distribution. Instead, our results should be considered correlative because there are many dimensions over which occupations, tasks, and workers might respond to technology exposure. Future research should address these margins of adjustment and, in particular, assess the extent to which the labor market response to technology is driven by changes *across* occupations or changes *within* occupations.

Lastly, the patent-task matching algorithm could be improved to produce more-accurate and more-relevant matches. This could involve bringing in

Abbreviations

ACS	American Community Survey
AI	artificial intelligence
AIPD	Artificial Intelligence Patent Dataset
BERT	Bidirectional Encoder Representations from Transformers
GPT-4	Generative Pre-trained Transformer 4
LLM	large language model
LOESS	locally estimated scatterplot smoothing
NLP	natural language processing
O*NET	Occupational Information Network
USPTO	U.S. Patent and Trademark Office

additional data on job tasks, such as data acquired from the Occupational Requirements Survey from the Bureau of Labor Statistics, to supplement tasks in the O*NET. Similarly, supplementing task data with other O*NET datasets on work activities, work styles, and work skills could provide more information for the algorithm to draw from. Finally, additional subsets of the patent data, such as those for robots or LLMs, could provide a more detailed picture of technology exposure.

Notes

1 NLP is a field of AI that focuses on enabling computers to understand, interpret, and generate human language in similar ways to humans.

2 These AI technologies are defined in the section on datasets that follows.

3 GPT-4 is the fourth installment of OpenAI's LLM. According to OpenAI, it can take both texts and images as input, process them, and achieve human-level performance in many professional and academic settings.

LLMs are a subset of what are called *foundational models*, which are millions and billions of parameter models that are broadly trained on large datasets and can be fine-tuned to perform many downstream tasks (Bommasani et al., 2022). LLMs are trained on large sets of online texts and books from which the model learns patterns in a written language. Then, the models are optimized for certain tasks, such as text classification, summarization, or generation.

4 *Transformers* are a type of neural network architecture designed for sequence modeling that is made up of a stack of self-attention layers (Tunstall, von Werra, and Wolf, 2022). Each token in a transformer sequence is similar to a word in a book. The attention mechanism allows each token to "look back" at the other tokens in the sequence to see whether there is any context that could help the token in question understand its meaning. It is no different than when someone is reading a book and they come across a word they do not quite understand: They might look at the previous words or sentence to understand the context that would help them understand what is being talked about. In this way, attention allows transformers to learn long-range dependencies in text.

5 We also explored using patent titles and abstracts for the matching process. In theory, more input text could allow for better patent-task matches. However, we did not find that the matches with the patent titles and abstracts were meaningfully different from those generated with just the patent titles. Additionally, the empirical analysis that follows did not differ in a meaningful way when using the matches with titles and abstracts.

6 In assessing the quality of matches, we used the cosine similarity as a determining factor. After observation and analysis, a value of 0.75 was selected, resulting in a combination of both good and subpar matches. It is important to note that the classification of a "good" match was based solely on our own judgment. We also assessed match quality with a cosine similarity of 0.80. This threshold resulted in better-quality matches but fewer overall matches. It should be noted, however, that the results of the analysis remained similar regardless of whether a matching threshold of 0.75 or 0.80 was utilized.

7 O*NET importance scores are all rated on a 1 (low importance) to 5 (high importance) scale. When surveyed, respondents are asked to rate the importance of different tasks based on a common Likert scale.

8 O*NET provides additional data on the importance of different skills to each occupation. Similar to task ratings, skill importance ratings are measured on a 1-to-5 Likert scale.

[9] We calculate the exposure category separately for each technology group because raw exposure scores are not directly comparable across technology groups. Specifically, our classification is based on the following expression in which we calculate exposure categories separately for each technology group z:

$$exposure\ category_z = \begin{cases} \dfrac{cumulative_{i,z} - \overline{cumulative_z}}{\sigma_z} > 2, & \text{Extremely Exposed} \\ \dfrac{cumulative_{i,z} - \overline{cumulative_z}}{\sigma_z} \in (1,2), & \text{Highly Exposed} \\ \dfrac{cumulative_{i,z} - \overline{cumulative_z}}{\sigma_z} \in (0,1), & \text{Mildly Exposed} \end{cases}$$

[10] We use the occ1990dd occupation classification, which was developed by Dorn (2009), for consistent occupation titles across survey waves; we use the O*NET to occ1990dd crosswalk, which was provided by Webb (2019).

Because the crosswalk between O*NET occupations and census occupations can be many-to-one, we calculate exposure at the census occupation level as the average across O*NET occupations in situations in which multiple O*NET occupations map to the same census occupation.

[11] Specifically, the routine-intensity index is equal to

$$RI_o = \left[\left(nonroutine\ cognitive_o + nonroutine\ manual_o\right) - \left(routine\ cognitive_o + routine\ manual_o\right) \right].$$

[12] We have data on employment and routine intensity for 261 occupations in 1980 and 2019. In total, we have data on employment for 275 occupations in both periods, but we do not have data on the routine-intensity measure for 14 occupations.

References

Acemoglu, Daron, Gary W. Anderson, David N. Beede, Cathy Buffington, Eric E. Childress, Emin Dinlersoz, Lucia S. Foster, Nathan Goldschlag, John C. Haltiwanger, Zachary Kroff, Pascual Restrepo, and Nikolas Zolas, "Automation and the Workforce: A Firm-Level View from the 2019 Annual Business Survey," National Bureau of Economic Research, working paper 30659, 2022.

Acemoglu, Daron, and David Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings," *Handbook of Labor Economics*, Vol. 4, Part B, 2011.

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo, "Artificial Intelligence and Jobs: Evidence from Online Vacancies," *Journal of Labor Economics*, Vol. 40, No. S1, 2022.

Anthropic, "Claude," webpage, undated. As of July 3, 2023: https://claude.ai/

Autor, David H., "Why Are There Still So Many Jobs? The History and Future of Workplace Automation," *Journal of Economic Perspectives*, Vol. 29, No. 3, 2015.

Autor, David H., and David Dorn, "The Growth of Low-Skill Service Jobs and the Polarization of the U.S. Labor Market," *American Economic Review*, Vol. 103, No. 5, August 2013.

Autor, David H., Frank Levy, and Richard J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics*, Vol 118, No. 4, 2003.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al., "On the Opportunities and Risks of Foundation Models," arXiv, 2108.07258v3, July 12, 2022.

Brynjolfsson, Erik, and Andrew McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, W. W. Norton & Company, 2014.

Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock, "What Can Machines Learn, and What Does It Mean for Occupations and the Economy?" *American Economic Association Papers and Proceedings*, Vol. 108, May 2018.

Challenger, Gray, and Christmas, Inc., "May 2023 Layoffs Jump in May on Tech, Retail, Auto; YTD Hiring Lowest Since 2016," blog post, June 1, 2023. As of July 3, 2023: https://www.challengergray.com/blog/may-2023-layoffs-jump-on-tech-retail-auto-ytd-hiring-lowest-since-2016/

Chui, Michael, Eric Hazan, Roger Roberts, Alex Singla, Kate Smaje, Alex Sukharevsky, Lareina Yee, and Rodney Zemmel, "The Economic Potential of Generative AI: The Next Productivity Frontier," McKinsey Digital, June 14, 2023.

Deming, David J. "The Growing Importance of Social Skills in the Labor Market," *Quarterly Journal of Economics*, Vol. 132, No. 4, 2017.

Dorn, David, *Essays on Inequality, Spatial Interaction, and the Demand for Skills*, dissertation, University of St. Gallen, September 2009.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models," arXiv, 2303.10130, August 21, 2023.

FactSet, "Document Search Online Data Tool," webpage, undated. As of July 3, 2023: https://my.apps.factset.com/document-search/all-documents/

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux, "Occupational Tasks and Changes in the Wage Structure," Institute for the Study of Labor, IZA Discussion Paper No. 5524, February 2011.

Frey, Carl Benedikt, and Michael A. Osborne, "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change*, Vol. 114, January 2017.

Giczy, Alexander V., Nicholas A. Pairolero, and Andrew A. Toole, "Identifying Artificial Intelligence (AI) Invention: A Novel AI Patent Dataset," *Journal of Technology Transfer*, Vol. 47, No. 2, 2022.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, 2016.

Google, "Bard," webpage, undated. As of July 3, 2023: https://bard.google.com/

Kogan, Leonid, Dimitris Papanikolaou, Lawrence D. W. Schmidt, and Bryan Seegmiller, "Technology, Vintage-Specific Human Capital, and Labor Displacement: Evidence from Linking Patents with Occupations," National Bureau of Economic Research, working paper 29552, 2021.

Lopez Garcia, Italo, Nicole Maestas, and Kathleen J. Mullen, "The Changing Nature of Work," University of Michigan Retirement and Disability Research Center, working paper 2020-415, November 2020.

Manyika, James, Michael Chui, Mehdi Miremadi, Jacques Bughin, Katy George, Paul Willmott, and Martin Dewhurst, "A Future That Works: AI, Automation, Employment, and Productivity," McKinsey Global Institute Research, January 12, 2017.

Meindl, Benjamin, Morgan R. Frank, and Joana Mendonça, "Exposure of Occupations to Technologies of the Fourth Industrial Revolution," arXiv, 2110.13317, October 25, 2021.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient Estimation of Word Representations in Vector Space," *arXiv:1301.3781*, September 17, 2014.

Montobbio, Fabio, Jacopo Staccioli, Maria Enrica Virgillito, and Marco Vivarelli, "Labour-Saving Automation and Occupational Exposure: A Text-Similarity Measure," Laboratory of Economics and Management, working paper 2021/43, 2021.

OpenAI, "ChatGPT," webpage, undated. As of July 3, 2023: https://chat.openai.com/

Rothman, Denis, *Transformers for Natural Language Processing: Build Innovative Deep Neural Network Architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and More*, Packt Publishing, 2021.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Megan Schouweiler, and Matthew Sobek, "IPUMS USA: Version 12.0," IPUMS, 2022. As of July 3, 2023: https://www.ipums.org/projects/ipums-usa/d010.v12.0

Tunstall, Lewis, Leandro von Werra, and Thomas Wolf, *Natural Language Processing with Transformers: Building Language Applications with Hugging Face*, O'Reilly, 2022.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention Is All You Need: Advances in Neural Information Processing Systems," arXiv, 2706.03762, June 12, 2017.

Wang, Shirui, Wenan Zhou, and Chao Jiang, "A Survey of Word Embeddings Based on Deep Learning," *Computing*, Vol. 102, No. 3, 2020.

Webb, Michael, *The Impact of Artificial Intelligence on the Labor Market*, Social Science Research Network, SSRN 3482150, 2019.

Acknowledgments

We would like to express our sincere gratitude to the RAND Corporation for funding this research as part of the 2022 RAND Initiated Research projects. We would also like to thank Jeffrey B. Wenger and Melanie A. Zaber for initiating this research stream as part of RAND Lowy Family Middle-Class Pathways Center. Additionally, we would like to thank Bryce Downing, Fernando Esteves, and Peter Schirmer for their contributions and feedback. Finally, we would like to thank Lisa Abraham and Prateek Puri for reviewing this report and for their thoughtful insights and suggestions.

About the Authors

Tobias Sytsma is an associate economist at RAND. His research focuses on international trade and supply chain frictions, natural disasters and climate change, economic development, and workforce development. He is a core researcher in the RAND Lowy Family Middle-Class Pathways Center, where he studies emerging technology, automation, and remote work. Sytsma has a Ph.D. in economics.

Éder M. Sousa is an aerospace and data engineer at RAND. He researches the acquisition of defense systems and focuses on propulsion and hypersonic systems, space science, technology and policy, machine learning and data science, and nuclear fusion energy. Sousa also researches the impact of automation and artificial intelligence on the labor force within RAND's Lowy Family Middle-Class Pathways Center. Sousa has a Ph.D. in aeronautics and astronautics.

About This Report

The rapid development of artificial intelligence (AI) has the potential to revolutionize the labor force with new generative AI tools that are projected to contribute trillions of dollars to the global economy by 2040. However, this opportunity comes with concerns about the impact of AI on workers and labor markets. As AI technology continues to evolve, there is a growing need for research to understand its implications for workers, firms, and markets. In this report, we aim to address this pressing need by exploring the relationship between occupational exposure and AI-related technologies, wages, and employment.

Using natural language processing (NLP) to identify semantic similarities between job task descriptions and U.S. technology patents awarded between 1976 and 2020, we evaluate occupation exposure to all technology patents in the United States and to specific AI technologies, including machine learning, NLP, speech recognition, planning and control, AI hardware, computer vision, and evolutionary computation.

Our findings suggest that exposure to both general technology and AI technology patents is not uniform across occupational groups, over time, or across technology categories. We estimate that up to 15 percent of U.S. workers were highly exposed to AI technology patents by 2019 and find that the correlation between technology exposure and employment growth can depend on the routineness of the occupation. This report contributes to the growing literature on the labor market implications of AI and provides insights that can inform policy discussions around this emerging issue.

RAND Education and Labor

This study was undertaken by RAND Education and Labor, a division of the RAND Corporation that conducts research on early childhood through postsecondary education programs, workforce development, and programs and policies affecting workers, entrepreneurship, and financial literacy and decisionmaking.

More information about RAND can be found at www.rand.org. Questions about this report should be directed to tsytsma@rand.org and esousa@rand.org, and questions about RAND Education and Labor should be directed to educationandlabor@rand.org.

Funding

Funding for this research was provided by gifts from RAND supporters and income from operations.

Image credits: cover: Feodora/Adobe Stock; p. 2: S/Adobe Stock; p. 5: Pinkypills/Getty Images; pp. 6, 16: Prostock-Studio/Getty Images.

For more information on this publication, visit www.rand.org/t/RRA2655-1.

www.rand.org

RR-A2655-1