Brookings Papers
ON ECONOMIC ACTIVITY

BPEA Conference Draft, September 25-26, 2025

Technology and Labor Markets: Past, Present, and Future; Evidence from Two Centuries of Innovation

Huben Liu (Northwestern Kellogg School of Management)

Dimitris Papanikolaou (Northwestern Kellogg School of Management)

Lawrence D. W. Schmidt (MIT Sloan School of Management)

Bryan Seegmiller (Northwestern Kellogg School of Management)

The Brookings Institution is committed to quality, independence, and impact. We are supported by a diverse array of funders. In line with our values and policies, each Brookings publication represents the sole views of its author(s).

Technology and Labor Markets: Past, Present, and Future; Evidence from Two Centuries of Innovation*

Huben Liu
Kellogg School of Management

Dimitris Papanikolaou
Kellogg School of Management and NBER

Lawrence D. W. Schmidt
MIT Sloan School of Management

Bryan Seegmiller
Kellogg School of Management

September 11, 2025

Abstract

We use recent advances in natural language processing and large language models to construct novel measures of technology exposure for workers that span almost two centuries. Combining our measures with Census data on occupation employment, we show that technological progress over the 20th century has led to economically meaningful shifts in labor demand across occupations: it has consistently increased demand for occupations with higher education requirements, occupations that pay higher wages, and occupations with a greater fraction of female workers. Using these insights and a calibrated model, we then explore different scenarios for how advances in artificial intelligence (AI) are likely to impact employment trends in the medium run. The model predicts a reversal of past trends, with AI favoring occupations that are lower-educated, lower-paid, and more male-dominated.

*We are grateful to Matthew Akuzawa and Weizhe Sun for excellent research assistance. We are grateful to William Kerr, Christina Patterson and Jón Steinsson for helpful comments and suggestions. Dimitris Papanikolaou and Bryan Seegmiller thank the The Financial Institutions and Markets Research Center and the Asset Management Practicum for their generous financial support.

Economists and workers alike have long worried about the prospect of technological displacement of labor.[1] Recent advances in artificial intelligence have re-ignited the perennial concern that technology will automate away most tasks performed by workers, leading to large declines in labor demand, depressed wages, and diminished job opportunities for workers. Yet systematic evidence on how technological advances shape labor demand over the long run remains limited. The difficulty is that even when technologies are labor-saving, their net impact is ambiguous: they directly substitute for labor in some tasks, but may also raise overall productivity, induce reallocation toward complementary tasks, and increase aggregate labor demand.

Our goal is twofold. First, we aim to tease apart these forces and understand the role that technological progress has played in shaping the demand for labor over the long run. Overall, we find that technological innovation has led to economically meaningful shifts in labor demand across occupations: it has consistently increased demand for occupations with higher education requirements, occupations that pay higher wages, and occupations with a greater fraction of female workers. Second, we explore the extent to which the experience over the last two centuries is informative about the role that advances in artificial intelligence (AI) are likely to play in shaping the composition of the labor force in the medium run. In sharp contrast to the past two centuries, our framework suggests that AI—by substituting primarily for cognitive tasks—will shift relative demand toward occupations with lower education, lower wages, and a greater share of male workers.

Our starting point is a simple theoretical model based on Hampole, Papanikolaou, Schmidt, and Seegmiller (2025) that links task-specific technological advances to overall labor demand for an occupation. The model nests both direct and indirect channels. A technology that improves a task-specific form of capital substitutes for labor in that task. If it applies broadly across all tasks of an occupation—as with the automatic telephone switching system that displaced operators—the result is a sharp reduction in labor demand. If it applies only to a narrow set of tasks—say, a tool that automates expense reporting for academic economists—the negative impact on labor demand is mitigated, or possibly reversed. Specifically, because workers optimally reallocate time across tasks, automation in one task has ripple effects: it frees up labor for other tasks within the occupation, and it can raise overall productivity in ways that boost labor demand in other, not directly affected, occupations.

The model implies that the labor market impact of a specific technology can be summarized by three statistics. First, the *mean* exposure of an occupation's tasks to technology improvements is,

[1] In 350 BCE, Aristotle wrote: "[If] the shuttle would weave and the plectrum touch the lyre without a hand to guide them, chief workmen would not want servants, nor masters slaves." In 1811, skilled weavers and textile workers (the Luddites) worried that mechanizing manufacturing (and the unskilled laborers operating the new looms) would rob them of their means of income. In 1930, Keynes worried about technological unemployment: "we are being afflicted with a new disease of technological unemployment...due to our discovery of means of economising the use of labor outrunning the pace at which we can find new uses for labor." More recently, a McKinsey report estimated that between 400 million and 800 million jobs could be lost worldwide due to new technologies by the year 2030.

in general, negatively related to demand for that occupation. Thus, a moderate improvement in a technology that is related to all the tasks of a particular occupation will lower demand—even a basic tractor will reduce demand for farm labor. Second, the degree to which the technology exposure is *concentrated* to a subset of tasks allows workers to reallocate their effort and thus increases their productivity and the labor demand for their occupation. Third, the degree of productivity improvements as a result of these technological advances directly determine the shift in labor demand across all indirectly-affected occupations

Measuring workers' exposure to technology presents two core challenges: identifying technological improvements and determining the extent to which they substitute for specific tasks. We begin by using patent documents as our measure of innovation. While not all inventions are patented, the patent record provides a consistent proxy for technological change dating back to the mid-19th century. Mapping technology to tasks is more subtle. Occupation-specific tasks evolve over time as the structure of work changes. To address this, we use a state-of-the-art large language model (LLM) with web-search capabilities to generate comprehensive task descriptions for U.S. Census occupations in each decade from 1850 to 2010. We then quantify exposure using modern natural language processing techniques: both patent text and task descriptions are embedded as numerical vectors, and their semantic similarity is computed via cosine distance. A task is classified as exposed to a new technology if its vector representation closely aligns with that of contemporaneous patents. This approach allows us to track the evolving overlap between new technologies and the tasks performed by workers over time.

Naturally, the direction of technological innovation can be endogenous to the current state of the market for specific types of workers. To strengthen our interpretation of the findings as identifying the causal effect of technology on labor demand, we develop a shift-share identification strategy inspired by Acemoglu, Akcigit, and Kerr (2016). Specifically, our identification strategy leverages the degree to which breakthrough technological advances (according to Kelly, Papanikolaou, Seru, and Taddy, 2021) in upstream technologies diffuses to innovations in downstream, labor-saving technologies. For example, our instrument leverages the proliferation of breakthrough technologies occurring in the upstream patent CPC technology class "G11" (Information storage), which led to downstream improvements in the technology class "G06" ("Computing, calculating, and counting") during the 1980s and 1990s; around the turn of the 20th century, our IV approach predicts that upstream breakthroughs in technology class "B61" (Railways) resulted in downstream innovations in technology class "B60" (Vehicles in general).

Consistent with the model, we find that the average exposure of an occupation's tasks to the technologies developed in a given period is significantly negatively related to subsequent employment growth. However, for a given level of mean exposure, labor demand for that occupation increases if the technology exposure is concentrated on a subset of tasks. Examining how these results vary across different periods, we find that our results are broadly consistent across periods, with the exception of the 1880 to 1920 period. This early coincides with a number of changes in the definitions of occupations, a shift that the Census occupation classifications fail to adequately capture. Thus, in the remainder of the paper we focus on the post–1900 period. Doing so allows us to also control for differential industry trends (the Census starts collecting information on industry in 1910). Our results remain similar if we include the interaction of industry and decade year effects—in which case we are comparing employment growth in differentially exposed occupations within a particular industry. Last, using the Census industry assignments starting in 1910, our estimates reveal a positive relation between increases in the rate of patents that are relevant for a particular industry and employment across all occupations in that industry.

Taken together, our findings paint a fuller picture of how technological change has shaped the relative labor demand for different occupations over the last century or more. On the one hand, technology has directly substituted for specific worker tasks, which ceteris paribus has decreased labor demand for these tasks. However, this does not necessarily imply that the demand for labor decreased, due to the presence of these two quantitatively relevant offsetting forces: labor-saving technologies allowed workers to direct their effort to tasks not substituted by technology, while the resulting productivity improvements increased the overall demand for labor for affected occupations relative to others.

That said, the relative importance of these forces varies across occupations. In particular, consistent with the prevailing view on job polarization (Autor and Dorn, 2013; Autor, Katz, and Kearney, 2006; Goos, Manning, and Salomons, 2014), we find that occupations at the middle of the skill (income) distribution have been significantly more exposed to technology in the post-1980 period than workers at either the top or the bottom of the distribution. Importantly, however, our results indicate that this pattern of technology-induced job polarization likely started earlier—possibly to the middle and even early 20th century. This is consistent with Bárány and Siegel (2018), who argue that labor market polarization can be dated to at least the post-1950 period, not just post-1980. Moreover, we even find some evidence of technology-related polarization slightly earlier, going back to the 1910-1960 subsample of our data. In sum, we find that job polarization is not just a post-1980 story. It has been going on for decades, and technology accounts for a substantial share. But consistent with prior work, the phenomenon intensifies after 1980, and our technology measures play a larger role in that period.

In addition, we sort occupations into employment-weighted quintiles based on their average educational attainment or share of workers who are female. Across the whole sample period and within each sub-period we analyze, we find a consistent pattern of technology-induced decline in the employment shares of less-educated occupations, and also of male-intensive occupations. Thus our measures of occupational technological exposure are consistent with the pattern of rising returns to education in the labor market (Katz and Murphy, 1992; Card and Lemieux, 2001; Goldin and Katz, 2008) and its relation to skill-biased technological change (Berman, Bound, and Machin, 1998; Krusell, Ohanian, Ríos-Rull, and Violante, 2000). Our results on technology-induced employment expansion in female-heavy occupations in turn may also speak to the long-run decline in male labor force participation, which dates back at least to the mid-20th century (Parsons, 1980).

A key advantage of our measure is that it also includes time-series variation in the direction of innovation over time. Examining the time series, we see that, prior to 1980, innovation was consistently associated with manual physical tasks; by contrast, the innovations of the late 20th/early 21st century have become relatively more related to cognitive tasks. This pattern is partly driven by the increased prevalence of breakthrough patents related to computers and electronics. Last, occupations that are associated with interpersonal tasks have consistently low exposures to innovation throughout the entire sample period.

A natural question is whether the substitutability between new technologies and worker tasks is stable across task types and over time. To address this, we decompose our mean task exposure measure into manual, cognitive, and interpersonal components and relate each to subsequent occupation-level employment changes. Three patterns emerge. First, exposure of manual tasks is consistently associated with employment declines, indicating that technological progress has persistently substituted for manual effort. Second, exposure of cognitive tasks shows a time-varying relation: before the ICT revolution, it is associated with employment gains, but after the 1960s the effect turns negative. This shift suggests that earlier innovations complemented cognitive work, while more recent advances—especially in computing and software—have substituted for it. Finally, exposure of interpersonal tasks shows no systematic link to employment changes, implying that social skills may provide some insurance against labor-saving technologies, consistent with their rising importance in the labor market documented by Deming (2017).

We also exploit variation across worker cohorts to study how the link between employment growth and technology exposure differs by age. The employment changes we document are not driven solely by younger cohorts avoiding entry into exposed occupations. Instead, we find a strong negative relationship between mean exposure and employment growth among incumbents, tracking each age group forward as it ages. In specifications without the exposure concentration measure, the magnitude of the exposure coefficients increases steadily from the youngest to the oldest cohorts, consistent with human capital having a vintage-specific component. When we include concentration, however, the pattern shifts: concentration coefficients are especially large for the youngest incumbents, suggesting that the productivity gains from task reallocation accrue disproportionately to younger workers.

In sum, our empirical results largely validate the model's predictions regarding the impact of technology exposure on labor demand. In the final part of the paper we use the structure of the model, with some additional assumptions, to explore the impact of advances in Artificial Intelligence (AI) on relative labor demand across occupations over the medium run. Overall, a robust pattern that emerges is that we expect that AI advances to partially reverse the trends in relative employment growth induced by technology during the 20th century. That is, we expect that AI will increase the relative demand for occupations with lower education requirements, lower pay, and lower share of female workers. These cross-sectional predictions obtain under the assumption that AI is likely to automate certain cognitive tasks performed by workers that do not require significant prior experience—in much the same way that mechanization in the 20th century substituted for manual tasks performed by workers.

Our analysis connects to a large literature on labor-substituting technological change. One influential strand emphasizes the role of automation in displacing routine tasks, showing that occupations with higher routine-task intensity have been more exposed to recent labor-saving advances (Autor et al., 2006; Acemoglu and Autor, 2011; Goos et al., 2014). A second strand develops direct measures of labor-saving technologies and quantifies their labor-market consequences.[2] Our contribution is to bring a long-run perspective: rather than focusing on recent decades or specific technologies, we trace how successive waves of innovation since the mid-19th century have reshaped labor demand across occupations.

Closest to our work is Hampole et al. (2025) who leverage detailed data on worker resumes and firm job postings to examine the impact of AI on labor demand during the 2010 to 2020 period. Using their model as a guide, we leverage data on patents to construct measures of the mean and concentration of technology exposure at the task level over the last two centuries. Given the long time span of our analysis, the granularity of our exposure measures is significantly lower—we can only construct our exposure measures at the occupation level. Nevertheless, we reach some similar conclusions: occupations with high mean exposure to labor-saving technologies experience declines in labor demand, while the degree to which the exposure to technology exposure is concentrated to a subset of tasks had an offsetting effects. Unlike Hampole et al. (2025), we cannot observe technology adoption at the firm level, which limits our ability to estimate the degree of productivity spillovers in the data.

1 Theoretical Framework

To guide measurement, we first begin with a simple model based on Hampole et al. (2025). A central implication of their framework is that the effect of technological change on labor demand depends

[2] An incomplete list includes Webb (2020); Jiang, Tang, Xiao, and Yao (2021); Acemoglu and Restrepo (2022); Humlum (2019); Dauth, Findeisen, Suedekum, and Woessner (2021); Koch, Manuylov, and Smolka (2021); de Souza and Li (2023); Kogan, Papanikolaou, Schmidt, and Seegmiller (2023); Mann and Püttmann (2023); Jiang, Park, Xiao, and Zhang (2025); Dechezleprêtre, Hémous, Olsen, and Zanella (2021); Aghion, Antonin, Bunel, and Jaravel (2020).

on how capital-augmenting shocks are distributed across the tasks that define an occupation. When improvements are broad-based—raising capital productivity across most tasks—capital substitutes directly for labor, leading to a decline in occupational labor demand. By contrast, when advances are uneven and concentrated in a subset of tasks, workers shift effort toward the remaining tasks, raising their marginal productivity and potentially increasing labor demand overall.

1.1 Setup

A single final consumption good (the numeraire) is produced using a nested constant elasticity of substitution (CES) production function. In the outer nest, aggregate output $\bar{Y}$ is a CES composite of the output Y_I of a continuum of different industries indexed by I,

$$\bar{Y} = \left(\int_{\mathcal{I}} \alpha_I^{\frac{1}{\theta}} Y_I^{\frac{\theta-1}{\theta}} dI \right)^{\frac{\theta}{\theta-1}}, \tag{1}$$

where θ captures the elasticity of substitution across industries and α_I is a weight capturing an industry's importance in the final good. Here, α_I can represent the level of industry TFP or the number of products produced by the industry—see Hampole et al. (2025) for more details.

Each industry produces its output Y_I by combining the output of many occupations,

$$Y_I = \left(\int_{\mathcal{O}} \alpha(o,I)^{\frac{1}{\chi}} Y(o,I)^{\frac{\chi-1}{\chi}} \right)^{\frac{\chi}{\chi-1}}. \tag{2}$$

The parameter χ governs the elasticity of substitution across different occupations within an industry, and $\alpha(o,I)$ is a weight capturing the importance of occupation o in industry I output. Relative to Hampole et al. (2025), we simplify by assuming that firms are perfectly competitive within industries. Hence, all firms price at marginal cost and earn zero profits in equilibrium.

Workers in occupation o employed in industry I produce output $Y(o,I)$ as a CES composite of different tasks,

$$Y(o,I) = \left(\sum_{j \in \mathcal{J}(o,I)} \alpha(j)^{\frac{1}{\psi}} y(j)^{\frac{\psi-1}{\psi}} \right)^{\frac{\psi}{\psi-1}}, \tag{3}$$

where the relevant subset of tasks $\mathcal{J}(o,I)$ is occupation-specific. The CES parameter ψ governs the elasticity of substitution across tasks within a given job, controlling the extent to which tasks are complements ($\psi < 1$) or substitutes ($\psi > 1$) in production of $Y(o,I)$. We economize on notation at the task-level by suppressing the industry and occupation subscripts unless needed. As above, the weight $\alpha(j)$, which may vary by industry, occupation, and task, captures the importance of task j in production.

Each task j in job (o) is produced by a labor input $l(j)$ and a capital input $k(j)$,

$$y(j) = \left(\gamma_j \, l(j)^{\frac{\nu-1}{\nu}} + (1-\gamma_j) \, k(j)^{\frac{\nu-1}{\nu}} \right)^{\frac{\nu}{\nu-1}}. \tag{4}$$

Here, ν captures the elasticity of substitution between capital $k(j)$ and labor $l(j)$. In discussing comparative statics which obtain from the model, we assume that $\nu > \psi$. This assumption implies that reductions in the user cost of $k(j)$ specific to task j will likely be labor-saving.

As in Acemoglu and Restrepo (2022), $q(j)$ captures the quality-adjusted price of capital $k(j)$ that is specific to task j–i.e., the rate at which the final good can be transformed into capital specific to task j. Given our objective of capturing long-run shifts in the economy, capital is assumed to fully depreciate after production, and part of the final good is used an intermediate input in production of capital. Following Hampole et al. (2025) and Kogan et al. (2023), innovations impact the real economy by reducing the quality-adjusted price of capital $q(j)$

$$\Delta \log q(j) = -\varepsilon(j). \tag{5}$$

Each technological improvement is potentially applicable to several tasks within a job—defined as an occupation–industry pair. A given technological improvement that is applicable to job (o) can therefore be represented as an job-specific vector ε of weakly positive random variables. If $\varepsilon(j) > 0$, that implies the arrival of an improved (or cheaper) labor-saving technology that is specific to task j. For now, we are completely agnostic about the joint distribution of these technology improvements.

The last piece is labor supply. The first order condition for labor supply at the occupation-industry level satisfies

$$N(o, I) = \alpha_I \, \alpha_I(o) \; (W(o, I))^{\zeta} \tag{6}$$

where $W(o, I)$ is the equilibrium wage offered to workers in industry I and occupation o.

In addition to choosing how many efficiency units of labor to allocate to each occupation-industry cell, the Hampole et al. (2025) framework allows workers to optimally choose the fraction of time to allocate to each task in order to maximize total earnings. The effective supply of labor by workers in task j is equal to

$$l(j) = \alpha(j)^{\beta} h(j)^{1-\beta}, \tag{7}$$

where $\beta \in (0, 1)$ captures the degree of decreasing returns to effort at the task level. As $\beta \to 1$, the allocation of efficiency units across tasks is fixed at the exogenous level $\alpha(j)$. By contrast, as $\beta \to 0$, workers can frictionlessly reallocate time across tasks. Lowering β increases the flexibility of workers to respond to changes in their productivity across tasks by reallocating time. For each efficiency unit of labor, we normalize the total number of hours a worker can supply across all tasks is equal to one. Given the above, the optimal hours supply of a worker in job (o, I) to task j is equal to

$$h(j) = \frac{\alpha(j) w(j)^{\frac{1}{\beta}}}{\sum_{k \in \mathcal{J}(o,I)} \alpha(k) w(k)^{\frac{1}{\beta}}}, \tag{8}$$

where $w(j)$ is the occupation-specific wage in task j. Thus, incorporating the optimal hours choice in (8), a worker's total earnings in job (o) per efficiency unit equals

$$W(o,I) \equiv \sum_{j\in\mathcal{J}(o,I)} \alpha(j)^{\beta} h(j)^{1-\beta} w(j) = \left(\sum_{j\in\mathcal{J}(o,I)} \alpha(j) w(j)^{\frac{1}{\beta}} \right)^{\beta}, \tag{9}$$

which depends on her allocation of time and the (job-specific) task prices $w(j)$.

1.2 Model Implications

With our model in hand, we can analyze the impact of a shift in ε on equilibrium earnings and employment. Approximating around the symmetric equilibrium in which the labor share is constant across tasks, Hampole et al. (2025) derive a simple linear equation for employment growth due to changes in technology,

$$\Delta \log N(o,I) \approx \underbrace{\zeta\, \eta_m m(\varepsilon) + \frac{\zeta}{2\,\beta}\, \eta_o^2\, C(\varepsilon)}_{\text{Direct effects}} + \underbrace{\Delta \log \alpha_I + \zeta\, \eta_z\, \Delta_\varepsilon \log Z_I}_{\text{Industry Spillovers}} + \underbrace{\frac{\zeta\, \eta_z}{\theta - \chi} \Delta_\varepsilon \log \bar{\Omega}}_{\text{Aggregate Spillovers}}\,, \tag{10}$$

Equation (10) serves as the foundation of our empirical analysis. We next discuss its key components.

The first two terms in (10) capture the direct effect of technology improvements on labor demand, which depend on two key sufficient statistics. The first statistic is

$$m(\varepsilon) \equiv \sum_{j\in\mathcal{J}(o,I)} \frac{\alpha(j)}{\sum_{k\in\mathcal{J}(o,I)} \alpha(k)} \varepsilon(j). \tag{11}$$

that is, the task importance-weighted mean improvement of the technology across all tasks within the industry-occupation cell. The impact of the occupation's mean exposure (11) on labor demand depends on the elasticity of labor demand ζ and

$$\eta_m \equiv \quad -\frac{s_k(\nu - \chi)}{\zeta + \nu s_k + \chi(1 - s_k)} \tag{12}$$

where s_k is the capital share, which is assumed to be equal across all tasks.

The sign of η_m depends on whether the elasticity of substitution between capital and labor ν exceeds the elasticity of substitution χ across occupations within industries. To see why, consider a capital improvement that is specific to task j. This improvement directly substitutes for labor in task j, with ν capturing the elasticity of substitution between capital and labor. At the same time, the productivity of the occupation has increased, and thus labor demand for its output; and χ is equal to the elasticity of demand for that occupation (Hicks, 1932). The resulting impact on labor demand (and wage earnings) a function of the sign of $\nu - \chi$, that is, whether the decline in labor demand due to improvements in labor-saving technology greater than the increase in labor demand for the occupation as its productivity increases.

The second term in (10) captures the extent to which technological improvements are concentrated in specific tasks. This term depends on the degree to which technology improvements are concentrated in specific tasks,

$$C(\varepsilon) \equiv \sum_{j \in J} \frac{\alpha(j)}{\sum_{k \in J} \alpha(k)} \Big(\varepsilon(j) - m(\varepsilon)\Big)^2. \tag{13}$$

as well as

$$\eta_o = \quad -\frac{s_k \beta(\nu - \psi)}{(1-\beta) + \beta(\nu s_k + \psi(1-s_k))}, \tag{14}$$

which captures the impact of $\varepsilon(j)$ on the wage paid for task j $w(j)$ relative to the effect on $\varepsilon(j)$ on the wage paid on other tasks $j' \neq j$.

This concentration effect emerges due to two forces. First, in the model, workers can optimally respond to changes in labor productivity by reallocating time across tasks, with the scope for reallocation being inversely related to β. Second, log wages are a convex function of the vector of task-level wages, so Jensen's inequality implies that (appropriately-weighted) mean preserving spreads in task prices $\log w(j)$ increase occupation wages and labor demand. The more that productivity improvements are concentrated in a subset of tasks, both effects become larger quantitatively. The concentration effect is unambiguously positive, partially offsetting the (typically negative) direct effects associated with $m(\varepsilon)$.

The final two terms in (10) capture productivity spillovers. The third term captures increases in labor demand due to technology improvements. This term in turn depends on two economic forces. First, changes in α_I, for instance due to the creation of new products. Second, changes in industry productivity Z_I, as the cost of production falls. The extend to which changes in industry productivity increase labor demand depends on

$$\eta_z \equiv \frac{\partial \log w(j)}{\partial \log Z_I} = \frac{\theta - \chi}{s_k \, \nu + s_l \, \chi + \zeta}. \tag{15}$$

As industry I becomes more productive, it will increase its labor demand for occupations. At the same time, however, whether the demand for the industry's output increases, or decreases, depends on θ. If the elasticity of substitution across industries θ is greater than the elasticity of substitution across occupations within a given firm χ, then an increase in industry productivity leads to an increase in the wage of each task. The final term in (10) captures aggregate labor demand and supply effects.

In the next section, we will construct direct empirical proxies for the first three terms in (10). Most importantly, our empirical design will include calendar year fixed effects; therefore, we cannot directly estimate the last term in (10). Given this 'missing intercept' problem, our empirical regressions can only identify relative employment shifts across occupations and industries. This issue becomes even more salient in the pre–1910 sample—given that the Census only included industry codes in 1910. This implies that, for the pre-1910 period, both the third and the fourth terms (10) are absorbed.

2 Measurement

Our model in the previous section implies that the impact of a specific technology on labor demand for a specific occupation can be summarized by three objects: the average improvement in the labor-saving technology across the tasks performed by workers in that job (11), the degree to which these improvements are concentrated in specific tasks (13), and the impact of industry spillovers—the term involving Z_I in equation (10). In this section, we describe how we construct empirical analogues of these objects in the data.

2.1 Data Sources

Our board goal is to use the textual description of these innovations to relate these innovations to the tasks that workers do. Here, we briefly describe the sources of data for our empirical analysis. We relegate all details to the Online Appendix.

Technology Improvements

We measure technology improvements using patents. We obtain the textual description of innovations of patents for the 1836–2024 period from Google patents and PatentsView. Since the quality of the extracted text varies considerably over time, we use a modern large language model (LLM) to obtain a concise description of the key new innovation from each patent document. To represent these patent summaries into numerical data, we use embeddings provided by OpenAI.[3] Appendix B.1.1 and B.1.2 provide further details.

We next construct a coarse measure of which industries are particularly likely to see productivity improvements from technology p. In particular, as part of the LLM-generated summary in the previous section, we also query the LLM to provide a likely industry of use for each patent. Using the same OpenAI embeddings as above, we compute the cosine similarity between this industry-of-use descriptions from the LLM and the textual descriptions of the industries from the Census (using the time-consistent "ind1950" Census industry scheme). We assign a patent as relevant to a given Census industry if the Census industry title is the most textually similar to the patent's industry of use description (based on cosine similarity of the embeddings generated for the patent industry of use and Census industry title); we further assign the patent to the industry if the textual similarity is in

[3]We use the *openai: text-embeddings-3-small* model. Embeddings are geometric representations of the semantic meaning of text and have the property that the cosine similarity of two embedding vectors have a similar meaning (see, e.g. Mikolov, Sutskever, Chen, Corrado, and Dean, 2013).

the top 1% of all industry-patent similarity pairs for patents issued in that decade.[4] Appendix B.1.3 provides further details.

As part of some of the descriptive analysis in the next section, we group technologies into broad categories over different time periods. To do so, we estimate a k-means clustering algorithm on the document embeddings in order to separate patents into 20 groups over three distinct technology periods (1850–1920, 1920–1980, and 1980–present). Then, we provide a large number of example summaries from each cluster to the OpenAI GPT-o3 model and ask it to provide concise labels summarizing each technology. See Appendix B.1.4 for further details.

Tasks

A key challenge in our long-run empirical analysis is that the scope and function of occupations has changed over time. Given our use of the Census data, an important constraint is the Census definitions of occupations (which change over time). Unfortunately, a detailed description of what these occupations do is not available—the earliest vintage of the Dictionary of Occupational Titles starts in 1939. To overcome this challenge, we employ a specialized LLMs trained to understand and execute web search queries (*gpt-4o-search-preview*), and direct the LLM to provide a list of tasks performed by each Census occupation, in each decade from 1850 to 2010, in the style of occupation task descriptions in O*NET. The result of this query is a combination of 84,393 task–occupation–decades, spanning the years 1850 to 2010. On average, the search query returns approximately 14 tasks per occupation in each decade. Appendix Table A.1 lists some examples. As before, we represent these tasks as numerical vectors using the same embeddings from OpenAI that we use for the technology descriptions in the previous section. Appendix B.2.1 provides further details on the procedure.

These LLM-sourced task descriptions play a key role in our empirical analysis. To validate these descriptions, we compare their semantic meaning with either the Dictionary of Occupational Titles (DOT) or O*NET counterparts—whenever these are available. Specifically, for each occupation in the 1940, 1980, and 2010 decades, we compute the (cosine) similarity between the average task embedding of the LLM-sourced task descriptions and the average task embedding of its DOT or O*NET counterpart. We then compare the distribution of these similarity scores to a placebo distribution where we randomly compare the LLM tasks with a different occupation in the DOT or O*NET. As we see in Appendix Figure A.1, the task descriptions we obtain from the LLM are fairly similar to their DOT or O*NET counterparts from the same period and is significantly different from the placebo distribution. Interestingly enough, the LLM tasks from the 2010 period

[4]For example, the LLM indicates that the industry of use for US patent 6,009,696 ('Harvester head for dried-on-the-vine raisins') is "most likely to be used in the agricultural industry, specifically in the production of vine crops such as raisins, olives, and other tree crops." Based off the textual similarity between this sentence and the Census industry description, we match this patent to Census industry 356, 'agricultural machinery and tractors'.

are particularly close to the O*NET tasks, which suggests that these tasks were part of the LLM's search query. In brief, using this procedure allows us to measure worker tasks during periods where official sources are not easily available.

In addition, we use the same LLM (*gpt-4o-search-preview*) to obtain a classification of these tasks into three categories: manual, cognitive, and interpersonal. To help the LLM classify each task, we first manually classify a list of 41 distinct occupational work activities from O*NET into a cognitive, manual, or interpersonal category. We then provide the LLM with these categorizations as context for how to categorize a task into one of the three categories. Approximately 45% of the task–occupation–year triplets are classified as manual, 34% as cognitive, and 21% as interpersonal tasks. See Appendix B.2.2 for details.

Employment

We obtain employment counts using the IPUMS Census extracts. We extract information on gender, age, occupation, industry, and labor force participation. We restrict the sample to census respondents aged 15 to 75 that report that they are employed. We exclude members of the armed forces and occupation codes indicating a non-occupational response. Following Katz and Margo (2014), we include both men and women in our analysis; however, we also investigate the robustness of our findings to restricting to men. We compute employment in a specific occupation (or industry, when available) using the Census respondent weights. To compute employment growth across decades, we aggregate the decade–specific definition of census occupations into time-consistent classifications. Appendix B.3 contains further details.

2.2 Measuring Technology Exposure

The next step is to estimate the direction of technological progress, specifically, the extent to which the tasks of a particular occupation are exposed to technological innovation in a given period.

Task Technology Exposure

We measure the exposure of a task performed by a given occupation to a particular technology using textual similarity between the summary of patent documents and the LLM-sourced tasks described above. For each pair of patents p and tasks j in a particular decade T, we compute the cosine similarity between their OpenAI text embeddings. We consider a task j exposed to technology p if its cosine similarity exceeds a particular threshold: as our baseline, we consider the

95th percentile of the distribution.[5] Thus, we compute the probability that task j is exposed to technology improvements in the period from T to $T+H$ as

$$\text{Exposure}^H_{j,T} = \frac{1}{|P_{T,T+H}|} \sum_{p \in P_{T,T+H}} \mathbf{1}(\text{similarity}_{p,j} > p95). \tag{16}$$

Equation (16) measures the direction of technological progress in a given decade: the average exposure of task j to technologies in patents issued between years T and $T+H$. Through the lens of the model, we can interpret (16) as being proportional to the degree of technology improvements $\varepsilon(j)$ that are specific to task j in a particular period.

Occupation Technology Exposure

The next step involves extrapolating from task exposures to technology to the exposure of a particular occupation to technology. The model provides a useful guide—equations (11) and (13)–which state that an occupations exposure to technology is a function of not only its average task exposure to technology, but also the extent to which this exposure is concentrated in particular tasks. We construct the direct analogues of (11) and (13) in the data as

$$\text{Mean Exposure}^H_{o,T} = \frac{1}{|J(o,T)|} \sum_{j \in J(o,T)} \text{Exposure}^{\text{H}}_{j,T} \tag{17}$$

and

$$\text{Exposure Concentration}^H_{o,T} = \frac{1}{|J(o,T)|} \sum_{j \in J(o,T)} \left(\text{Exposure}^{\text{H}}_{j,T} - \text{Mean Exposure}^H_{o,T} \right)^2. \tag{18}$$

Given our timing convention, equations (17) and (18) measure the flow of innovation across one or two decades.[6]

Here, there are two important caveats that we should keep in mind. First, the measures of occupation exposure vary only at the occupation level; there is no variation across industries. The reason is that our assignments of patents to sectors using LLMs is a bit too coarse to accurately compute (17) and (18) at the sector–occupation level.

Second, our measure of task exposure (16) that is used to construct (17) and (18) measures

[5] Time-variation in our measure is confounded by changes in language over time. To address this, we de-mean patent–task similarity scores by year and compute percentile cutoffs separately for tasks in each Census decade. Specifically, for a task j linked to a Census occupation in decade T, we calculate the 95th percentile cutoff using the distribution of similarity scores across patents issued from 20 years before to 20 years after the start of decade T. For example, for 1940 Census tasks, the cutoff is based on patents issued between 1920 and 1959. Using alternative thresholds (e.g., the 90th or 99th percentile) yields nearly identical results. Because the cutoffs are decade-specific, low-frequency time-series variation in our measure should be interpreted with caution. However, secular shifts in language are less likely to confound the cross-sectional dimension of our exposure measures.

[6] The fact that occupation definitions change over time implies that we need to subtly modify (17) and (18) to map them into time-consistent Census occupation codes. In the cases when multiple decade–specific occupation codes map into a single time-consistent occupation code (using the Census occ1950 scheme), we compute weighted averages of (17) and (18), using the share of individuals in each aggregated occupation.

the direction of technological progress in a given decade. What is missing from these expressions, is a measure of the *intensity* of technological progress. Given our use of patent data, it is quite challenging to identify shifts in the intensity of innovation that is directed to specific tasks—and potentially specific sectors. To overcome this limitation, we will also explore the relation between technology exposure and employment growth period-by-period.

Industry Spillovers

Given the long historical period we study, obtaining consistent measures of industry output or productivity is difficult. We therefore rely on patents, which we assign to industries using the LLM-based procedure described above. Because the industry matching text-based and Census industry descriptions are sparse, there are some industries which persistently get matched to a large share of patents (such as ind1950 code 399, "Miscellaneous Manufacturing Industries"), while others get matched to a small share. This creates large level differences in the flow of patents matched to an industry. To account for this we instead rely on *growth rates* in the number of patents assigned to a given industry. For 20-year employment changes, our industry spillover measure is

$$\text{Spill}^{20}_{I,T} = \log(\text{Matched Patents}_{T \to T+20,I}) - \log(\text{Matched Patents}_{T-20 \to T,i}) \tag{19}$$

where Matched $\text{Patents}_{T \to T+20,i}$ is the number of patents issued between the years T to $T+20$ that are matched to industry I. The measure for the 10-year horizon $\text{Spill}^{10}_{I,T}$ is defined analogously. Accordingly, we map the broad changes in technology-related industry productivity implied by the model to the growth in the number of patents that are applicable to that industry.

A Shift-share IV

The direction of innovation is likely endogenous to the state of the labor market in a particular period. To address this concern we build a shift-share instrumental variable that builds on Acemoglu et al. (2016). In particular, our shift-share identification strategy leverages the extent to which variation in the arrival of breakthrough technologies in an 'upstream' technology class leads to the development of 'downstream' (labor-saving) technologies in other technology classes. Our identification assumption is that the development of these 'upstream' technologies is unrelated to the labor market conditions of the occupations exposed to the 'downstream' technologies.

The construction of our shift-share instrument entails two steps. The first step involves predicting the arrival of the number of patents $N_{c,t}$ in a given technology class c at time t using a Poisson regression. Our predictor is constructed based on previous breakthrough innovations in other tech classes that occurred in the past,

$$\lambda_{c,t} = \sum_{\tau} \sum_{c' \neq c} \Omega_{c' \to c,t,\tau} \times I_{c',t-\tau}. \tag{20}$$

The predicted number of downstream patents $\lambda_{c,t}$ depends on two objects. First, $\Omega_{c' \to c,t,\tau}$ is a technology diffusion matrix constructed based on the textual similarity of patents: its elements are the average similarity of patents in technology class c to patents in technology class c' for patents issued in tech class c at time t and tech class c' at time $t - \tau$; $\tau = 5, \ldots, 20$ represents a diffusion lag for innovation to propagate from tech class c' to c. When constructing Ω, we set its diagonal to zero: we only use spillovers to class c from other technology classes c'. Second, $I_{c,t}$ is equal to the intensity of breakthrough patents in class c and year t, measured by the number of patents that are breakthroughs according to Kelly et al. (2021) granted in year t and in technology class c relative to total patents in year t and tech class c. We estimate the Poisson regression at an annual frequency, allowing the estimated coefficients to vary by decade T, and then sum the predicted number of patents each decade. Overall (20) is a strong predictor of subsequent patenting in tech class c, with an average t-stat of 4.2 across the decade-by-decade coefficients. See Appendix C.4 for further details.

The second step involves estimating the likelihood that a patent p from technology class c is related to task j,

$$\alpha_{j,T} = \frac{1}{|P_{c,T-10}|} \sum_{p \in P_{c,T-1}} \mathbf{1}(\text{similarity}_{p,j} > P95). \tag{21}$$

When constructing the exposure shares (21), we use all patents issued in the previous decade.

Putting the pieces together, our shift-share measure for the exposure probability (16) of task j to technological innovation in period T is given by

$$Z^H_{j,T} = \sum_c \alpha_{j,T} \times \frac{\widehat{N}^H_{c,T}}{|P_{T,T+H}|}. \tag{22}$$

Our instrument is equal to the sum across technology classes of the product of the predicted shifts in the direction of innovation (21) times the predicted share of innovation in class c in decade T over the horizon H—the fitted values from the Poisson regression above divided by the total number of patents during this period. Using (22), we proceed to construct instruments for our mean (17) and concentration (18) measures, by replacing (16) with (22) in their definition.

Last, we instrument for industry-level technological spillovers as follows. The first step is to predict the level of patenting in the industry at the H-year horizon based on innovation in upstream technologies as

$$\text{Predicted Patents}_{T \to T+H,I} = \sum_c \Gamma_{I,c,T-10} \times \widehat{N}^H_{c,T}, \tag{23}$$

where $\Gamma_{I,c,T-10}$ is the probability that a patent in time period T belonging to tech class c is relevant for industry I given our textual mapping of patents to industries. We compute the growth rate of (23) at the H-year horizon to instrument for our industry spillover measure (19).

2.3 Examples

A key advantage of our measure is that it is available for essentially all years for which patents are available, and thus allows us to study very different technologies across long periods of time. Here, we first discuss some specific examples and then examine which major innovations are driving our exposure measures during different periods.

Examples from Specific Inventions

As an early example from the 19th century, consider the Bessemer Process (U.S. patents 16,082 and 49,055), the first method for mass-producing steel inexpensively. The occupations most exposed to these patents were in the iron and steel industries. Another example is the invention of the sewing machine. U.S. Patent 276,146, issued in 1883, is most closely linked to sewing machine operators, followed by carpet makers and thread makers. Other occupations had lower mean exposure overall but concentrated exposure in specific tasks, such as *tailors and tailoresses*, *shirt, cuff, and collar makers pattern makers*. These examples illustrate how major breakthroughs in the Second Industrial Revolution directly transformed manual occupations in textiles and metalworking.

A particularly vivid case of skill displacement is the cylinder machine (U.S. patent 814,612), a major innovation in the window-glass industry during the late 19th century and included by Kelly et al. (2021) in their list of breakthrough patents. Within a generation, the process of producing window glass shifted from hand-blowing to mechanization, and entire artisanal branches of the industry disappeared. As Jerome (1934) documents, wages of blowers and gatherers fell by 40 percent, and multiple skilled trades were eliminated. We identify *glass-works operatives* as being among the most related occupations to this patent. Moving to the early 20th century, Jerome describes the Barber-Colman warp-tying machine (patent 1,115,399), which could replace the work of about fifteen hand operators and be run by a single tender. In our methodology, the closest occupations include *thread makers*, *sewing machine operators*, *lace makers*, and *carpet makers*. Another example is the drawing-in machine (patent 1,364,091), which displaced five or six manual drawers-in with one operator and an assistant, again mapping closely to textile occupations. Other cases from this period further illustrate the link between particular patents and occupational exposure: the refrigerator (1,276,612), linked to traders and dealers in ice; the modular combine harvester (4,846,198), linked to farm laborers; the cotton picker spindle (2,716,320), linked to textile weavers; the washing machine (2,758,461), linked to laundresses; and the automatic telephone exchange systems (1,146,583 and 1,439,723), linked to telephone operators.

Not all examples from this period represent labor-saving innovations. Some patents describe technologies that support cognitive tasks and could plausibly complement labor rather than substitute for it. For instance, patent 1,138,792 (a calculating machine) relates to the task ‘prepare financial statements, such as balance sheets and profit and loss statements, by aggregating data from various ledgers and journals,' performed by accountants and auditors. Likewise, patent 2,655,941 pertains to the task "diagnose mechanical or hydraulic system failures to determine necessary repairs,' performed by airplane mechanics and repairmen. These cases underscore that exposure in our measure does not necessarily imply substitution, but can also capture technologies that enhance or complement human tasks.

Finally, in the second half of the 20th century, labor-saving patents shifted toward white-collar and service work. For example, U.S. patent 5,911,135 ("System for managing financial accounts by a priority allocation of funds among accounts") is most closely linked to *financial officers* and *accountants and auditors*. Patent 5,828,979 ("Automatic train control system and method") is related to *locomotive operating occupations* and *railroad conductors and yardmasters*. Patents 5,696,906 and 5,592,560, which cover e-commerce technologies, are most closely associated with billing clerks and sales occupations.

Which Technologies are Driving Exposure?

We next ask which types of technologies drive our exposure measures. To do so, we split the sample into three periods and assign patents within each to 20 broad technology groups, using the clustering and labeling procedure described in Section 2.1 and Appendix B.1.4. Figure 1 reports, for each period, the share of overall exposure (blue bars) and the share of patents (orange bars).

1850–1920. The left panel covers the Second Industrial Revolution. The largest source of exposure is woodworking machinery, reflecting the importance of agriculture and related industries early in the period. Railroads and textile/manufacturing machines are also key contributors, consistent with the replacement of artisanal production by mechanized factory production. By contrast, categories such as Packaging and Containers and Household and Leisure Devices account for a sizable share of patents but little worker exposure.

1920–1980. The middle panel shows a similar imbalance. Mechanization technologies continue to dominate worker task exposure, well beyond their patent share. Meanwhile, several important innovations—Industrial Chemical Processes, Chemical Compositions and Polymers, Electronics and Communications Systems—generate many patents but little direct exposure, suggesting they reshaped production without directly displacing tasks at the occupational level.

1980–present. The right panel highlights the ICT era. Software, Networking, and Security Systems emerge as the dominant source of worker task exposure, especially late in the sample. Earlier in the period, Printing, Paper, and Copying contributed disproportionately. Across the period, IT-related categories consistently drive exposure. Other breakthroughs, including Advanced Circuits and Signal Processing and Pharmaceutical and Biotech Innovations, account for large patent volumes but a relatively small share of task exposure. A key difference from earlier eras is that exposure now increasingly comes from white-collar, cognitive technologies, in contrast to the manual, physical technologies that dominated earlier periods. We examine these occupational differences in more detail below.

3 Technology Exposure and Employment

Here, we discuss our main empirical findings regarding the impact of technology on the employment shares of different occupations. Section 3.1 presents our first set of results which focus on identifying the direct effect of technology on labor demand, abstracting from productivity spillovers. Section 3.2 estimates the impact of productivity spillovers. Section 3.3 examines what our estimates imply for changes in the composition of employment across occupations by education, average occupational wage level, and gender. Section 3.4 considers heterogeneity in the link between exposure and employment growth by different task types and across different age groups.

3.1 Technology Exposure and Occupation Labor Demand

We begin our analysis using the full sample of occupation–decade employment counts spanning 1850 to 2020. Given the absence of consistent industry identifiers in the data, we are only able to identify the first two terms of equation (10). Thus, initially, we will abstract away from measuring the productivity effect of technology on labor demand.

Baseline Results

We first examine the impact of the mean and concentrated technology exposure of occupation o to technology improvements in decade beginning year T over a horizon of H of ten or twenty years. In particular, we estimate

$$\log\left(\frac{N_{o,T+H}}{N_{o,T}}\right) = \beta \, \text{Mean Exposure}^{H}_{o,T} + \gamma \, \text{Exposure Concentration}^{H}_{o,T} + c\, \mathbf{\Gamma}_{o,t} + \varepsilon_{o,T}. \tag{24}$$

Equation (24) is the direct empirical analogue of equation (10) in the model. Here, $N_{o,T}$ denotes the employment of occupation o in Census decade T. Given our timing convention, our ten-year specification relates the flow of innovation that occurred, for instance between years 1850 to 1860 to the growth in employment between 1850 to 1860. Depending on the specification, the vector of controls $\mathbf{\Gamma}$ includes calendar year effects; the share of occupation o in total employment in period T; and the occupation's employment growth over the last decade. Standard errors are clustered at the occupation level. Table 1 shows the estimated coefficients β and γ across different specifications.

First, we see that occupations with greater average exposure to technological improvements experience significant declines in employment growth relative to occupations with lower exposure. The coefficient β is consistently negative and statistically significant across all specifications, with magnitudes increasing at longer horizons. The OLS and IV estimates are comparable. Economically, these effects are substantial: a one-standard-deviation rise in mean task exposure predicts a 11 to 13 percentage point employment decline over a decade—for comparison, the standard deviation of occupation employment growth is approximately 75 percent across a decade.

Second, controlling for mean exposure, increased concentration of technology exposure across tasks partially offsets these employment declines. That is, we see that the estimated coefficient γ is now consistently positive across different specifications and horizons, though it is somewhat less precisely estimated than the β coefficient. In terms of magnitudes, holding the degree of mean exposure constant, a one-standard deviation increase in concentration is followed by approximately a 6 to 7 percentage point increase in employment over the next decade.

Third, the OLS and IV specifications are largely comparable across specifications. In general, there are two reasons to expect these two estimates to differ. First, there could be classical measurement error, which would tend to attenuate the OLS coefficients towards zero. Second, if labor-saving innovations are endogenously directed towards certain occupations that are expected to experience increased labor scarcity, then the OLS coefficient could be larger in magnitude. However, the fact that the IV coefficients are slightly larger than the OLS estimates suggests that this is type of selection is likely driving our estimates. Instead, it may be consistent with IV estimates improving attenuation error, or with different forms of selection (for example, innovation may instead be targeted on average towards occupations that are otherwise growing).

Fourth, interpreting the signs of these coefficients through the lens of the model, we see that the model indeed implies a positive estimate of γ. By contrast, a negative estimate of β strongly suggests that our measure of task exposure identifies technologies that substitute for labor in these tasks—that is, the elasticity of substitution between capital and labor ν is greater than the elasticity of substitution across tasks ψ.

Fifth, our baseline specification (24) focuses on contemporaneous effects of technology exposure on employment growth. Even though we estimate long differences, delayed adoption may generate lagged responses. To account for this, we re-estimate equation (24) using employment growth over longer horizons as the dependent variable, while holding exposure measures fixed at $H = 10$. Given the serial correlation of our mean and exposure measures (90 and 71 percent, respectively, at the 10-year horizon), we also include their lagged values to control for persistence in technological development. Figure 2 shows that estimated coefficients rise with the horizon but level off after about five decades, though standard errors widen given the limited sample. These results suggest that our baseline estimates likely understate the full extent of labor reallocation induced by technology.

Last, recall from our discussion in Section 2.2, our mean and exposure measures only identify the direction of technological innovation in a particular decade. Absent a direct measure of the intensity of technological innovation that is consistently available over the entire period, we then proceed to re-estimate equation (24) every Census decade. In Figure A.2, we plot the estimated coefficients β and γ in each cross-section that correspond to column (5) in the IV panel of Table 1.

Examining Figure A.2, we see that the estimated coefficients β and γ have consistent signs and magnitudes, with the notable exception being the 1880–1920 period, during which they are not really statistically different from zero. Appendix Table A.3 confirms this conclusion, by re-estimating equation (24) across three subsamples: corresponding to technology improvements in the 1850–1920, 1910–1970, and 1960–2020 period. Examining the table, we see that our estimates are statistically weaker during the 1850–1920 period. However, they are significant and of similar magnitude in both the 1910–1970 and 1960–2020 subsamples.

Several explanations stand out for why our results are weaker in the 1880—1920 sample, which spans the Second Industrial Revolution. First, technological advances of this era may have been more complementary to labor than in later decades, particularly for skilled white-collar work (Goldin and Katz, 1998). When $\nu \approx \chi$, the model predicts coefficients β and γ near zero, consistent with weak estimates. Second, technology may have affected labor demand beyond direct substitution in production. Improvements in transportation that expanded market access (Donaldson and Hornbeck, 2016), likely boosted demand for manufactured goods. In the context of our model, these changes would manifest as shifts in $\alpha_I(o)$ in equation (2). Third, many forces beyond technology likely drove large shifts in both labor supply and demand during 1880—1910. Standardization of tasks and the division of labor expanded the pool of workers able to produce goods, while immigration between 1870 and 1914 supplied many laborers suited for factory work (O'Rourke, Taylor, and Williamson, 1996; Abramitzky and Boustan, 2017). Finally, measurement is noisier in this period. Census occupation codes were less consistent, and the harmonized series are coarse. Factory jobs often received distinct occupation codes from pre-mechanization artisan work. The 1910 Census marked a shift to a more standardized, function-based classification system, improving longitudinal comparability. For these reasons, in the rest of the paper we focus on the post-1910 period.

Robustness Checks

We perform several robustness checks to our main analysis. First, our baseline results pool male and female workers. One concern is that long-run changes in female labor force participation could affect our findings. Appendix Table A.5 shows this is not the case: restricting the sample to men yields qualitatively and quantitatively similar results.

Second, we explore the sensitivity of our results if we were to obtain the tasks performed by each occupation using alternative LLMs. Our baseline uses *gpt-4o-search-preview*; we also generate tasks with GPT4o and Llama 403b. Appendix Tables A.6 to A.8 show that these models perform somewhat worse during 1850–1920, but lead to comparable results in 1910–1970 and 1960–2020.

Third, we re-estimate our baseline regression weighting occupations by their employment shares, right-winsorized at the top 5 percent to avoid domination by very large occupations. Appendix Table A.4 shows the results are similar to the unweighted baseline.

Fourth, our empirical measures of the mean and concentration in task exposure are fairly highly correlated in the cross-section—occupations with high degrees of technology tend to have concentrated task exposures. This raises the question of how do the estimates of β change if we just focus on the first-order effects. As we see in Appendix Table A.2, including only the mean exposure measure leads to smaller estimates of β: a one-standard deviation increase in mean exposure is now followed by approximately a 6 to 7 percentage point decline in employment.

Fifth, we test whether our measures simply capture broad secular shifts in labor demand, such as the decline of manual work. As Appendix Figure A.2 shows, manual tasks are more exposed to technology than other task types, so falling demand for manual labor could confound our results. We therefore add controls for each occupation's initial share of manual and cognitive tasks in the 1910–2020 sample. Appendix Table A.9 shows that coefficients on mean exposure shrink modestly and become less precise, but remain significantly different from zero. Coefficients on concentration, by contrast, fall sharply and lose precision, suggesting this measure partly reflects exposure differences already embedded in broad task categories.

Finally, we address the possibility that our estimates reflect industry-level trends such as the secular decline of manufacturing. We re-estimate equation (24) at the occupation-industry-decade level, restricting attention to the post-1910 period when industry codes are available. Appendix Table A.10 shows that including industry trends produces somewhat smaller coefficients but otherwise similar results.

3.2 Estimating the Impact of Technology Spillovers on Labor Demand

We next turn our attention to estimating the impact of productivity spillovers on labor demand. Doing so requires a modification to our empirical analysis: we now estimate our coefficients at the occupation–industry–decade level. Since industry classifications are only reliably available in the Census post–1910, we restrict attention to this period. In particular, we estimate

$$\log\left(\frac{N_{o,I,T+H}}{N_{o,I,T}}\right) = \beta \, \text{Mean Exposure}^{H}_{o,T} + \gamma \, \text{Exposure Concentration}^{H}_{o,T} + \delta \, \text{Spill}^{H}_{I,T} + c \, \mathbf{\Gamma}_{o,,I,t} + \varepsilon_{o,T}, \tag{25}$$

where $\text{Spill}^{H}_{I,T}$ is our measure of productivity spillovers in industry I in year T and horizon H. Now, $N_{o,I,T}$ denotes the employment of occupation o in Census decade T in industry I, where industries refer to the Census industry classifications (ind1950). As in our baseline results, the vector of controls $\mathbf{\Gamma}$ includes calendar year effects and the share of occupation o in total employment in period T. In addition, we include sector level (or sector interacted with decade) fixed effects. Sectors correspond to 16 broad industry categories derived from the "ind1950" classification. Interacting these sector level fixed effects with decade fixed effects allows us to partial out broad trends across sectors (for example, services vs manufacturing) while still at the same time identifying productivity spillovers across the more granular industries within each sector.

Table 2 presents our estimates of β, γ and δ. Examining the table, we note that the coefficient δ of employment growth is positively and significantly related to our productivity spillover measure. Focusing on Panel A that reports the OLS estimates, a one standard deviation increase in our spillover measure is associated with approximately a 8 to 10 percentage point increase in employment growth over the next decade, and a 17 to 20 percent increase over the next two decades. Examining the IV estimates in Panel B, we note that our instrument for spillovers is quite weak in the 10 year period, hence we view these estimates as not reliable. The instrument is somewhat stronger at a horizon of 20 years, in which case the IV estimates are close to the OLS estimates. Last, the coefficients β and γ are both strongly statistically significant across specifications, though their magnitudes a bit smaller than our baseline estimates.

The estimated coefficient δ is related to the third term in equation (15) capturing industry spillovers. This term is a function of the extent to which new technologies increase the number of new products at the industry level—the α_I term—and also depends on the relative value of the elasticity of substitution across industries θ compared to the elasticity of substitution across occupations within an industry χ, together with the magnitude of productivity improvements at the industry level.

Last, we explore the extent to which these estimates are consistent across different subsamples. We find that our estimates are broadly comparable across the 1910–1970 and 1960–2020 period, as we can see in Appendix Tables A.12 and A.13. Further restricting the sample to only male workers leads to very similar results, as we see in Appendix Table A.11.

3.3 Implications For Shifts in Composition of Labor Demand

So far, our estimates show how technology shifts relative labor demand across occupations. We now ask what these estimates imply for the changing composition of labor across different types of occupations. To do this, we re-estimate equation (25) decade by decade, focusing on a twenty-year horizon and plotting the resulting OLS estimates corresponding to column (8) of Table 2.[7] This approach lets us account for the fact that the mean and exposure measures capture the direction of innovation but not necessarily its intensity. Appendix Figure A.3 shows the estimates over time. The only real outlier is 1910–1940, when coefficients are especially large—consistent with Field (2003), who argues this was the most technologically intensive period of the 20th century. This

[7] The IV estimates for the mean and concentration exposure measures are close to the OLS results. The IV for the spillover measure is somewhat underpowered, especially in decade-by-decade estimation, so those coefficients are less stable.

decade-by-decade approach highlights not just the direction of innovation, but also the intensity of technology's impact on employment over time.

Next, we use the time-varying estimates of β, γ, and δ from equation (25) to construct predicted employment share growth at the occupation level:

$$\begin{aligned}\text{Tech-Predicted Growth}_{o,I,T\to T+20} =& \widehat{\beta}\ \text{Mean Exposure}^{H}_{o,T} \\ &+ \widehat{\gamma}\ \text{Exposure Concentration}^{H}_{o,T} + \widehat{\delta}\ \text{Spill}^{H}_{I,T}\end{aligned} \tag{26}$$

We compare this to actual employment share growth $\log(N_{o,I,T+20}) - \log(N_{o,I,T})$.

We then aggregate these occupation-level growth rates at a broader occupation group level using the start-of-period employment share of each occupation in the category. We group occupations in three ways: based on average educational attainment on each occupation; based on the female share of employment; and ten broad occupational groups sorted by average earnings, following Autor, Chin, Salomons, and Seegmiller (2024). To summarize across decades, we pool within group g and report averages of predicted and actual annual growth in employment shares. In all cases, we de-mean by subtracting the across-group average so that the pooled values sum to zero in each decade. This exercise allows us to assess how well technology explains long-run differences in employment growth by education, gender, and broad occupation group.

By Worker Skill (Education)

The mainstream view is that 20th-century technology was skill-biased: it raised demand for high-skill workers relative to low-skill ones (Goldin and Katz, 2008; Katz and Margo, 2014; Card and Lemieux, 2001). Berman et al. (1998); Krusell et al. (2000) formalize this pattern as skill-biased technical change. We test this idea by estimating predicted employment growth by education. We measure education requirements within occupations using the IPUMS 'edscore50' prior to 1980 and 'edscore90' afterward—the share of workers with at least one year of college within each Census occupation code. Panel A of Figure 4 plots the mean predicted employment growth by quintiles of education, for the full 1910–2020 sample. Appendix Figure A.4 breaks this into subperiods. Each panel also shows the realized employment changes.

Panel A of Figure 4 shows that, over the last century, technology raised demand for high-education occupations relative to low-education ones. The effect is large: our estimates imply that technological progress increased the relative demand for highly educated occupations by approximately 1.1 percentage points per year compared to the demand for occupations with lower levels of education. But the difference shrinks over time, falling to 0.9 percentage points in 1980–2020 as we see in Panel A of Table 3. A likely reason is the ICT revolution: in Section 3.1 we found evidence that ICT increasingly substitutes for cognitive skills. Since occupations with higher levels of education also tend to do more cognitive tasks than lower-educated occupations, this increased degree of substitutability likely led to relatively lower labor demand for these types of workers.

Panel A of Table 3 shows that over the full sample, the gap in realized employment growth is approximately 3 percentage points per year between the highest and the lowest occupation education score groups; our estimates of exposure to technological progress can account for a third of that gap over this period. Moreover, the top quintile of occupations have gained in employment share relative to the bottom quintile across all sub-samples; depending on the time period, technology explains between 27 to 47% of the change.

Our findings relate to Goldin and Katz (2008), who emphasize that the college wage premium fluctuated: high early in the century, compressed mid-century, then soaring after 1980. They attribute this to supply as well as demand. Our estimates show that technology consistently raised relative demand for educated occupations. But realized employment growth of educated occupations was largest in 1910–1960—precisely when our measures explain the least. This is consistent with supply-side factors (such as government subsidization of education post-WWII) causing a temporary increase in the growth of educated labor supply that outpaced pure technological considerations, leading to a temporarily deflated college premium.

By Occupation Wages

We next examine the degree to which our estimates can account for the observed degree of job polarization in the 20th century. Following Autor et al. (2024), we group occupations into ten broad categories, ordered left to right from low to high wage. Unlike their twelve-group scheme, we collapse the three small service categories into one 'low-skill services' group, given the small size of health services early in the panel. Panel B of Figure 4 covers the 1910–2020 period; Panel B of Table 3 and Appendix Figure A.5 split the sample.

Examining the blue bars in the middle, and consistent with the literature on polarization (Autor et al., 2006; Goos et al., 2014; Autor and Dorn, 2013), we find a decline in employment shares in low- and middle-skill occupations, like production, construction and transportation occupations, across the whole time period. Technology-predicted growth in employment shares tracks these declines closely. At the same time, low-skill services on the left and high-skill occupations—professionals, managers, and to a lesser extent sales and technicians—grow in relative terms. Again, technology exposure lines up with these shifts.

These patterns are consistent across sub-periods. Production declines in relative terms in every subsample, transportation in most. Our measures capture these patterns well. The post-1980 era shows the sharpest polarization, as highlighted in the literature, but we see that the process starts earlier. Consistent with Bárány and Siegel (2018), we find evidence of polarization in the post-1950 period, with a shift towards low-skill services and high-skill professionals (panel C of Appendix Figure A.5). Further we see in panel B of Appendix Figure A.5 that this pattern can even be traced earlier to the 1910-1960 period, suggesting that technology-related polarization has been an ongoing phenomenon for a longer period of time than previously recognized, even for up to a century.

To dig deeper, we focus on middle-skill, manual jobs—production, transportation, and construction—compared to low-skill services and to high-skill occupations (technicians, professionals, managers). Including clerical/admin with middle-skill or sales with high-skill leads to similar conclusions. Table 3, Panel B reports relative changes. Low-skill services grew 1.7 percent per year relative to middle-skill jobs, while technology predicts 0.8 percent. Technology thus explains about half of the shift. Across subperiods, it explains between 35 and 63 percent of the change, depending on the time period in question. In the bottom three rows of panel B, Table 3, we examine the other half of the polarization phenomenon: the shift towards high-skill occupations and away from these middle-skill occupations. From 1910–2020, high-skill jobs gained 2.1 percent per year relative to middle-skill jobs; technology explains 43 percent. By subperiod, the share explained ranges from 27 to 66 percent. In both cases, the contribution of technology is larger in the post-1980 era.

By Occupation Female Share

Female labor force participation has shifted dramatically over the past 150 years. As Ngai, Olivetti, and Petrongolo (2024) document, women's employment followed a U-shaped path: it fell as the economy moved out of farming, then rose again as employment shifted from manufacturing into services. Over the 20th century, the expansion of clerical, office, and later service jobs was central to rising female employment and hours (Goldin, 2006).

Our results show that technological progress over the past century has consistently favored white-collar service jobs over blue-collar production. A natural next question is whether these same forces can explain shifts in labor demand across occupations that differ by gender composition. We therefore compare predicted employment growth from our technology measures across occupations grouped by their initial female share. Panel C of Figure 4 covers 1910–2020; Appendix Figure A.6 examine subperiods separately.

Two patterns stand out. First, occupations with higher initial female shares grew substantially relative to male-dominated ones. Second, technology exposure explains much of this growth. Panel C of Table 3 shows that, from 1910 to 2020, female-intensive occupations gained about 1.6 percentage points per year relative to male-intensive occupations. Our measures explain about half of that—0.8 percentage points per year. Looking by subperiod, technology accounts for roughly one-third of the shift in both 1910–1960 and 1950–1990, and somewhat more than the observed shift during 1980–2020.

In sum, we find evidence that technology pulled women into the labor market, and it did so long before 1980. Our estimates show that demand shifted steadily toward female-intensive occupations as technology favored services over production. The mechanism we emphasize operates through labor demand, not just supply. Of course, technological change also likely boosted female labor supply by substituting for home production (see Greenwood, Seshadri, and Yorukoglu, 2005; Albanesi and Olivetti, 2016; Bose, Jain, and Walker, 2022). But the demand story is clear: occupations with more women grew faster, and technology explains a large share of that growth. This pattern links directly to our polarization results. As Ngai and Petrongolo (2017) argue, the service economy underpins rising female labor force participation. Our measures provide the technology-based mechanism for that argument, and they show the shift started as early as 1910–1960, not only after 1980.

3.4 Heterogeneity

We next turn to heterogeneity, asking how exposure differs across tasks and across age groups.

Heterogeneity by Task Types

Our analysis so far has implicity assumed that the elasticity of substitution between capital and labor is similar across different types of tasks. This may not be a particularly good assumption: technologies that are semantically similar to the description of interpersonal tasks are unlikely to have the same degree of substitution as technologies that are similar to manual tasks. Further, there probably reasons to think that these elasticities of substitution change over time.

To explore this idea further, we next decompose our mean exposure measure into three components, that is, the ones arising from the technology exposure of manual M, cognitive C, and interpersonal I tasks. Specifically, the mean exposure measure is equal to the sum of the exposure of the manual, cognitive and interpersonal mean task exposure, multiplied by their respective task shares,

$$\text{Mean Exposure}^{H}_{o,T} = \sum_{\tau \in \{M,C,I\}} \text{Task Share}^{\tau}_{o,T} \times \text{Mean Exposure}^{\tau,H}_{o,T}, \tag{27}$$

where each task-type $\tau\{M, C, I\}$ mean exposure is equal to

$$\text{Mean Exposure}^{\tau,H}_{o,T} = \frac{1}{|J(o,T,\tau)|} \sum_{j \in J(o,T,\tau)} \text{Exposure}^{\text{H}}_{j,T}, \tag{28}$$

and $J(o, T, C)$ denotes the tasks of type τ performed by occupation o in decade T. These measures are motivated by the model of Kogan et al. (2023), who show that they emerge as key sufficient statistics for worker exposure in a model with CES production and tasks that differ in terms of their elasticities of substitution between capital and labor (ν in our model).

Figure A.7 shows how the composition of which workers are most exposed has shifted across different task categories over the sample period. In particular, the figure shows the contribution of each task-type exposure measure 28 to the overall mean exposure measure (27), averaged across all occupations at each point in time. Examining the figure, a few points stand out. First, throughout the entire sample, occupations emphasizing manual tasks are overall much more exposed to automation than occupations emphasizing cognitive or interpersonal tasks. This is consistent with the traditional view that automation primarily affects workers in manufacturing industries (blue collar workers). However, the Information and Communications Technology (ICT) revolution has likely expanded the set of occupations that are effected to white collar workers. Indeed, as we see in the figure, there is an increasing trend in the exposure of occupations emphasizing cognitive tasks since the mid–1980s. Figure A.7 shows the breakdown of manual, cognitive, and interpersonal exposure across major clusters. ICT-related technologies, with software appearing at the top of the list, emerge among the set of technologies with the highest exposure to cognitive and interpersonal tasks.

Armed with this decomposition, we then re-estimate equation (24) and replace our mean exposure measure with its individual components. Table 4 presents the estimated coefficients. Panel *A* focuses on the post-1910 period, while Panels B and C report results separately for the 1910–1950 and 1960–2000 periods. Given that, in the context of the model, decomposing the exposure measure in this fashion to a large extent already captures the concentration of exposure to specific tasks, we report results with and without our baseline concentration measure.

Examining Table 4, we see two notable patterns. First, the exposure of manual tasks to technology is consistently negatively predictive of subsequent employment growth at the occupation level. Second, focusing on the entire post-1910 period in Panel A, we see that the exposure of cognitive tasks to technology is not significantly related to subsequent employment growth. The two subsequent panels B and C indicate why this is likely the case: in the 1910–1950 period, the exposure of cognitive tasks to technology is actually *positively* related to subsequent employment growth. By contrast, during the second period (1960–2000), we see that exposure to cognitive tasks is significantly *negatively* linked with employment growth. Throughout our sample, we see that the exposure of interpersonal tasks to technology—a small share of our overall exposure measure—are not related to subsequent employment growth. Perhaps not surprisingly, decomposing our mean measure into distinct task groups also implies that the relation between our overall task concentration measure and subsequent employment growth is less precisely estimated than before.

Through the lens of the model, these two patterns indicate that technology has consistently substituted for manual tasks throughout the 20th century. By contrast, in the pre-ICT period, technology served to complement cognitive tasks—in the context of the model, an elasticity of substitution between technology and congitive tasks that is smaller than the elasticity of substitution across tasks $\nu < \psi$. By contrast, the arrival of computers starting from 1960s and onwards implied that technology improvements such as computer software could also substitute for certain cognitive tasks as well—probably routine cognitive tasks, consistent with the evidence in Kogan et al. (2023); Hampole et al. (2025).

Heterogeneity by Age

Although the decennial Census provides repeated cross-sections, we can track cohorts over time and decompose employment growth by age group. Table 5 reports two sets of longitudinal comparisons. First, we study young workers–ages 16–25–over horizons $H \in 10, 20$:

$$\log\left(\frac{N_{o,age\in[16,25],T+H}}{N_{o,age\in[16,25],T}}\right) = \beta_{\text{entrants}}\,\text{Mean Exposure}^{H}_{o,T} + \gamma_{\text{entrants}}\,\text{Exposure Concentration}^{H}_{o,T} + c_{\text{entrants}}\,\mathbf{\Gamma}_{o,T} + \varepsilon_{o,entrants,T}. \quad (29)$$

Because these workers were at most 15 in year T, most are new entrants into the labor force. This specification therefore links our exposure measures to entry into exposed occupations. The first rows of Table 5 report β_{entrants} and γ_{entrants} across specifications and horizons.

Second, we follow cohorts as they age:

$$\log\left(\frac{N_{o,age\in[a+H,b+H],T+H}}{N_{o,age\in[a,b],T}}\right) = \beta_{\text{a-b}}\,\text{Mean Exposure}^{H}_{o,T} + \gamma_{\text{a-b}}\,\text{Exposure Concentration}^{H}_{o,T} + c_{\text{a-b}}\,\mathbf{\Gamma}_{o,T} + \varepsilon_{o,a-b,T} \qquad \text{for } [a,b] \in \{[16,25],[26,45],[46,65]\} \quad (30)$$

This specification measures the effect of exposure on employment growth for young, middle-aged, and older incumbents. Controls include year fixed effects, lagged employment shares, and the lagged average age of workers in the occupation.

Technological transitions may be particularly benign in a world in which reallocation can take place mostly through changes in entry of new workers, as these workers may be most easily able to switch between potential jobs without displacing existing skills and expertise. Stated differently, a model with switching costs which increase with experience (with age acting as a noisy proxy for experience), would likely predict employment elasticities that are declining in age (see, e.g., Kambourov and Manovskii, 2009; Cortes, Jaimovich, Nekarda, and Siu, 2020, for related evidence). If this was indeed the case, we would expect our estimate β_{entrants} to be larger in magnitude than our estimates for overall occupational employment and the analogous coefficient to be smaller for incumbents and decline with age.

Examining Table 5, we see that the impact of our exposure measures for entrants are somewhat larger than the pooled coefficients in Table 1, which is consistent with the presence of switching costs. Among incumbents, mean exposure coefficients decline sharply with age, with older cohorts significantly more negatively affected. The impact of exposure concentration declines with age, suggesting that suggests younger workers benefit more from opportunities for task reallocation.

Overall, these results show that technology-induced reallocation is not borne mainly by new entrants, but falls heavily on incumbents—especially older workers. The fact that the impact of mean exposure is increasing with age is inconsistent with the idea that switching costs increase with experience—unless these are switches into non-employment. By contrast, this pattern is consistent with the existence of vintage-specific human capital: new technologies displace existing skills and expertise, and older workers are more likely to hold those vulnerable skills. This evidence complements Kogan et al. (2023), who find similar displacement effects in the post-1980 period using data on individual worker earnings.

4 Implications for the Impact of AI on Labor Demand

Our results so far suggest that the model in Section 1 provides a useful framework for understanding how technological advances shape labor demand. We now turn to the implications of this framework for advances in AI. Forecasting is inherently difficult, so this exercise requires assumptions about which tasks AI is most likely to substitute for, the key model parameters, and the magnitude of spillovers. In Section 4.1, we lay out these choices, drawing where possible on the estimates from the previous section. Section 4.2 reports the resulting implications for labor demand, and Section 4.3 discusses the main caveats.

4.1 Calibration

Here, we briefly discuss our parametrization of the model. We relegate all details to Appendix D.

AI and Worker Tasks. The first step is to take a stand on which tasks AI is likely to substitute for. In our baseline, we assume AI can perform all cognitive tasks that do not require substantial experience, as measured by the O*NET 'specific vocational preparation' (SVP) score. Tasks requiring less than one year of SVP are treated as fully exposed; tasks requiring one to five years are treated as 50 percent exposed, facing half the capital price shock of a fully exposed task. Tasks are classified into SVP categories using an LLM prompt following Kogan et al. (2023), described in the appendix. Roughly 5 percent of tasks fall in the highest category (more than five years of training) and 32 percent require one to five years, implying that about one-third of tasks are exposed to AI substitution in this baseline. As a robustness check, we also use the task-level generative AI exposure measures from Eloundou, Manning, Mishkin, and Rock (2023). We take their preferred β version of exposure, assigning 50 percent exposure to tasks with a score of 0.5 and full exposure to tasks with a score of 1.

AI and Productivity Improvements. Next, we take a stand on the decline in the quality-adjusted price of "AI capital." We assume its cost will fall at the same rate as computer hardware. Caunedo, Jaume, and Keller (2023) estimate that computer prices declined by about 13 percent per year between 1984 and 2015. We assume the price of AI capital will decline by the same amount each year over the next decade.

AI and New Products. AI may also shift labor demand by enabling firms to introduce new products (Babina, Fedyk, He, and Hodson, 2024). To capture this channel, we allow for shocks to industry-level shifts in α_I, which we assume are driven by the industry's rate of new AI-related patenting relative to its stock of existing technologies. To simulate industry-specific growth in α_I, we combine the probabilistic assignment of patents to NAICS industries from Goldschlag, Lybbert, and Zolas (2019), data on AI patent counts and growth rates from the USPTO AI Patent Database of Pairolero, Giczy, Torres, Islam Erana, Finlayson, and Toole (2025), and an estimate of the elasticity of new product creation with respect to patenting from Argente, Baslandze, Hanley, and Moreira (2025).

Model Parameters. For the key elasticities, we set the elasticity of substitution across occupations, χ, to 1.34 following Caunedo et al. (2023). In the baseline we also set the elasticity across industries, θ, equal to 1.34 so that spillovers operate only through changes in α_I, and as a robustness check we consider $\theta = 1.72$ from Kogan et al. (2023), which introduces spillovers through the industry cost-efficiency index Z_I. The elasticity of substitution across tasks is set to $\psi = 0.5$, consistent with Humlum (2019) and Acemoglu and Restrepo (2022), while the degree of decreasing returns at the task level, β, has no clear empirical counterpart; we take $\beta = 0.5$ as the baseline but note that $\beta = 0.75$ yields similar results. The task-level elasticity of substitution between AI capital and labor, ν, is a central parameter in our calibration. We identify it from the empirical relation between employment growth and technology exposure, using the ratio of the estimated β coefficient to the square root of the estimated γ coefficient. This ratio is increasing in ν and independent of the magnitude of the technology shock ε. Targeting the OLS estimates from the 10-year horizon in column (4) of Table 2 yields a value of $\nu = 4.63$. Last, we calibrate the across-market elasticity of labor supply ζ to 0.42, following and Berger, Herkenhoff, and Mongey (2022).

Other Assumptions. The last step in parameterizing the model requires us to map the observed distribution of employment shares to the model. We do so using BLS OEWS data on occupation–industry employment shares, assuming an initial steady state with equal capital intensity across occupations so that wage bill shares match the 2024 distribution. To discipline the remaining task-level CES parameters, we assume symmetric initial capital prices and set the labor share to $s_l = 60\%$, following Karabarbounis and Neiman (2013).

4.2 Model Predictions

We focus on the model's predictions for shifts in relative labor demand across occupations. Figure 5 and the first column of Table 6 present the baseline results, grouping occupations by educational attainment (Panel A), hourly wages (Panel B), and female employment share (Panel C). Overall, the model implies that AI will raise relative demand for lower-educated, lower-paid, and more male-dominated occupations, partly reversing the patterns of technological change during the 20th century shown in Figure 4.

The effects are economically meaningful. Over the next decade, labor demand for high-education occupations is projected to fall by 0.65 percent per year relative to low-education jobs, though the occupations requiring average levels of education are projected to contract the most. Relative to mid-wage occupations, demand is expected to decline by 0.59 percent annually for managers, 0.29 percent for sales and professionals, and 0.85 percent for clerical and technical jobs. Finally, occupations with higher female shares are predicted to contract relative to male-dominated ones by about 0.53 percent per year.

Table 6 demonstrates that these predictions are robust to alternative parameterizations. Eliminating the possibility that AI stimulates new product creation—and hence suppressing industry-level increases in labor demand—yields nearly identical results, as shown in column (2) and Figure 5. Column (3) shows that using the task-exposure measure from Eloundou et al. (2023) produces qualitatively similar shifts, though the magnitudes are notably larger, particularly for sales and professional occupations, which are predicted to face an annual decline in labor demand of 0.82 percent relative to middle-wage jobs. Finally, column (4) indicates that relaxing the assumption $\theta = \chi$ has essentially no effect on the implied reallocation patterns.

4.3 Caveats

Our simple model makes a number of simplifying assumptions and omits several important economic forces. As such, our analysis comes with several important caveats which we discuss here.

New tasks. Advances in AI can lead to the creation of new tasks. This possibility is absent in our model, partly because LLM-sourced task descriptions are too crude to analyze trends in the creation of new tasks. Depending on which occupations perform these new tasks, this possibility can be an additional driver of cross-occupational shifts in labor demand.

AI as complement to Labor. Our model has a sharp distinction between AI capital and human workers. Our baseline assumption is that AI will substitute for certain cognitive tasks that are currently performed by workers. An alternative, perhaps highly speculative, view is that some of these technological advances can enhance the effective quantity of labor. For instance, advances in AI can be similar to improvements in human education—workers could pay a fee to download new skills and expertise immediately to their technology enhanced brains. In this case, AI can be a labor-augmenting technology that could complement certain types of tasks, similar to the findings for the non-routine exposure for the 1980-2010 period in Kogan et al. (2023).

Capital vs Wage Income. The prediction that AI advances are likely to reduce income inequality pertain solely to labor not capital income. To the extent that financial claims to the new technological advances in AI are not widely owned across households (as in the model of Kogan, Papanikolaou, and Stoffman, 2020) advances in AI can lead to an increase in income inequality at the right tail.

Skill Displacement. Most importantly, our unit of analysis is a specific occupation, not an individual worker. Even if advances in AI increase demand for a specific occupation, individual workers can be left behind if they lack the skills to fully take advantage of these new technologies—consistent with the evidence in Kogan et al. (2023). To the extent that younger workers face a lower cost of acquiring new skills than older workers—or simply have a larger incentive to—we would expect younger workers to be less adversely impacted relative to older workers from AI-induced declines in labor demand, consistent with results in Table 5.

5 Conclusion

Using new measures of technology exposure that are implied by our model, we document that advances in technology over the last century or more have led to meaningful shifts in labor demand across affected occupations: labor-saving technology advances have consistently increased the relative labor demand for occupations with higher education requirements, higher wages, and higher fraction of female workers.

References

Abramitzky, R. and L. Boustan (2017). Immigration in american economic history. *Journal of economic literature 55*(4), 1311–1345.

Acemoglu, D., U. Akcigit, and W. R. Kerr (2016). Innovation network. *Proceedings of the National Academy of Sciences 113*(41), 11483–11488.

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics, Volume 4. Amsterdam: Elsevier-North*, pp. 1043–1171.

Acemoglu, D. and P. Restrepo (2022). Tasks, automation, and the rise in US wage inequality. *Econometrica 90*(5), 1973–2016.

Aghion, P., C. Antonin, S. Bunel, and X. Jaravel (2020). What are the labor and product market effects of automation?: New evidence from france. Technical report, Cato Institute.

Albanesi, S. and C. Olivetti (2016). Gender roles and medical progress. *Journal of Political Economy 124*(3), 650–695.

Argente, D., S. Baslandze, D. Hanley, and S. Moreira (2025, September). Patents to Products: Product Innovation and Firm Dynamics. Working paper, Yale; Federal Reserve Bank of Atlanta; Compendium Labs; Northwestern University. Accessed via URL: https://www.davidargente.com/_files/ugd/b05023_c3467658dd6c4af2b3cd225e3e9780cc.pdf.

Autor, D., C. Chin, A. Salomons, and B. Seegmiller (2024, 03). New frontiers: The origins and content of new work, 1940–2018*. *The Quarterly Journal of Economics 139*(3), 1399–1465.

Autor, D. H. and D. Dorn (2013, August). The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review 103*(5), 1553–97.

Autor, D. H., L. F. Katz, and M. S. Kearney (2006, May). The Polarization of the U.S. Labor Market. *American Economic Review 96*(2), 189–194.

Babina, T., A. Fedyk, A. He, and J. Hodson (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics 151*, 103745.

Berger, D., K. Herkenhoff, and S. Mongey (2022). Labor market power. *American Economic Review 112*(4), 1147–1193.

Berman, E., J. Bound, and S. Machin (1998, 11). Implications of skill-biased technological change: International evidence*. *The Quarterly Journal of Economics 113*(4), 1245–1279.

Bose, G., T. Jain, and S. Walker (2022). Women's labor force participation and household technology adoption. *European Economic Review 147*, 104181.

Bárány, Z. L. and C. Siegel (2018, January). Job polarization and structural change. *American Economic Journal: Macroeconomics 10*(1), 57–89.

Card, D. and T. Lemieux (2001, 05). Can falling supply explain the rising return to college for younger men? a cohort-based analysis*. *The Quarterly Journal of Economics 116*(2), 705–746.

Caunedo, J., D. Jaume, and E. Keller (2023, June). Occupational exposure to capital-embodied technical change. *American Economic Review 113*(6), 1642–85.

Cortes, G. M., N. Jaimovich, C. J. Nekarda, and H. E. Siu (2020). The dynamics of disappearing routine jobs: A flows approach. *Labour Economics 65*, 101823.

Dauth, W., S. Findeisen, J. Suedekum, and N. Woessner (2021, 05). The Adjustment of Labor Markets to Robots. *Journal of the European Economic Association 19*(6), 3104–3153.

de Souza, G. and H. Li (2023, November). Robots, Tools, and Jobs: Evidence from Brazilian Labor Markets. Working Paper Series WP 2023-42, Federal Reserve Bank of Chicago.

Dechezleprêtre, A., D. Hémous, M. Olsen, and C. Zanella (2021). Induced automation: evidence from firm-level patent data. *University of Zurich, Department of Economics, Working Paper* (384).

Deming, D. J. (2017, 06). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics 132*(4), 1593–1640.

Donaldson, D. and R. Hornbeck (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics 131*(2), 799–858.

Eloundou, T., S. Manning, P. Mishkin, and D. Rock (2023). Gpts are gpts: An early look at the labor market impact potential of large language models.

Field, A. J. (2003). The most technologically progressive decade of the century. *American Economic Review 93*(4), 1399–1413.

Goldin, C. (2006, May). The quiet revolution that transformed women's employment, education, and family. *American Economic Review 96*(2), 1–21.

Goldin, C. and L. Katz (2008). *The Race Between Education and Technology.* Harvard University Press.

Goldin, C. and L. F. Katz (1998, August). The origins of technology-skill complementarity. *The Quarterly Journal of Economics 113*(3), 693–732.

Goldschlag, N., T. Lybbert, and N. Zolas (2019). An 'algorithmic links with probabilities' crosswalk for uspc and cpc patent classifications with an application towards industrial technology composition. *Economics of Innovation and New Technology 28*(6), 589–609.

Goos, M., A. Manning, and A. Salomons (2014, August). Explaining job polarization: Routine-biased technological change and offshoring. *American Economic Review 104*(8), 2509–26.

Greenwood, J., A. Seshadri, and M. Yorukoglu (2005, 01). Engines of liberation. *The Review of Economic Studies 72*(1), 109–133.

Hampole, M., D. Papanikolaou, L. D. Schmidt, and B. Seegmiller (2025, February). Artificial intelligence and the labor market. Working Paper 33509, National Bureau of Economic Research.

Hicks, J. (1932). *The Theory of Wages.* Macmillan.

Humlum, A. (2019). Robot adoption and labor market dynamics. *Princeton University.*

Jerome, H. (1934). *Mechanization in Industry.* NBER.

Jiang, W., J. Park, R. J. Xiao, and S. Zhang (2025, February). Ai and the extended workday: Productivity, contracting efficiency, and distribution of rents. Working Paper 33536, National Bureau of Economic Research.

Jiang, W., Y. Tang, R. J. Xiao, and V. Yao (2021). Surviving the fintech disruption. Technical report, National Bureau of Economic Research.

Kambourov, G. and I. Manovskii (2009). Occupational specificity of human capital. *International Economic Review 50*(1), 63–115.

Karabarbounis, L. and B. Neiman (2013, 10). The Global Decline of the Labor Share*. *The Quarterly Journal of Economics 129*(1), 61–103.

Katz, L. F. and R. A. Margo (2014, January). Technical change and the relative demand for skilled labor: The united states in historical perspective.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963-1987: Supply and demand factors. *The Quarterly Journal of Economics 107*(1), 35–78.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021, September). Measuring technological innovation over the long run. *American Economic Review: Insights 3*(3), 303–20.

Keynes, J. M. (1930). *Economic Possibilities For Our Grandchildren*, Volume 9 of *The Collected Writings of John Maynard Keynes*, pp. 321–332. Royal Economic Society.

Koch, M., I. Manuylov, and M. Smolka (2021, 01). Robots and Firms. *The Economic Journal 131*(638), 2553–2584.

Kogan, L., D. Papanikolaou, L. D. Schmidt, and B. Seegmiller (2023, November). Technology and labor displacement: Evidence from linking patents with worker-level data. Working Paper 31846, National Bureau of Economic Research.

Kogan, L., D. Papanikolaou, and N. Stoffman (2020). Left behind: Creative destruction, inequality, and the stock market. *Journal of Political Economy 128*(3), 855–906.

Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica 68*(5), 1029–1053.

Mann, K. and L. Püttmann (2023, 05). Benign Effects of Automation: New Evidence from Patent Texts. *The Review of Economics and Statistics 105*(3), 562–579.

Mikolov, T., I. Sutskever, K. Chen, G. Corrado, and J. Dean (2013). Distributed representations of words and phrases and their compositionality. *CoRR abs/1310.4546*.

Ngai, L. R., C. Olivetti, and B. Petrongolo (2024, May). Gendered change: 150 years of transformation in us hours. Working Paper 32475, National Bureau of Economic Research.

Ngai, L. R. and B. Petrongolo (2017, October). Gender gaps and the rise of the service economy. *American Economic Journal: Macroeconomics 9*(4), 1–44.

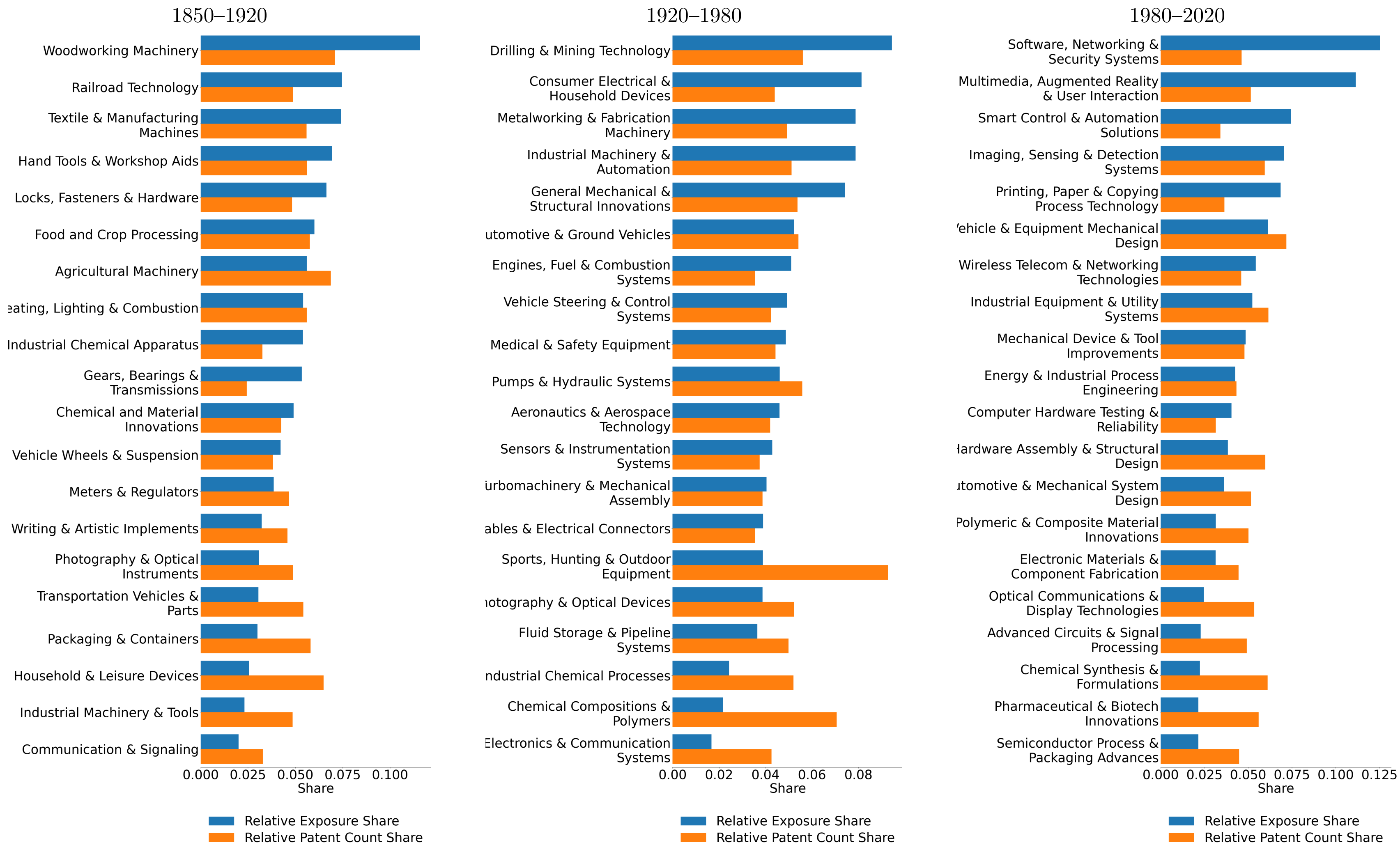
O'Rourke, K. H., A. M. Taylor, and J. G. Williamson (1996). Factor price convergence in the late nineteenth century. *International Economic Review*, 499–530.

Pairolero, N. A., A. V. Giczy, G. Torres, T. Islam Erana, M. A. Finlayson, and A. A. Toole (2025). The artificial intelligence patent dataset (aipd) 2023 update. *The Journal of Technology Transfer*, 1–25.

Parsons, D. O. (1980). The decline in male labor force participation. *Journal of Political Economy 88*(1), 117–134.

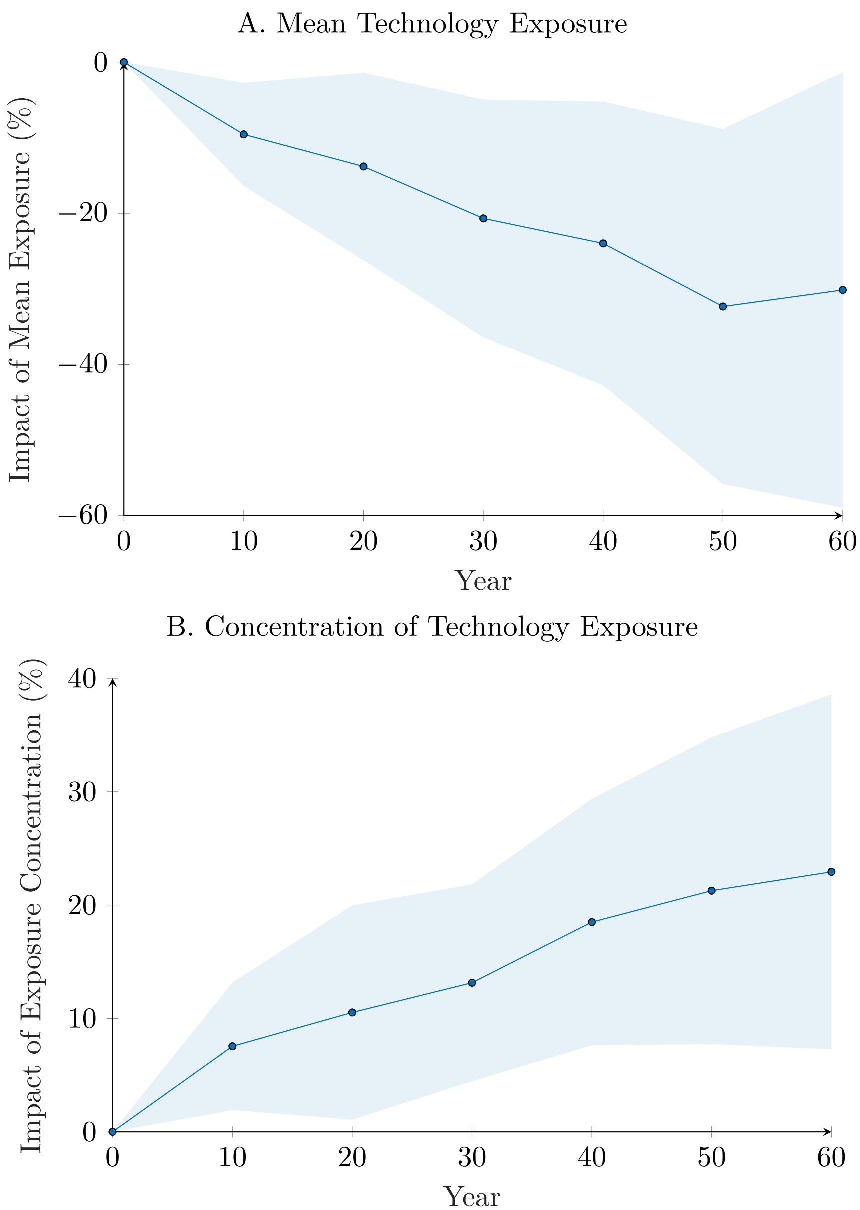
Webb, M. (2020). The impact of artificial intelligence on the labor market.

Figure 1: Composition of Overall Technology Exposure, by Technology Clusters



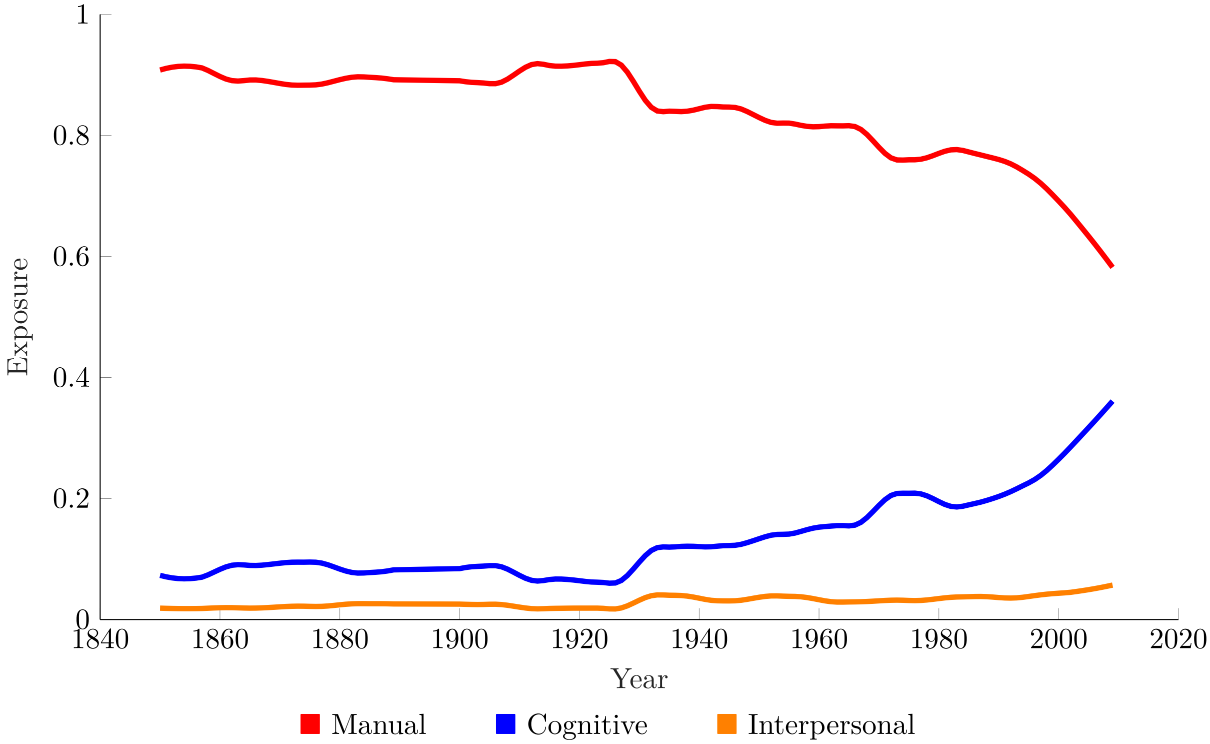
Note: This figure plots the composition of technology exposure across 20 major technology clusters identified for the 1980-2025 period. Specifically, we use a k-means algorithm to separate technologies into 20 clusters, then use a LLM to generate short labels for each. Plot at the left illustrates the share of total occupational exposure and the fraction of patents associated with each cluster, in orange and blue, respectively. Clusters are sorted vertically based on descending contributions to occupational exposure. The plot at right shows the fraction of patents coming from each of the clusters across years in our sample.

Figure 2: Technology Exposure and Employment Growth, Dynamic Effects



Note: This figure plots the coefficients from a long-horizon version of equation (24), in which the dependent variable is long-horizon growth in education employment. The specification uses our 10-year exposure measures as independent variables, and controls for one lag of the employment share and one lag of our exposure measures. We report the IV coefficients. Shaded bands represent 95 percent confidence intervals.

Figure 3: Composition of Overall Technology Exposure, by Task Type

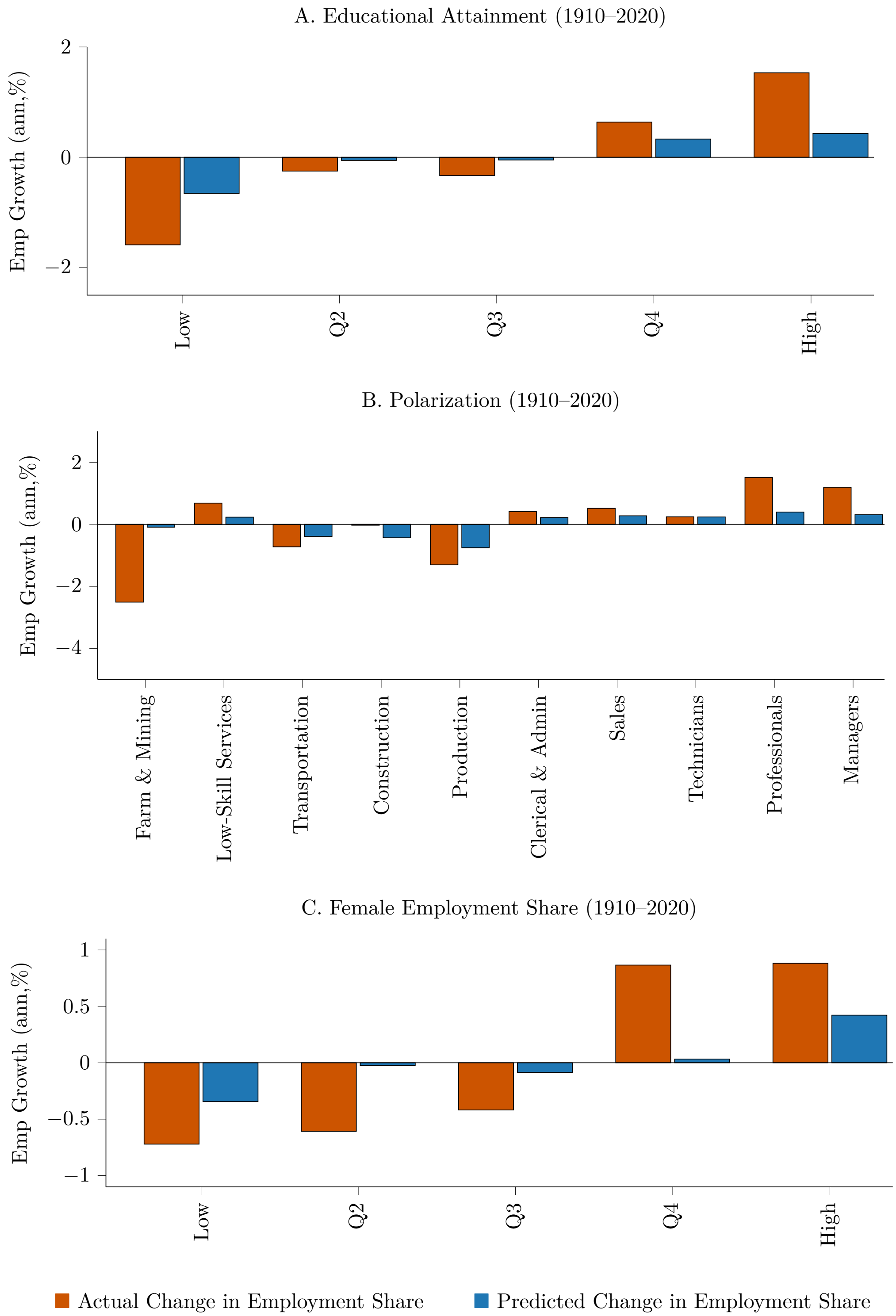


Note: This figure plots the composition of technology exposure by each task type $\tau \in \{\text{Manual, Cognitive, Interpersonal}\}$. The composition of each type-τ task, c_τ, is defined as the share of all valid patent-task link that are contributed by type-τ tasks. Specifically, we define

$$c_\tau = \frac{\sum_{j \in J(T,\tau)} \sum_{p \in P_t} \mathbf{1}(\text{similarity}_{p,j} > p95)}{\sum_{j \in J(T)} \sum_{p \in P_t} \mathbf{1}(\text{similarity}_{p,j} > p95)},$$

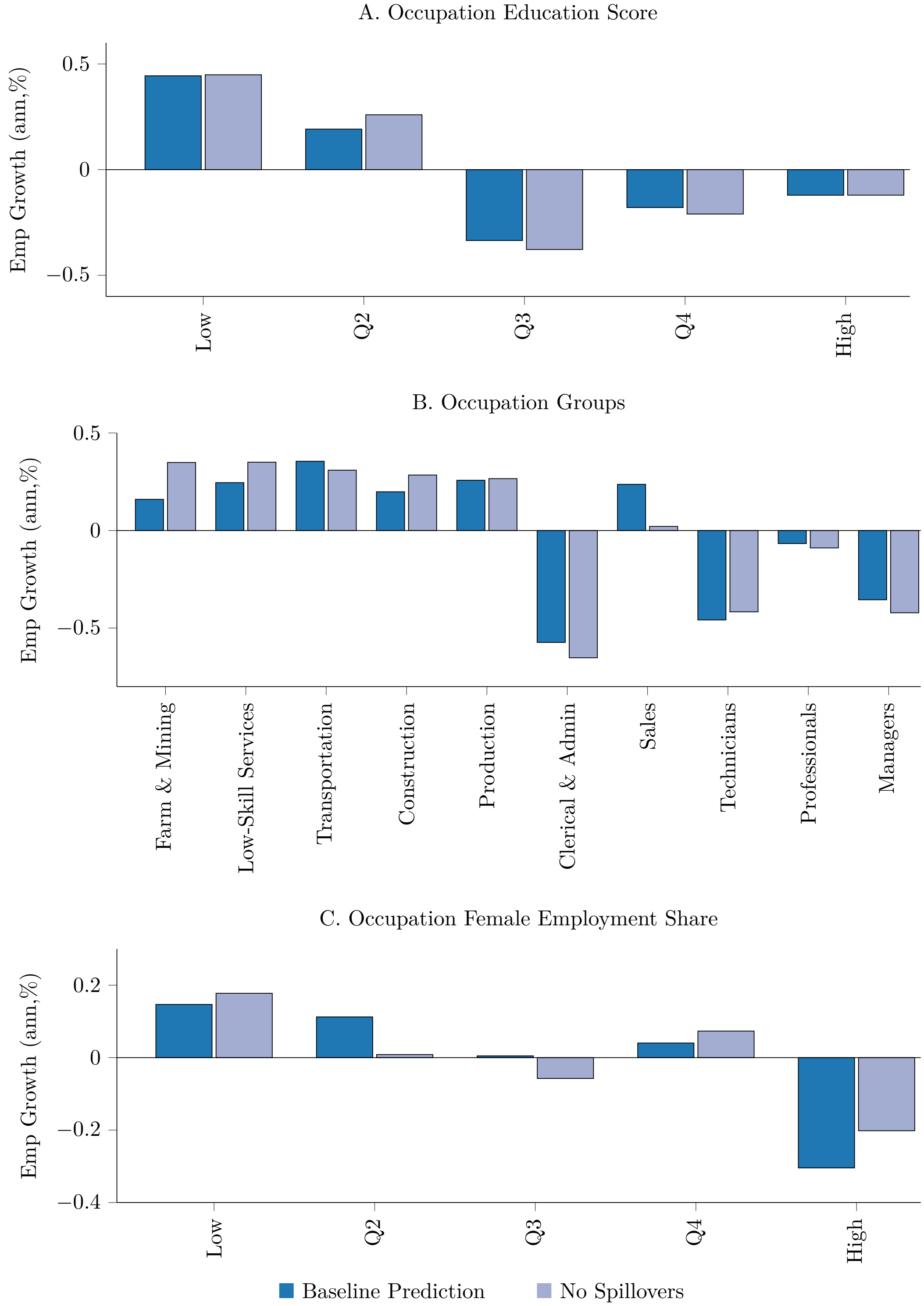
where P_t is the set of all patents issued in year t, $J(T)$ is the set of all tasks in decade T containing year t, and $J(T,\tau)$ contains all type τ tasks in $J(T)$. The figure plots c_τ for years $t \in [1850, 2010]$. The plot is LOWESS-smoothed with parameter 0.1.

Figure 4: Technology Exposure and Shifts in Labor Demand Across Occupations



Note: This figure plots actual and technology-predicted average growth rates in employment shares based off estimates of equation (A.3), and also by occupational educational rank, following the procedure described in section 3.3 of the main text. In Panel A we sort occupations into yearly employment-weighted quintiles based off educational attainment (IPUMs variable edscor50 for years before 1980 and edscor90 for 1980 onwards). In Panel B, group occupations into broad time-consistent groups following Autor et al. (2024); occupation groups are sorted from left to right based off their average wages. In Panel C we group occupations based on the share of female workers in the occupation.

Figure 5: AI and Shifts in Labor Demand



Note: This figure plots predicted annualized relative employment growth from our AI model simulations described in section 4 of the main text, for occupations categorized by education score, broad occupation group, and share of workers who are female.

Table 1: Technology Exposure and Employment Growth, Direct Effects

	A. OLS					
	10-yr Horizon			20-yr Horizon		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Task Exposure	-13.2***	-13.2***	-11.6***	-21.2***	-21.0***	-21.9***
	(1.92)	(1.93)	(2.26)	(3.33)	(3.34)	(3.56)
Concentration in Task Exposure	6.54***	6.36***	7.16***	8.90***	8.47**	12.8***
	(1.91)	(1.91)	(2.38)	(3.31)	(3.32)	(3.82)
Obs	3,212	3,212	2,452	3,166	3,166	2,410
R^2 (Within)	0.019	0.026	0.046	0.024	0.035	0.068
	B. IV					
	10-yr Horizon			20-yr Horizon		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Task Exposure	-13.5***	-13.5***	-11.4***	-23.7***	-23.8***	-21.6***
	(2.08)	(2.09)	(2.42)	(3.61)	(3.63)	(3.78)
Concentration in Task Exposure	6.38***	6.28***	6.31**	10.0***	9.82***	10.3**
	(2.12)	(2.12)	(2.62)	(3.65)	(3.66)	(4.19)
Obs	3,212	3,212	2,452	3,166	3,166	2,410
R^2 (Within)						
F stat (Exposure)	6,606	6,624	5,554	4,282	4,288	3,663
F stat (Concentration)	2,222	2,229	1,865	1,487	1,491	1,302
Year FE	X	X	X	X	X	X
Employment Share, Lag		X	X		X	X
Employment Share, Lag growth			X			X

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,T+H}}{N_{o,T}}\right) = \beta\,\text{Mean Exposure}^{H}_{o,T} + \gamma\,\text{Exposure Concentration}^{H}_{o,T} + c\,\mathbf{\Gamma}_{o,t} + \varepsilon_{o,T}.$$

for decades T spanning from 1850–2000, excluding 1890. The variables of interest are Mean Exposure$^{H}_{o,T}$, technology mean exposure, and Exposure Concentration$^{H}_{o,T}$, technology exposure concentration (both normalized to unit standard deviation). Controls $\mathbf{\Gamma}_{o,t}$ include year fixed effects, lagged employment share $N_{o,T}$, and lagged employment growth $\log\left(\frac{N_{o,T}}{N_{o,T-10}}\right)$. The controls included in each regression specification are denoted by X. Coefficients are multiplied by 100. The top panel reports the estimated coefficients using OLS, while the bottom panel reports the IV estimates. Standard errors (in parentheses) are clustered by occupation. Columns 1–3 show regressions with $H = 10$, and columns 4–6 show regressions with $H = 20$.

Table 2: Technology Exposure and Employment Growth, Industry Controls and Innovation Spillovers

	A. OLS							
Employment Growth (%)	10 Years				20 Years			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-4.22***	-4.20***	-7.96***	-8.02***	-7.23***	-7.19***	-15.0***	-15.0***
	(1.23)	(1.22)	(2.05)	(2.05)	(2.24)	(2.24)	(3.35)	(3.36)
Concentration in Task Exposure			4.53**	4.64**			9.51***	9.56***
			(1.86)	(1.86)			(2.76)	(2.77)
Industry Spillover	8.16***	10.6***	8.17***	10.6***	17.3***	20.4***	17.3***	20.4***
	(2.45)	(1.78)	(2.45)	(1.78)	(3.37)	(3.39)	(3.37)	(3.39)
N	137,192	137,192	137,192	137,192	126,463	126,463	126,463	126,463
R^2 (Within)	0.005	0.006	0.006	0.007	0.013	0.013	0.014	0.015

	B. IV							
Employment Growth (%)	10 Years				20 Years			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-4.82***	-4.96***	-8.08***	-8.25***	-9.55***	-9.47***	-16.2***	-16.2***
	(1.22)	(1.21)	(2.06)	(2.06)	(2.15)	(2.15)	(3.21)	(3.23)
Concentration in Task Exposure			4.06**	4.10**			8.44***	8.59***
			(1.88)	(1.90)			(2.83)	(2.84)
Industry Spillover	44.5**	43.3**	44.5**	43.3**	15.3*	19.1**	15.3*	19.1**
	(20.39)	(17.18)	(20.39)	(17.18)	(7.94)	(7.40)	(7.94)	(7.40)
N	135,637	135,637	135,637	135,637	125,956	125,956	125,956	125,956
R^2 (Within)								
F stat (Exposure)	4,151	4,168	2,892	2,903	2,295	2,370	1,618	1,676
F stat (Concentration)			1,100	5,384			760	796
F stat (Spillover)	4	3	7	29	24	21	16	40
Year FE	X		X		X		X	
Sector FE	X		X		X		X	
Year × Sector FE		X		X		X		X
Employment Share, Lag	X	X	X	X	X	X	X	X

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,I,T+H}}{N_{o,I,T}}\right) = \beta\,\text{Mean Exposure}^{H}_{o,T} + \gamma\,\text{Exposure Concentration}^{H}_{o,T} + \delta\,\text{Spill}_{I,T} + c\,\mathbf{\Gamma}_{o,I,t} + \varepsilon_{o,I,T},$$

for decades T spanning 1910–2000. Controls $\mathbf{\Gamma}_{o,I,t}$ include year fixed effects, broad-sector fixed effects (or year × sector fixed effects), and lagged employment share. Sectors which correspond to 16 broad industry categories derived from the "ind1950" classification: "Agriculture, Forestry, and Fishing"; "Mining"; "Construction"; "Durable Goods Manufacturing"; "Non-Durable Goods Manufacturing"; "Transportation"; "Telecommunications"; "Utilities and Sanitary Services"; "Wholesale Trade"; "Retail Trade"; "Finance, Insurance, and Real Estate"; "Business and Repair Services"; "Personal Services"; "Professional and Related Services"; and, "Public Administration". Coefficients correspond to a one-standard-deviation change in the independent variable and are multiplied by 100. Standard errors (in parentheses) are clustered by occupation and industry. Panel A reports the estimated coefficients using OLS, while Panel B reports the IV estimates. Columns 1, 2, 5, and 6 show estimates excluding exposure concentration measure, while the others include it. Columns 1–4 show regressions with $H = 10$, while the others show regressions with $H = 20$.

Table 3: The role of technology in accounting for observed shifts in labor composition

A. Occupation Education Level	Full Sample	Sub-periods		
	1910-2020	1910-1960	1950-1990	1980-2020
Employment Growth (ann, %)				
Realized: High - Low	3.1	4.0	3.1	2.0
Predicted: High - Low	1.1	1.1	1.2	0.9
% of Gap Explained	34.8	27.5	39.4	47.4
B. Occupation Wages	**Full Sample**	**Sub-periods**		
	1910-2020	1910-1960	1950-1990	1980-2020
Employment Growth (ann, %)				
Realized: Low-skill - Middle-skill	1.7	1.0	1.3	2.5
Predicted: Low-skill - Middle-skill	0.8	0.4	0.7	1.6
% of Gap Explained	49.6	35.1	58.0	63.2
Employment Growth (ann, %)				
Realized: High-skill - Middle-skill	2.1	1.7	2.5	2.1
Predicted: High-skill - Middle-skill	0.9	0.5	1.1	1.4
% of Gap Explained	43.3	27.2	44.4	65.8
C. Occ. Female Employment Share	**Full Sample**	**Sub-samples**		
	1910-2020	1910-1960	1950-1990	1980-2020
Employment Growth (ann, %)				
Realized : High - Low	1.6	1.6	2.4	0.8
Predicted: High - Low	0.8	0.5	0.7	1.2
% of Gap Explained	47.8	31.1	29.9	145.2

Note: In panel A of this table we compare the average annualized actual and technology-predicted relative employment growth of occupations in the top employment-weighted quintile of educational attainment to the bottom quintile. In panel B, we compare the average actual and technology-predicted growth of middle-skill occupations relative to either low-skill services or high-skill occupations, using an occupational categorization inspired by Autor et al. (2024). Finally, in panel C, we compare the relative actual and technology-predicted employment growth of occupations in the top employment-weighted quintile to those in the bottom quintile based off the share of workers who are female. Predicted employment growth comes from estimates of A.3. See section 3.3 of the main text for more details.

Table 4: Technology Exposure and Employment Growth, Subsamples, Heterogeneous Effects by Task Exposure

	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Full Sample (1910–2020)								
Mean Task Exposure, Manual	-2.15***	-4.70***	-2.92***	-6.33***	-2.29***	-5.00***	-2.85***	-6.30***
	(0.26)	(0.51)	(0.46)	(0.82)	(0.26)	(0.51)	(0.49)	(0.86)
Mean Task Exposure, Cognitive	0.74	1.42	-0.59	-1.50	0.58	0.85	-0.37	-1.47
	(0.69)	(1.38)	(1.01)	(1.81)	(0.67)	(1.30)	(1.06)	(1.91)
Mean Task Exposure, Inter	-0.41	-2.53	-1.20	-4.20	-0.59	-2.86	-1.17	-4.03
	(2.22)	(4.68)	(2.23)	(4.80)	(2.26)	(4.59)	(2.26)	(4.71)
Concentration in Task Exposure			10.3**	22.5**			7.57	18.2*
			(5.25)	(9.09)			(5.67)	(9.89)
Obs	2,436	2,375	2,436	2,375	2,436	2,375	2,436	2,375
B. 1910–1970								
Mean Task Exposure, Manual	-1.77***	-4.29***	-2.46***	-6.07***	-1.94***	-4.54***	-2.55***	-5.84***
	(0.40)	(0.75)	(0.74)	(1.18)	(0.41)	(0.77)	(0.83)	(1.30)
Mean Task Exposure, Cognitive	5.29***	7.32***	3.97**	3.70	5.08***	6.64***	3.94*	4.06
	(1.50)	(2.64)	(2.00)	(3.03)	(1.34)	(2.29)	(2.01)	(2.95)
Mean Task Exposure, Inter	-1.07	-6.62	-1.48	-7.44	-1.45	-6.09	-1.87	-6.78
	(2.75)	(7.33)	(2.77)	(7.15)	(2.82)	(7.05)	(2.88)	(6.95)
Concentration in Task Exposure			9.14	23.8*			8.28	17.8
			(9.16)	(13.12)			(10.83)	(15.76)
Obs	1,116	1,153	1,116	1,153	1,116	1,153	1,116	1,153
C. 1960–2020								
Mean Task Exposure, Manual	-2.57***	-5.07***	-3.09***	-5.85***	-2.78***	-5.36***	-3.26***	-6.36***
	(0.33)	(0.63)	(0.54)	(0.98)	(0.33)	(0.62)	(0.57)	(1.07)
Mean Task Exposure, Cognitive	-1.31**	-2.11*	-2.14**	-3.35	-1.52**	-2.49*	-2.31**	-4.13*
	(0.63)	(1.27)	(1.06)	(2.20)	(0.65)	(1.29)	(1.15)	(2.42)
Mean Task Exposure, Inter	0.69	4.29	-0.33	2.68	0.12	2.23	-0.62	0.69
	(4.17)	(6.88)	(4.00)	(7.21)	(4.24)	(7.29)	(4.08)	(7.73)
Concentration in Task Exposure			7.44	10.9			6.93	14.0
			(5.38)	(10.23)			(5.91)	(11.89)
Obs	1,320	1,222	1,320	1,222	1,320	1,222	1,320	1,222
Year FE	X	X	X	X	X	X	X	X
Employment Share, Lag	X	X	X	X	X	X	X	X

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,T+H}}{N_{o,T}}\right) = \sum_{\tau} \beta_\tau \left[\text{Mean Exposure}^{H,\tau}_{o,T} \times \text{Task Share}^{\tau}_{o,T}\right] + c\,\mathbf{\Gamma}_{o,t} + \varepsilon_{o,T}, \quad \tau \in \{M, C, I\}.$$

Panel A, B and C estimates coefficients using T spanning 1910–2000, 1910–1950 and 1960–2000 (all ranges inclusive), respectively. The variables of interest are Mean $\text{Exposure}^{H,\tau}_{o,T}$ ($\tau \in \{M, C, I\}$), technology mean exposure for different task types. Controls $\mathbf{\Gamma}_{o,t}$ include year fixed effects and lagged employment share $N_{o,T}$. Columns 1–4 show OLS regressions, while the others show IV regressions. Columns 1, 2, 5, and 6 show regressions with $H = 10$, while the others show regressions with $H = 20$.

Table 5: Technology Exposure and Employment Growth, Age and Cohort Effects

Employment Growth, by Age and Cohort	10-yr Horizon				20-yr Horizon			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Entrants × Mean Task Exposure	-9.05***	-15.3***	-9.95***	-14.7***	-18.4***	-32.4***	-20.8***	-32.6***
	(1.37)	(2.58)	(1.39)	(2.73)	(2.39)	(4.10)	(2.40)	(4.28)
Entrants × Concentration in Task Exposure		7.40**		5.77*		16.8***		14.6***
		(2.67)		(2.88)		(3.92)		(4.28)
16-25 × Mean Task Exposure	-2.35	-14.6***	-5.18*	-16.1***	-8.76**	-26.6***	-13.4***	-29.3***
	(2.23)	(3.98)	(2.26)	(4.20)	(3.09)	(5.31)	(3.09)	(5.50)
16-25 × Concentration in Task Exposure		14.6***		13.2**		21.4***		19.6***
		(4.05)		(4.35)		(5.43)		(5.62)
26-45 × Mean Task Exposure	-6.70***	-12.4***	-7.28***	-11.9***	-15.4***	-26.2***	-17.0***	-25.7***
	(1.22)	(2.35)	(1.23)	(2.42)	(2.11)	(3.65)	(2.10)	(3.76)
26-45 × Concentration in Task Exposure		6.79**		5.56*		12.9***		10.8**
		(2.25)		(2.39)		(3.66)		(3.84)
46-65 × Mean Task Exposure	-11.4***	-17.3***	-12.1***	-16.9***	-34.7***	-52.5***	-36.1***	-50.3***
	(1.13)	(2.05)	(1.16)	(2.12)	(2.49)	(4.08)	(2.48)	(4.17)
46-65 × Concentration in Task Exposure		6.99***		5.85**		21.2***		17.4***
		(1.94)		(2.09)		(3.94)		(4.13)
Estimator	OLS	OLS	IV	IV	OLS	OLS	IV	IV
N	9,680	9,680	9,680	9,680	9,407	9,407	9,407	9,407
Cohort-Year FE	X	X	X	X	X	X	X	X
Cohort × Employment Share, Lag	X	X	X	X	X	X	X	X
Cohort × Avg Occ Age, Lag	X	X	X	X	X	X	X	X
P(Mid = Young, Mean)	0.009	0.403	0.213	0.127	0.001	0.898	0.075	0.292
P(Mid = Young, Concentration)		0.002		0.005		0.006		0.004
P(Old = Young, Mean)	0.000	0.415	0.000	0.812	0.000	0.000	0.000	0.000
P(Old = Young, Concentration)		0.019		0.031		0.972		0.639

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,age\in[a+H,b+H],T+H}}{N_{o,age\in[a,b],T}}\right) = \beta_{\text{a-b}}\,\text{Mean Exposure}^H_{o,T} + \gamma_{\text{a-b}}\,\text{Exposure Concentration}^H_{o,T} + c_{\text{a-b}}\,\mathbf{\Gamma}_{o,T} + \varepsilon_{o,a-b,T}$$

for $[a,b] \in \{[16,25],[26,45],[46,65]\}$ (row 3–8), with the exception of row 1–2 (entrants), where the dependent variable is $\log\left(\frac{N_{o,age\in[16,25],T+H}}{N_{o,age\in[16,25],T}}\right)$. Controls $\mathbf{\Gamma}_{o,t}$ include year fixed effects, lagged employment share, and lagged average workers age within occupations. The coefficients are estimated separately for each cohorts. Coefficients correspond to a one-standard-deviation change in the independent variable and are multiplied by 100. Standard errors (in parentheses) are clustered by occupation. The final four rows of the table display p-values on various tests for coefficient equality.

Table 6: Model-implied impact of AI on labor composition

	Model Predictions (annualized percent rates)			
	Baseline	No Product Innovation	Eloundou et al. Exposures	Industry Prod. Spillovers ($\theta = 1.72 > \chi = 1.34$)
	(1)	(2)	(3)	(4)
A. Occupation Education Level				
High - Low	-0.65	-0.71	-1.21	-0.63
B. Occupation Wages				
Managers - Middle-skill	-0.59	-0.70	-1.05	-0.59
Sales/Professionals - Middle-skill	-0.29	-0.42	-0.82	-0.28
Clerical/Technicians - Middle-skill	-0.85	-0.95	-1.12	-0.85
C. Occupation Female Employment Share				
High - Low	-0.53	-0.57	-0.56	-0.54

Note: This table examines aggregate outcomes implied by our model simulation of the impact of artificial intelligence. In the first column we impose our baseline assumption that $\theta = \chi$ and allow for AI-related product innovation. In the succeeding columns we differ from the baseline by dropping product innovation; replacing our baseline assumption that cognitive tasks are exposed to AI with the Eloundou et al. (2023) measure of exposure to generative AI; allowing for additional technology spillovers by calibrating $\theta > \chi$ using an estimate for θ taken from Kogan et al. (2023). Growth rates are in annualized terms and in percent rates. We assume AI capital prices follow the same average decline as those observed for user cost of computer hardware (13.96% per year) over the 1984-2015 period, as calculated by Caunedo et al. (2023). See section 4 of the main text and the appendix for further details.

Appendix

A Model Appendix

Section A.1 contains the key equilibrium conditions characterizing the solution of our model. We then discuss how we compute the solution of our quantitative exercises related to AI in section A.2

A.1 Key Model Equations

Here, we reproduce the key equations which are necessary for solving for the equilibrium of our model. For further details, including detailed derivations, we refer the model to Hampole et al. (2025). Our model is essentially nested in theirs with two minor changes. First, we assume perfect competition in labor and product markets, which implies that the terms reflecting markups and monopsony wedges (Θ and $\mathcal{M}(j)$ in Hampole et al. (2025)) both equal one and can be omitted from the labor demand equation. We now proceed to reproduce the main equations of the model.

The industry level CES:

$$\bar{Y} = \left(\int_{\mathcal{I}} \alpha_I^{\frac{1}{\theta}} Y_I^{\frac{\theta-1}{\theta}} dI \right)^{\frac{\theta}{\theta-1}}. \tag{A.1}$$

The occupation level CES:

$$Y_I = \left(\int_{\mathcal{O}} \alpha_I(o)^{\frac{1}{\chi}} Y(o,I)^{\frac{\chi-1}{\chi}} \right)^{\frac{\chi}{\chi-1}}. \tag{A.2}$$

The task level CES:

$$Y(o,I) = \left(\sum_j \alpha_o(j)^{\frac{1}{\psi}} y(j)^{\frac{\psi-1}{\psi}} \right)^{\frac{\psi}{\psi-1}}, \tag{A.3}$$

The labor/capital level CES:

$$y(j) = \left(\gamma_j \, l(j)^{\frac{\nu-1}{\nu}} + (1-\gamma_j) \, k(j)^{\frac{\nu-1}{\nu}} \right)^{\frac{\nu}{\nu-1}}. \tag{A.4}$$

Effective Labor Supply:

$$\ell(j) = \alpha(j)^{\beta} h(j)^{1-\beta}. \tag{A.5}$$

Optimal Hours Allocation

$$h(j) = \frac{\alpha(j) w(j)^{\frac{1}{\beta}}}{\sum_{k \in J} \alpha(k) w(k)^{\frac{1}{\beta}}}. \tag{A.6}$$

Occupation Level Efficiency

$$X(o,I) = \left[\sum_{j \in J} \alpha(j) p(j)^{1-\psi} \right]^{-\frac{1}{1-\psi}} = P(o,I)^{-1} \tag{A.7}$$

Firm Level Efficiency

$$Z_I \equiv \left(\int_{\mathcal{O}} \alpha_I(o) P(o,I)^{1-\chi}\right)^{-\frac{1}{1-\chi}} = P_I^{-1} \tag{A.8}$$

Firm Inverse Demand Curve

$$Y_I = \alpha_I Z_I^{\theta}\, \bar{Y} \tag{A.9}$$

Marginal Cost of task

$$p(j) = \left(a_j\, w(j)^{1-\nu} + b_j\; q(j)^{1-\nu}\right)^{\frac{1}{1-\nu}} \tag{A.10}$$

Demand for occupation o output

$$Y(o,I) = \alpha_I(o) P(o,I)^{-\chi}\, Z_I^{-\chi}\, Y_I \tag{A.11}$$

Demand for task j output:

$$y(j) = \alpha_I \alpha_I(o) \alpha(j) p(j)^{-\psi} X(o,I)^{\chi-\psi} Z_I^{\theta-\chi} \bar{Y}. \tag{A.12}$$

Demand for task j labor

$$l(j) = \alpha_I \alpha_I(o) \alpha(j) \frac{a_j}{w(j)^{\nu}} \left(a_j w(j)^{1-\nu} + b_j\, q(j)^{1-\nu}\right)^{\frac{\nu-\psi}{1-\nu}} \times X(o,I)^{\chi-\psi}\, Z_I^{\theta-\chi} \bar{Y}. \tag{A.13}$$

Demand for task j capital

$$k(j) = \alpha_I \alpha_I(o) \alpha(j) \frac{b_j}{q(j)^{\nu}} \left(a_j w(j)^{1-\nu} + b_j\, q(j)^{1-\nu}\right)^{\frac{\nu-\psi}{1-\nu}} \times X(o,I)^{\chi-\psi}\, Z_I^{\theta-\chi} \bar{Y}. \tag{A.14}$$

Labor Supply for task j

$$L_o(j) = N(o,I)\, \alpha(j)^{\beta} h(j)^{1-\beta} = N(o,I)\, \alpha(j) w(j)^{\frac{1-\beta}{\beta}} \left(\sum_{k \in J} \alpha(k) w(k)^{\frac{1}{\beta}}\right)^{\beta-1} \tag{A.15}$$

Worker supply for occupation o

$$N(o,I) = \alpha_I \alpha_I(o) W(o,I)^{\zeta} = \alpha_I\, \alpha_I(o) \left(\frac{W(o,I)}{\bar{W}}\right)^{\zeta}. \tag{A.16}$$

where

$$\bar{W} = \left[\int \int \alpha_I\, \alpha_I(o) W(o,I)^{\zeta+1} dI do\right]^{\frac{1}{\zeta+1}}. \tag{A.17}$$

Which depends on total wages for occupation o

$$W(o,I) \equiv \sum_{j \in J_o} \alpha(j)^{\beta} h(j)^{1-\beta} w(j) = \left(\sum_{j \in J} \alpha(j)\, w(j)^{\frac{1}{\beta}}\right)^{\beta} \tag{A.18}$$

Labor market clearing equation: labor supply equal to labor demand.[8]

$$w(j)^{\frac{1}{\beta}} \left(\sum_{j \in J_o} \alpha(j) w(j)^{\frac{1}{\beta}} \right)^{\beta - 1 + \zeta \beta} \bar{W}^{-\zeta} \varphi = a_j \, w(j)^{1-\nu} \left(a_j w(j)^{1-\nu} + b_j \, q(j)^{1-\nu} \right)^{\frac{\nu - \psi}{1-\nu}} X(o, I)^{\chi - \psi} Z_I^{\theta - \chi} \bar{Y}. \tag{A.19}$$

A.2 AI counterfactual calculations

To compute the implied changes and employment and wages associated with our AI counterfactual, we use the following procedure. We assume that the economy begins in a symmetric steady state in which $w(j)$, $q(j)$, and γ_j are the same across all tasks. In this case, we can then pin down all alpha parameters using the observed allocation of labor across industries and occupations, with the total amount of labor normalized to 1. Without loss of generality, we separately identify γ_j from $q(j)$ by normalizing that $w(j) = q(j)$, then picking the common task price which is consistent with a marginal cost of per-task output of 1 and our target for the capital share. We then pin down φ to match the level of aggregate labor supply associated with this initial allocation.

Next, to solve for the new steady state, we look for a growth rate of aggregate wages and output which satisfies the equilibrium conditions specified above. Rather than rely on the loglinear approximation presented in the text, we solve the equilibrium conditions exactly. This process is greatly facilitated by the fact that equation (A.19) is the same across all tasks that have the same $\varepsilon(j)$ within the same occupation-industry cell.

B Data Construction

B.1 Obtaining technology descriptions from patent data

B.1.1 Patent data

Our patent data comes from two sources. The pre-1976 patents (1836 - 1975) are scraped from Google patents. Specifically, from year 1836 to 1975 (including), there are in total 3,924,149 patents (according to PatentMetrics). Post-1976 patents are downloaded directly from PatentsView. PatentsView contains the full text of all patent documents from 1976 to 2024. There are in total 8,206,179 patents that have a numeric patent number (we focus on these patents). We construct patent text data by combining the patent title, abstract, brief summary, detailed description, and claim (use the order as described).

[8]Here φ is an aggregate labor supply normalizing constant which plays no role except to ensure that initial task wages are constant and equal to capital prices in symmetric initial equilibrium for the model simulation, which we discuss in appendix D.

B.1.2 Extracting technology descriptions from patent documents

Our goal is to use the textual description of these innovations to relate them to the tasks that workers do. Unfortunately, however, the quality of the extracted text of patent documents varies considerably across years. Therefore, we use a modern large language model (LLM) to obtain a concise description of the key new innovation from each patent document. In particular, we feed each patent document into *llama-3.1-8B-Instruct*, along with the following prompt:

> Here is a description of an invention from a patent issued in [PATENT ISSUE YEAR HERE]. Summarize the following: 1. what the invention does; 2. what is the main innovation; 3. how it can be used in production at that time; 4. which industry or industries are most likely to make use of this invention. Give detailed answers to each question, but each answer should be no more than four sentences long. Please avoid extrapolating too much. Put your answer to the four questions within a tuple as (1 &&&&& [ANSWER 1] 2 &&&&& [ANSWER 2] 3 &&&&& [ANSWER 3] 4 &&&&& [ANSWER 4]). In other words, use &&&&& to split answers to different questions, and replace [ANSWER #] with the appropriate answer to each question. All output should be contained within the tuple (do not output any other things). [PATENT TEXT HERE].

The large language model we use has a maximum input capacity of approximately 131,000 tokens. For patents exceeding this limit, we truncate the content and retain only the first 100,000 tokens. These longer patent texts are more common in patents issued after 2000. If the LLM returns an empty summary, we use *llama-3.1-70B-Instruct* to re-summarize those problematic summaries after the first round. We are able to do so for all except for 3,141 out of all 3,924,149 pre-1976 patents. Among them, there are 1215 patents that have empty patent text that can not be scrapped, 3 patents that have 404 website error. Among all 8,206,179 post-1976 patents, we are able to extract valid summary for all except for 30,477 (0.37%) patents.

We then obtain a vector representation of these summaries using *openai: text-embeddings-3-small* model. Thus, each patent can be represented as a vector of length 1,536. Note that summaries include the brief summary of invention ([Answer 1] in the query), the main innovation of the patent ([Answer 2] in the query), the production use of the patent ([Answer 3] in the query). We use the industry of use of the patent ([Answer 4] in the query) to assign industry code.

B.1.3 Patents to Industries

We use the same LLM model (*llama-3.1-8B-Instruct*) to obtain a coarse measure of which industries are particularly likely to see productivity improvements from technology p. As we note above, our LLM prompt includes a request for providing a likely industry of use. We use this to obtain a coarse measure of which industries are particularly likely to see productivity improvements from technology p. We obtain the vector representation of the industry-of-use descriptions using the same embedding model and compute the cosine similarities with the textual descriptions of the ind1950 industries from Census.

We assign a patent as relevant to a given Census industry if the Census industry title is the most textually similar to the patent's industry of use description (based on cosine similarity of the embeddings generated for the patent industry of use and Census industry title); we further assign the patent to the industry if the textual similarity is in the top 1% of all industry-patent similarity pairs for patents issued in that decade. In particular, for each calendar year, we determine a similarity threshold by taking the 99th percentile of scores between 5,000 randomly sampled patents' embeddings and all ind1950 entries (or, if fewer than 5,000 patents exist in a given year, the entire patent set). We then calculate the pairwise similarity between every patent embedding and each industry embedding: an industry is assigned to a patent if it is either the most similar industry overall or if its similarity exceeds the year-specific 99th-percentile threshold. Because the Census "ind1950" classification is available for all Census years in our sample, we rely on this industry classification scheme throughout.

B.1.4 Clustering Patents

To categorize technological innovations over time, we cluster patents into 20 distinct categories each time period using vector representations derived from the patent summaries. Recognizing that the nature of technological advancements evolves across historical periods, we split the sample into three eras: 1850–1920, 1920–1980, and 1980–present. For each period, we apply the mini-batch K-means algorithm to cluster all patents within that timeframe. To assign meaningful labels to each cluster, we sample 500 patents per cluster and use GPT-4o to generate descriptive group names based on the sampled content.

B.2 Tasks

B.2.1 Creating Time-varying Task Data

Next, we obtain a time varying description of the key tasks performed by each occupation in each decade in our sample. In particular, for each Census occupation in each decade from 1850 to 2010, we submit the following query to *gpt-4o-search-preview*

> You are a labor economist and an expert on economic history studying the evolution of work. Describe the list of tasks that the occupation [occupation title] would perform in their day to day job in [year]. Be very careful to make your answer specific to that time period. Your answer should be a list of tasks in JSON format, and should be written in the style of occupation task descriptions in the O*NET database. Please be sure to avoid duplicating the same task within your list; each task should be conceptually distinct. At the end please briefly provide a comprehensive and highly specific list of your sources as well. [Formatting instructions follow]

We constrain the search context to the United States and configure the search context size to its maximum setting. For each decade–occupation pair, the LLM generates a set of task titles and corresponding descriptions. We concatenate each title with its description to form the final task text representation. The result of this query is a combination of 84,393 task–occupation–decades, spanning the years 1850 to 2010 (in 10 year intervals, excluding year 1890 as the Census microdata for that year are not available). On average, the search query returns approximately 14 tasks per occupation in each decade. We then represent these tasks as numerical vectors using the *openai: text-embeddings-3-small* embeddings from OpenAI. For robustness, we perform this exercise using two alternative LLMs: gpt-4o-2024-11-20 (referred to as gpt-4o) and Meta-Llama-3.1-405B-Instruct (Llama). Our baseline is gpt-4o-search-preview, due to its integrated web search capabilities.

These task descriptions obtained from the LLM will play a key role in our empirical analysis. To ensure that these descriptions are indeed reliable, we perform a validation exercise using the occupation task descriptions from the Dictionary of Occupational Titles (DOT) and O*NET. In particular, we compare the semantic meaning of all these task descriptions between the LLM and the DOT or O*NET counterparts, whenever these are available. Thus, for each occupation in the 1940, 1980, and 2010 decades, we compute the average task embedding of the LLM task descriptions and compute its cosine similarity with the average task embedding of its DOT or O*NET counterpart. We then compare the distribution of these similarity scores to a placebo distribution where we randomly compare the LLM tasks with a different occupation in the DOT or O*NET. As we see in Appendix Figure A.1, the task descriptions we obtain from the LLM are fairly similar to their DOT or O*NET counterparts from the same period, with the within-occupation cosine similarity between LLM tasks and DOT/O*NET having a distribution that is substantially shifted to the right relative to the cross-occupation placebo. Interestingly enough, the LLM tasks from the 2010 period are particularly close to the O*NET tasks, which suggests that these tasks were part of the LLM's search query.

B.2.2 Task Classification - Manual, Cognitive, Interpersonal

We further use the same LLM we used to extract our occupation task description to classify tasks into three categories: manual, cognitive, and interpersonal. To do so, we first obtain a complete list of broad work activities from O*NET (edition 28.3). O*NET work activities are a less specific categorization of the content of work than the detailed occupation tasks. There are 41 distinct work activities, which we partition into manual, cognitive, and interpersonal categories. Our approach is similar to Acemoglu and Autor (2011), who also inspect and group O*NET work activities into different task types in order to broadly understand the task content of occupations. We list these work activities falling under a given category to give textual context for the large language model when classifying a given task. We submit the following prompt to *gpt-4o-search-preview*

> You are an expert labor economist. Your job is to read descriptions of a task performed by labor and classify whether the task is most likely a "manual task", "interpersonal task" or "cognitive task". A manual task would be expected to emphasize at least some of the following characteristics: Performing General Physical Activities; Handling and Moving Objects; Controlling Machines and Processes; Operating Vehicles, Mechanized Devices, or Equipment; Repairing and Maintaining Mechanical Equipment; Repairing and Maintaining Electronic Equipment; Inspecting Equipment, Structures, or Materials; Drafting, Laying Out, and Specifying Technical Devices, Parts, and Equipment. A interpersonal task would be expected to emphasize at least some of the following characteristics: Communicating with Supervisors, Peers, or Subordinates; Communicating with People Outside the Organization; Establishing and Maintaining Interpersonal Relationships; Assisting and Caring for Others; Selling or Influencing Others; Resolving Conflicts and Negotiating with Others; Performing for or Working Directly with the Public; Coordinating the Work and Activities of Others; Developing and Building Teams; Training and Teaching Others; Guiding, Directing, and Motivating Subordinates; Coaching and Developing Others; Providing Consultation and Advice to Others; Staffing Organizational Units. A cognitive task would be expected to emphasize at least some of the following characteristics: Getting Information; Monitoring Processes, Materials, or Surroundings; Identifying Objects, Actions, and Events; Estimating the Quantifiable Characteristics of Products, Events, or Information; Judging the Qualities of Objects, Services, or People; Processing Information; Evaluating Information to Determine Compliance with Standards; Analyzing Data or Information; Making Decisions and Solving Problems; Thinking Creatively; Updating and Using Relevant Knowledge; Developing Objectives and Strategies; Scheduling Work and Activities; Organizing, Planning, and Prioritizing Work; Documenting/Recording Information; Interpreting the Meaning of Information for Others; Working with Computers; Performing Administrative Activities; Monitoring and Controlling Resources. #################### Here is the task text: [TASK TEXT HERE] Based on the above definition, would this be considered a manual task, interpersonal task, or cognitive task? Give a manual/interpersonal/cognitive/uncertain answer, along with a brief (no more than one sentence) explanation. Your answer should be in the form of a tuple (answer, explanation). Only include this tuple in your response.

Among all the 84,393 task–occupation–year triplets, 45% are classified as manual, 34% as cognitive, and 21% as interpersonal.

B.2.3 Task Classification - Specific Vocational Preparation

For our AI simulations in section 4 of the main text, we assume that cognitive tasks are exposed to AI. We classify cognitive tasks in O*NET using the exact prompt listed above. However, we also assume that high-expertise cognitive tasks are not exposed to AI. We measure expertise by mapping it the concept of “Specific Vocational Preparation" as defined by O*NET (see `https://www.onetonline.org/help/online/svp`.). In particular, we use a prompt originally constructed by Kogan et al. (2023)–who demonstrate that the categories generated from this prompt are strongly predictive of an occupation's O*NET's job zone category, which captures occupational skill and training required. We send the following prompt to the OpenAI GPT-4o Search model:

> Specific Vocational Preparation is the amount of lapsed time required by a typical worker to learn the techniques, acquire the information, and develop the facility needed for average performance in a specific job-worker situation.Tell me whether attaining proficiency in the below occupation task requires A) an extensive amount (more than 5 years); B) a fair amount (1 to 5 years); C) a moderate amount (3 months to 1 year); or D) very little (less than 3 months) of specific vocational preparation; and, explain your reasoning in one sentence. You must format the answer in a tuple '(A/B/C/D, one-sentence reasoning)'. The occupation task is: < Insert Task Description >

We classify 4.9% of O*NET tasks into the more than 5 year category; 32.4% into the 1 to 5 year category; 39.3% into 3 months to 1 year; and 23.4% to fewer than 3 months.

B.3 Employment

We obtain employment counts using the IPUMS Census extracts.

B.3.1 IPUMS Population Sample

We use the full population sample for the period 1850–1940, and use the 1% sample for 1950, 5% sample for 1960, 1% state Form 1 sample for 1970, 5% state sample for 1980 and 1990. Due to the smaller sample sizes for the ACS, in 2010, we use 2008-2012 ACS 5-year sample and take all data for years from 2018-2012, as an average centered at year 2010. Similarly for 2020, we use 2018-2022 ACS 5-year sample, as an average centered at year 2020.

We use USA IPUMS census data to extract information related to population, including gender, age, occupation, industry, labor force participation, education, weeks/hours worked, and income:

Decade(s)	Data Source and Sample Description
1850–1940	Full population sample
1950	1% sample
1960	5% sample
1970	1% state Form 1 sample
1980	5% state sample
1990	5% state sample
2010	2012 ACS 5-year sample
2020	2022 ACS 5-year sample

We restrict the sample to census respondents aged 15 to 75 that report that they are employed. (census variable labforce = 2 for years 1850–1920, and empstat = 1 for other years). We exclude members of the armed forces and occupation codes indicating a non-occupational response, such as helping at home (occ1950 codes 975–999). Following Katz and Margo (2014), we include both men and women in our analysis; however, we also investigate the robustness of our findings to restricting to men. We compute employment in a specific occupation (or industry, i.e. ind1950 or in2d, when available) using the Census responded weights (census variable perwt).

B.3.2 Consistent Occupation Codes

To compute employment growth across decades, we aggregate the decade–specific definition of census occupations into time-consistent classifications between the two start- and end-of-period Census years involved in the employment change. For occupation classifications, we rely on the 1950 Census occupation coding scheme (IPUMs variable “occ1950") for employment change observations with start-of-period Census years before 1980. We use a revision of the Census 1990 scheme created by Autor et al. (2024) (called “occ1990dd_18") for employment change observations with start-of-period years 1980 and later. Because there is less resolution of “occ1950" codes available in IPUMs data in the year 1940 or post-1980, we convert 1950 occupations into a collapsed scheme (also introduced by Autor et al. (2024) and called “occ1950rj") for any employment changes where the start-of-period or end-of-period includes 1940, or if the end-of-period of the employment change is post-1980.

The detailed table with our time-consistent occupation coding scheme is below:

Employment StartYear	Employment EndYear	OccScheme
1850	1860	occ1950
1850	1870	occ1950
1860	1870	occ1950
1860	1880	occ1950
1870	1880	occ1950
1880	1900	occ1950
1900	1910	occ1950
1900	1920	occ1950
1910	1920	occ1950
1910	1930	occ1950
1920	1930	occ1950
1920	1940	occ1950
1930	1940	occ1950rj
1930	1950	occ1950
1940	1950	occ1950rj
1940	1960	occ1950rj
1950	1960	occ1950
1950	1970	occ1950
1960	1970	occ1950
1960	1980	occ1950rj
1970	1980	occ1950rj
1970	1990	occ1950rj
1980	1990	occ1990
1980	2000	occ1990
1990	2000	occ1990
1990	2010	occ1990
2000	2010	occ1990
2000	2020	occ1990

We use matching weights $w_{o' \to o}$ to construct a weighted crosswalk from the year-specific occupation code o' to the time-consistent code o we use above. We take $w_{o' \to o}$ to be the share of individuals in a time-consistent occupation o who are also assigned to the year-specific occupation o' in the given Census decade based off Census employment counts. As we explain in the next section, we use these to constructed weighted versions of time-consistent occupational technology exposure mean and concentration across tasks.

C Measurement

Here, we discuss our measurement approach in detail.

C.1 Task exposure

Let T denote a decade period (e.g. 1940-49, 1950-59); j, an occupation task; p, a patent, and $\text{similarity}_{p,j}$ the cosine similarity between the OpenAI text embeddings for patent p and task j. Our goal is to determine whether a given patent exposes a particular task. We impose sparsity in this relationship, by assuming that only pairs sufficiently in the right tail of the distribution of $\text{similarity}_{p,j}$ are related. For our baseline we impose a 95th percentile cutoff. Denote this 95th percentile cutoff in the distribution of task-patent similarity scores as $P95$.

Due to trends in language, we de-mean similarity scores by year and compute percentile cutoffs separately for the tasks in each Census decade. Specifically, for a task j which is relevant to a Census occupation in decade T, we compute the $P95$ percentile cutoff of the similarity between all patents issued in the 20 years before and after the start of Census decade T and decade T tasks. For example, for tasks in Census year 1940, the $P95$ threshold is calculated from the similarity score distribution between all patents issued in 1920 1959 and all decade 1940 tasks.

Since it will be computationally costly to get the percentile, we take a 1/10000 sample from all patent - task pairs except for decade 2000. For decade 2000, there are around 10,000 tasks and 3 million patents (1980 - 2019) in the window, which forms a similarity sequence of length 30 billion. We take a 1/1000000 sample from it to compute $P95$.

We compute two versions of technological exposure of task j in decade T, mapped to either a 20-year or 10-year forward-looking horizon ($H = 10$ or $H = 20$). Let P_H denote the set of patents issued in the time window $[T, T+H)$ (so when $T = 1940$, P_{10} includes patents issued from 1940-1949, inclusive, and P_{20} includes patents issued from 1940-1959, inclusive) and N_H the number of patents in P_H. The exposure of the task j to technology is simply the probability it is relevant to patents issued over that horizon:

$$\text{Exposure}^{H}_{j,T} = \frac{1}{N_H} \sum_{p \in P_H} \mathbf{1}(\text{similarity}_{p,j} > p95) \tag{A.20}$$

C.2 Constructing Occupation-Level Measures of Exposure

The model labor demand equation (10) implies that the means and variances of task-level exposure are a sufficient statistic for occupational employment growth. We therefore take the means and variances (which is our measure of exposure concentration) of task-level exposure across all tasks j which are relevant to Census occupation o. This yields our measures of occupational average task exposure to technology, and Exposure Concentration in occupational exposure to technology. Specifically, let $J(o,T)$ denote the set of tasks j relevant to occupation o in Census decade beginning in T. We have

$$\text{Mean Exposure}^{H}_{o,T} = \frac{1}{|J(o,T)|} \sum_{j \in J(o,T)} \text{Exposure}^{\text{H}}_{j,T} \tag{A.21}$$

and

$$\text{Exposure Concentration}^{H}_{o,T} = \frac{1}{|J(o,T)|} \sum_{j \in J(o,T)} \left(\text{Exposure}^{\text{H}}_{j,T} - \text{Mean Exposure}^{H}_{o,T}\right)^2 \tag{A.22}$$

for $H = 10, 20$.

To use the measure to examine employment across time, as discussed in Section B.3.2, we map year-specific occupation-level exposure to time-consistent occupation scheme using (**??**).

There is one nuance to consider for occupation-level exposures. While we observe and predict employment-growth using the cross-time consistent schemes discussed previously in the appendix, we generate task exposures for year-specific titles corresponding to time-varying Census occupation coding schemes. In practice, the year-specific Census occupation codes for which we construct LLM-generated occupational tasks may map to more than one of the time-consistent codes in the list of time-consistent occupation codes used for each year. In such a case, we weight tasks for a year-specific code according to its employment weight within the time-consistent code in that year: if a year-specific occupation o' has 5 tasks in decade T, and it makes up 10% of overall employment in a time-consistent occupation o in decade T, then a single task j in occupation o' will receive a $10\% \times 1/5 = 0.5\%$ weight in computing occupation o overall mean exposure in decade T. Similarly, the same weight is applied when calculating exposure concentration across all tasks.

More formally, if a time-specific Census occupation o' has $J_{o'}$ tasks, in a given decade, and a share $w_{o' \to o}$ in time-consistent code o come from year-specific census code o', then each task j in occupation o' receives weight $\omega_j \equiv (1/J_{o'}) \times w_{o' \to o}$ in computing the occupation mean and concentration in equations (A.21) and (A.22). That is, in practice we calculate

$$\text{Mean Exposure}^{H}_{o,T} = \sum_{j \in J(o,T)} \omega_j \times \text{Exposure}^{\text{H}}_{j,T} \tag{A.23}$$

and

$$\text{Exposure Concentration}^{H}_{o,T} = \sum_{j \in J(o,T)} \omega_j \times \left(\text{Exposure}^{\text{H}}_{j,T} - \text{Mean Exposure}^{H}_{o,T}\right)^2 . \tag{A.24}$$

C.3 Task Type Decomposition of Occupational Exposure - Manual, Cognitive, and Interpersonal

As discussed above in Section B.2.2, we classify each task to one of the three types: manual, cognitive and interpersonal. We use these categories to compute the technology exposure to the manual, cognitive or the interpersonal part of the occupation. The measure construction is almost the same as above, with the difference of using only tasks that are of speicific type for any occupation. Specifically, denote task type $\tau \in \{\text{manual}, \text{cognitive}, \text{interpersonal}\}$, Mean Exposure$^{\tau,H}_{o,T}$ the type τ exposure of occupation o in decade T, for H horizon. We use $J(o, T, C)$ for tasks relevant to occupation o in decade T which is classified as type τ.

$$\text{Mean Exposure}^{\tau,H}_{o,T} = \frac{1}{|J(o,T,\tau)|} \sum_{j \in J(o,T,\tau)} \text{Exposure}^{\text{H}}_{j,T} \tag{A.25}$$

We also define the type τ task share of occupation o in decade T, as the share of tasks classified as type τ for all tasks occupation o performs in decade T:

$$\text{Task Share}^{\tau}_{o,T} = \frac{|J(o,T,\tau)|}{|J(o,T)|} \tag{A.26}$$

Using the definition above, we see that the occupational mean exposure is equal to the sum of the exposure of the manual, cognitive and interpersonal mean task exposure, multiplied by their respective task shares:

$$\text{Mean Exposure}^{H}_{o,T} = \sum_{\tau} \text{Task Share}^{\tau}_{o,T} \times \text{Mean Exposure}^{\tau,H}_{o,T} \tag{A.27}$$

C.3.1 Productivity Spillover

We compute the industry spillover measure using the growth rates in the number of patents issued to each industry. The industry assignment for each patent is according to section B.1.3.

Specifically, for H-year employment changes, our industry spillover measure is

$$\text{Spill}^{H}_{I,T} = \log(\text{Matched Patents}_{T \to T+H,I}) - \log(\text{Matched Patents}_{T-H \to T,i}) \tag{A.28}$$

where Matched Patents$_{T \to T+H,i}$ is the number of patents issued within window $[T, T+H)$ that are matched to industry I.

C.4 Constructing a shift-share IV

Our shift-share identification strategy builds on Acemoglu et al. (2016), and leverages how the arrival of breakthrough technologies diffuses to 'downstream' labor-saving technologies. The construction of the shift-share instrument entails two steps. The first step (the shift) involves predicting the arrival of breakthrough patents in a given tech class c at time t based on innovation in other tech classes in the past,

$$\lambda_{c,t,\tau} = \sum_{c' \neq c} \Omega_{c' \to c,t,\tau} \times I_{c',t-\tau}. \tag{A.29}$$

Here, $\Omega_{c' \to c,t,\tau}$ is a technology diffusion matrix constructed based on the textual similarity of patents: its elements are the average similarity of patents in technology class c to patents in technology class c' for patents issued in tech class c at time t and tech class c' at time $t - \tau$; hence τ represents a diffusion lag period for innovation to propagate from tech class c' to c. When constructing Ω, we set its diagonal to zero so that we only use spillovers to class c from other technology classes c'. $I_{c,t}$ gives the intensity of breakthrough patents in class c and year t, as measured by the share of patents that are breakthroughs at time t and in technology class c.

We then use a Poisson model to predict the number of patents issued in year t for tech class c, where the independent variable is the average value of $\lambda_{c,t,\tau}$ across diffusion lags $\tau = 5, \ldots, 20$. We allow the coefficients in the Poisson model to vary by issue decade T; this measure is a strong predictor of subsequent patenting in tech class c, with an average t-stat of 4.2 across the decade-by-decade coefficients. Finally, we aggregate yearly predicted patenting from the Poisson model to the decade level. Finally, we divide by the total patents issued in that decade to get the tech class predicted patenting share in decade T. Call the predicted tech class number of patents $\widehat{N}_{c,T}$ and the predicted share of total patenting $\widehat{S}_{c,T}$.

To construct the second step (the shares) we estimate the likelihood that a patent j from technology class c is related to task j within the window $[T - 10, T - 10 + H)$:

$$\alpha^H_{j,T} = \frac{1}{N^H_{c,T-10}} \sum_{p \in P^H_{c,T-10}} \mathbf{1}(\text{similarity}_{p,j} > P95). \tag{A.30}$$

where $P_{c,T-10}$ is the set of patents issued to tech class c in decade $T - 10$.

Our shift-share measure for the innovation exposure of task j is:

$$Z^H_{j,T} = \sum_c \alpha^H_{j,T} \times \widehat{N}^H_{c,T}, \tag{A.31}$$

Finally, we compute the mean and variance of $Z^H_{j,T}$ across all tasks j which are applicable to occupation o. This yields our two IVs for a 10-year horizon of analysis, which we denote $Z^{H,Mean}_{o,T}$ and $Z^{H,Concentration}_{o,T}$, respectively. The IVs at the 20-year horizon are constructed in the exact same manner, except we replace $\widehat{N}_{c,T}$ with the implied tech-class patenting aggregated over a two-decade forward-looking period instead of a single decade.

C.4.1 Industry Spillover IV

We instrument for industry-level technological spillovers as follows. Let $\Gamma_{I,c,T-H}$ be the probability that a patent in tech class c came from industry I based on our textual mapping of patents to industries in time period T, within time period $[T-H,T)$. We estimate $\Gamma_{I,c,T-H}$ by the number of patents in tech class c, divided by the number of patents in both tech class c and industry I, within time period $[T-H,T)$. We instrument for the predicted matched patents in the industry at the H-year horizon by taking

$$\text{Predicted Matched Patents}_{T\to T+H,I} = \sum_c \Gamma_{I,c,T-H} \times \hat{S}^H_{c,T}. \tag{A.32}$$

Then, following the definition of our industry spillover measure (A.28), we construct the IV for spillover effect at the 20-year horizon by taking the log change:

$$Spill^{IV,H}_{I,T} = \log(\text{Predicted Matched Patents}_{T\to T+H,I}) - \log(\text{Predicted Matched Patents}_{T-H\to T,I}) \tag{A.33}$$

D Model Simulation

Here, we discuss details of the model simulation.

D.1 Simulating new product creation

Our model simulation requires an assumption on the rate of AI-induced new product creation, captured by growth in the model's industry production share parameter α_I. To map this to the data, we first assume that patenting in AI generates new products. We use Pairolero et al. (2025) to label patents as AI-related. We include any patent that Pairolero et al. (2025) tag as related to machine-learning, computer vision, natural language processing, or speech recognition as AI-related, using the strictest 93% cutoff for tagging AI patents (see their paper for details). We then take a probabilistic mapping from patent 3-digit CPC technology classes to NAICS industries created by Goldschlag et al. (2019). We use the Goldschlag et al. (2019) mapping to 6-digit NAICS 2007 industries, and then sum across all 6-digit industries within a 2-digit NAICS sector to get the share of each patent assigned to that sector. We then sum across all patents for each industry and year to get the total number of patents assigned to that industry each year.

Next, we need a mapping of new patents to product creation. For this, we follow Argente et al. (2025), who report that the rate of new product creation within a firm follows the rate of new patent issuance (relative to the stock of existing patents) with an elasticity 0.0469. We create the stock of existing patents following Argente et al. (2025) by accumulating all patents assigned to the industry over the last 20-years, with a depreciation rate of 15%. The Pairolero et al. (2025) patent database coverage extends through the end of 2023, so we then create the rate of new AI patenting for each industry I as of 2023:

$$P_I = \frac{\text{New AI Patents Issued in } 2023_I}{\text{Stock of Existing Patents as of } 2023_I}. \tag{A.34}$$

Finally, since we simulate over a horizon of 10 years, we adjust P_I by calculating the expected total flow of new AI patents from the end of 2024 to 2034. To do this, we calculate the average annual growth rate in total AI patents from 2014 through 2023, which is $g \approx 6.9\%$. Extrapolating this same growth rate going forward, we then compute

$$\text{Predicted flow of New AI Patents from 2024 to } 2034_I = \\ \text{New AI Patents Issued in } 2023_I \times ((1+g)^{T+1} - (1+g))/g,$$

where we use the fact that $\sum_{t=1}^{T} K \times (1+g)^t = K \times ((1+g)^{T+1} - (1+g))/g$. We set $T = 11$, since we simulate 10 years forward using BLS data as of the end of 2024, but cannot observe AI patents after 2023. We call the resulting ratio

$$P_I^{Adjusted} \equiv \frac{\text{Predicted flow of New AI Patents from 2024 to } 2034_I}{\text{Stock of Existing Patents as of } 2023_I}. \tag{A.35}$$

Finally, taking the elasticity of new product creation to the rate of new innovation from Argente et al. (2025), we calculate the 10-year impact of AI on product innovation $\Delta \log \alpha_I = 0.0469 \times P_I^{Adjusted}$. Since Goldschlag et al. (2019) does not provide a mapping of patents to the BLS 2-digit sector 55 (management of companies and enterprises), we assign the cross-sector average of $P_I^{Adjusted}$ to that particular sector.

D.2 Other Assumptions.

Calibrating the model requires us to map the observed distribution of employment shares to the model. We focus on the distribution of labor across occupation $\times$ industry cells, which we take from the BLS OEWS. Given that we do not have information about the capital intensity of each occupation, we assume that the economy enters in an initial steady state in which all occupations have the same level of capital intensity. In such a case, we can then set the weights α_I and $\alpha(o, I)$ equal to the across industry and within industry, across occupation wage bill shares and exactly match the initial allocation of labor in the economy as of 2024 in the BLS data. To capture aggregate product innovation, in the post-shock equilibrium we no longer impose that the α_I weights sum to 1 across industries. To reduce the dimensionality of the simulation, we remove all occupation–industry pairs which make up less than 0.1% of an industry's BLS employment. We also drop a small set of occupations in the BLS data for which we cannot observe task-level exposure. Last, we assume symmetric initial task-level capital prices and that the task-level labor share is $s_l = 60\%$, which we take from the latest year of US data reported in Karabarbounis and Neiman (2013). This allows us to calibrate the remaining CES task-level income share parameter γ_j, which we assume is common across all tasks. Finally, the general equilibrium labor supply constant φ in equation (A.19) is pinned down by our assumption that all $w(j) = q(j) = c$ in the initial symmetric equilibrium, where the initial wage/capital price c is identified by the labor share s_l combined with the market clearing condition defined in (A.19).

References

Acemoglu, D., U. Akcigit, and W. R. Kerr (2016). Innovation network. *Proceedings of the National Academy of Sciences 113*(41), 11483–11488.

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics, Volume 4. Amsterdam: Elsevier-North*, pp. 1043–1171.

Argente, D., S. Baslandze, D. Hanley, and S. Moreira (2025, September). Patents to Products: Product Innovation and Firm Dynamics. Working paper, Yale; Federal Reserve Bank of Atlanta; Compendium Labs; Northwestern University. Accessed via URL: https://www.davidargente.com/_files/ugd/b05023_c3467658dd6c4af2b3cd225e3e9780cc.pdf.

Autor, D., C. Chin, A. Salomons, and B. Seegmiller (2024, 03). New frontiers: The origins and content of new work, 1940–2018*. *The Quarterly Journal of Economics 139*(3), 1399–1465.

Goldschlag, N., T. Lybbert, and N. Zolas (2019). An 'algorithmic links with probabilities' crosswalk for uspc and cpc patent classifications with an application towards industrial technology composition. *Economics of Innovation and New Technology 28*(6), 589–609.

Hampole, M., D. Papanikolaou, L. D. Schmidt, and B. Seegmiller (2025, February). Artificial intelligence and the labor market. Working Paper 33509, National Bureau of Economic Research.

Karabarbounis, L. and B. Neiman (2013, 10). The Global Decline of the Labor Share*. *The Quarterly Journal of Economics 129*(1), 61–103.

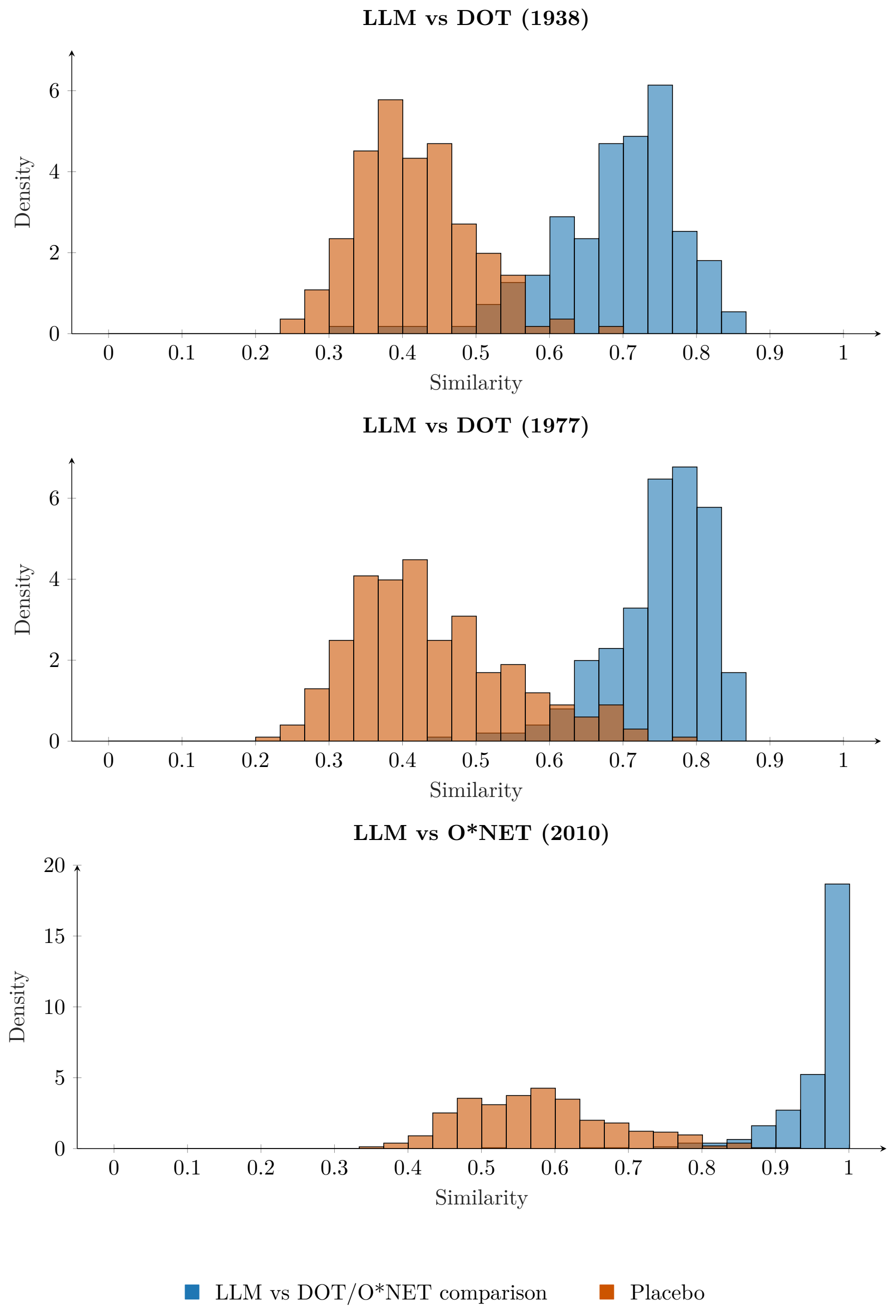
Katz, L. F. and R. A. Margo (2014, January). Technical change and the relative demand for skilled labor: The united states in historical perspective.

Kogan, L., D. Papanikolaou, L. D. Schmidt, and B. Seegmiller (2023, November). Technology and labor displacement: Evidence from linking patents with worker-level data. Working Paper 31846, National Bureau of Economic Research.

Pairolero, N. A., A. V. Giczy, G. Torres, T. Islam Erana, M. A. Finlayson, and A. A. Toole (2025). The artificial intelligence patent dataset (aipd) 2023 update. *The Journal of Technology Transfer*, 1–25.

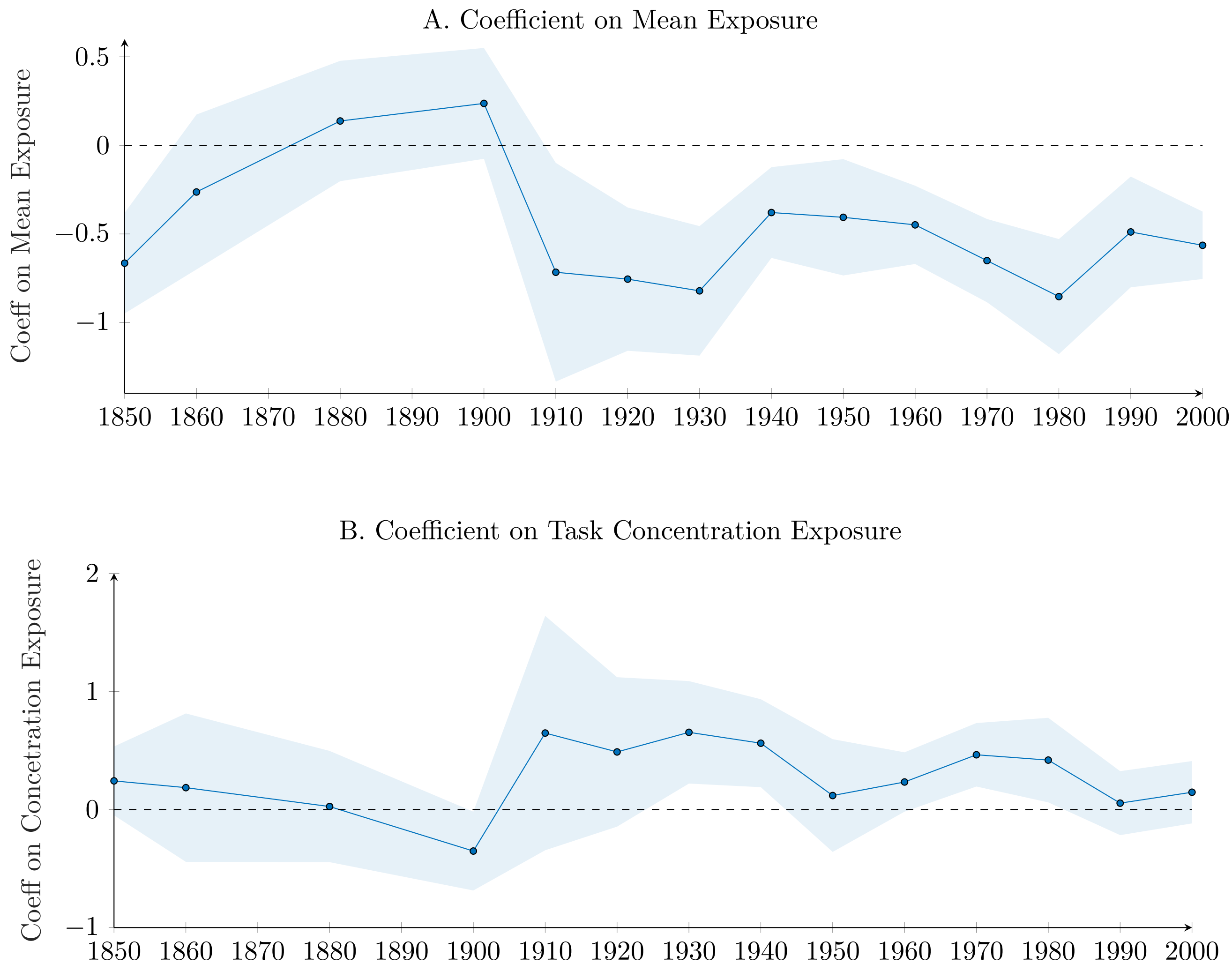
Appendix Figures and Tables

Figure A.1: Validation of LLM tasks using DOT/ONET



Note: The figure shows the distribution of cosine similarities between LLM (*gpt-4o-search-preview*) task embeddings and their DOT/O*NET counterparts. Task embeddings are averaged at the occupation level. We compare 1940 LLM tasks to 1939 DOT tasks, 1980 LLM tasks to 1977 DOT tasks, and 2010 LLM tasks to O*NET 2010 tasks. Because LLM tasks use year-specific occupation codes, whereas DOT uses occ1950rj/occ1990dd18, we match each occ1950rj/occ1990dd18 occupation to the closest year-specific occupation using the crosswalk weights. For O*NET 2010, we link SOC codes to 2010 occupation codes. A placebo distribution pairs each occupation with a random, non-matching occupation.

Figure A.2: Estimated Coefficients, by Decade

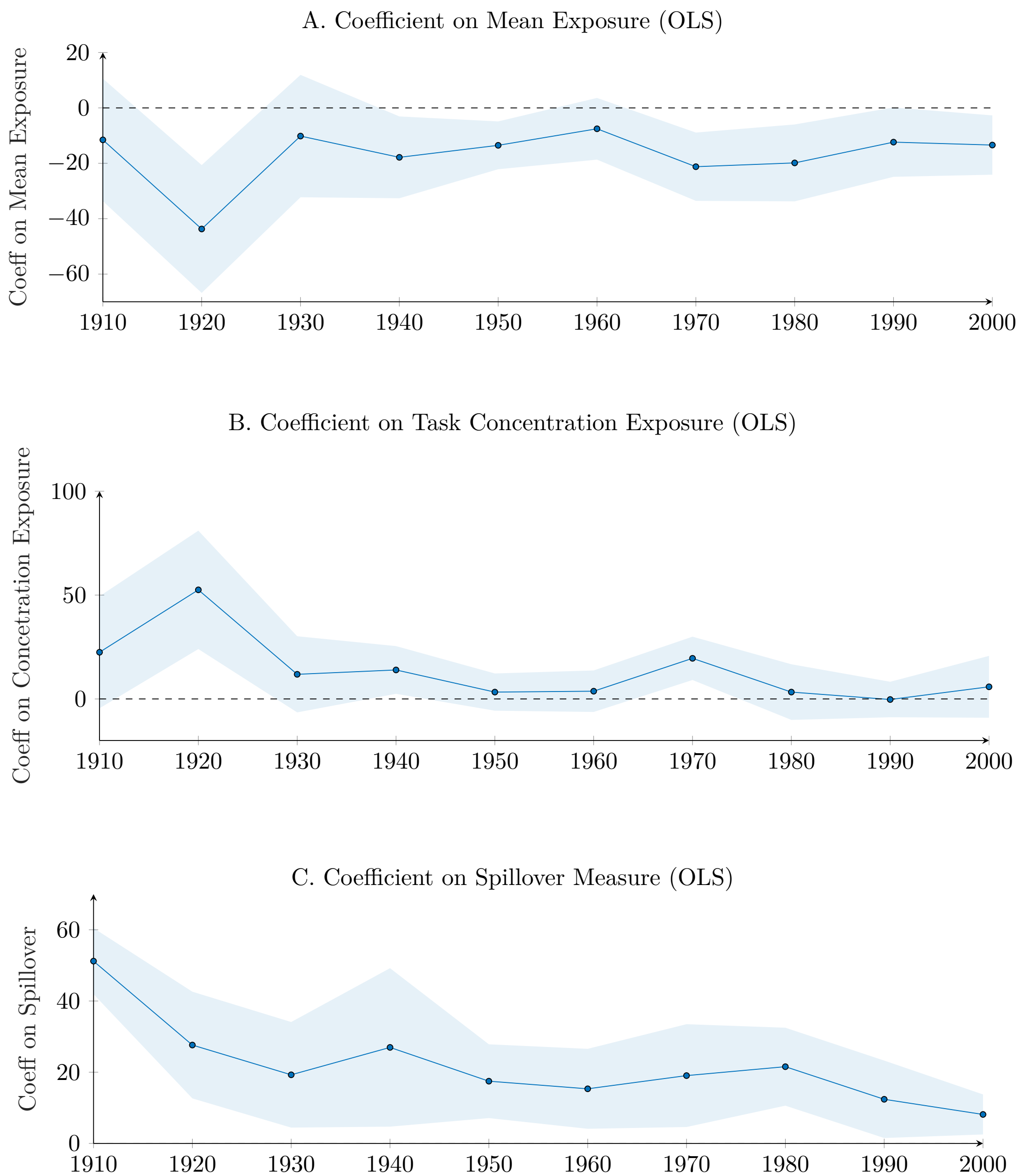


Note: This figure shows the coefficients of the following regression running for each decade T:

$$\log\left(\frac{N_{o,T+20}}{N_{o,T}}\right) = \beta\,\text{Mean Exposure}^{20}_{o,T} + \gamma\,\text{Concentrated Exposure}^{20}_{o,T} + c\,\mathbf{\Gamma}_{o,t} + \varepsilon_{o,T}.$$

Decades 1870 and 1890 are missing due to a lack of 1890 census data. Panel A shows IV coefficients of mean exposure, and Panel B shows IV coefficients of concentrated exposure. Controls $\mathbf{\Gamma}_{o,t}$ include a constant and lagged employment share $N_{o,I,T}$. Coefficients correspond to a unit standard deviation of the dependent variable and are multiplied by 100. Standard errors are clustered by occupation. Shaded bands represent 90 percent confidence intervals.

Figure A.3: Estimated Coefficients, by Decade

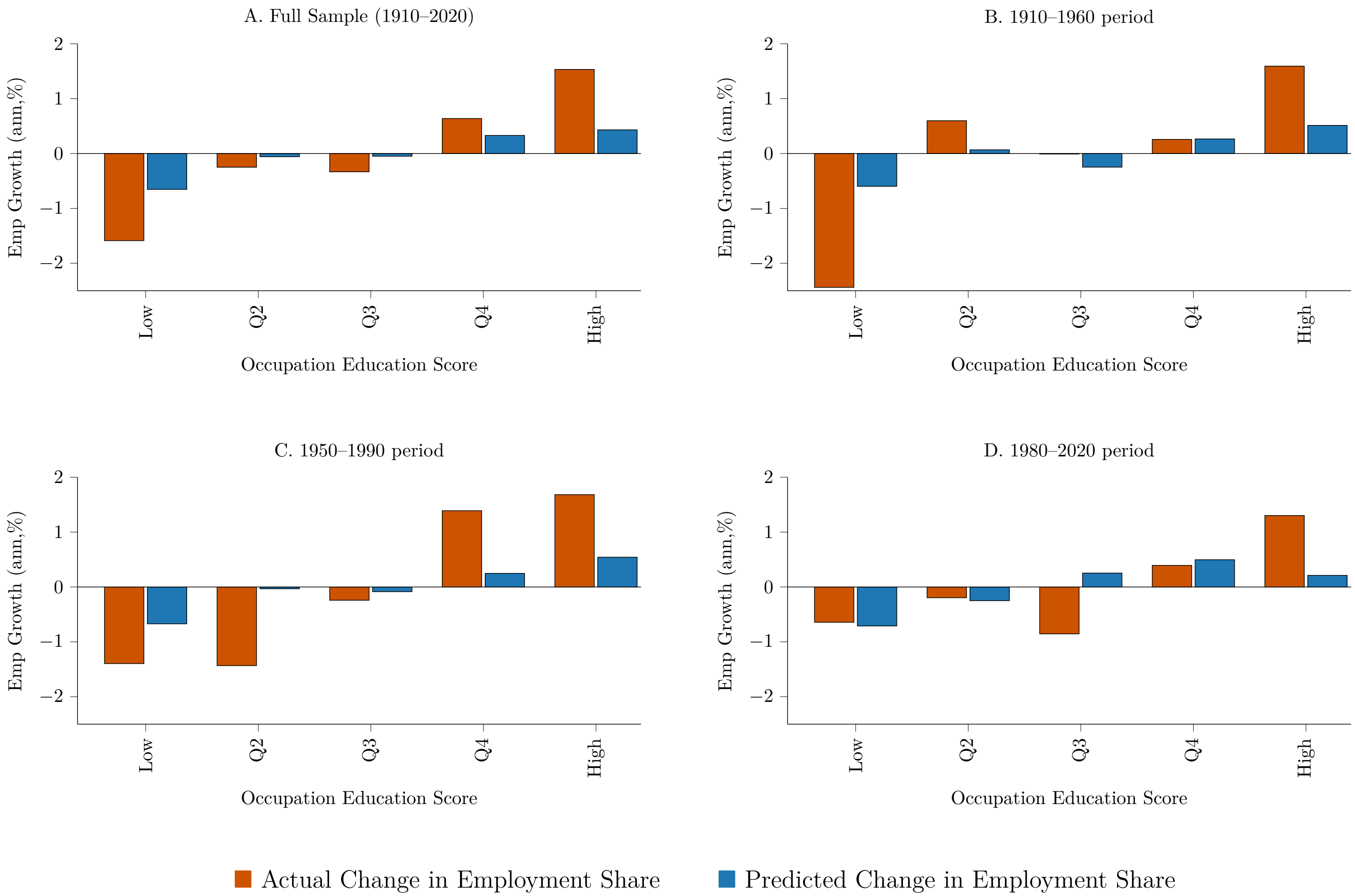


Note: Figure shows the coefficients of the following regression running for each decade T:

$$\log\left(\frac{N_{o,I,T+H}}{N_{o,I,T}}\right) = \beta \,\text{Mean Exposure}^{H}_{o,T} + \gamma \,\text{Exposure Concentration}^{H}_{o,T} + \delta \,\text{Spill}_{I,T} + c\,\mathbf{\Gamma}_{o,I,t} + \varepsilon_{o,I,T}$$

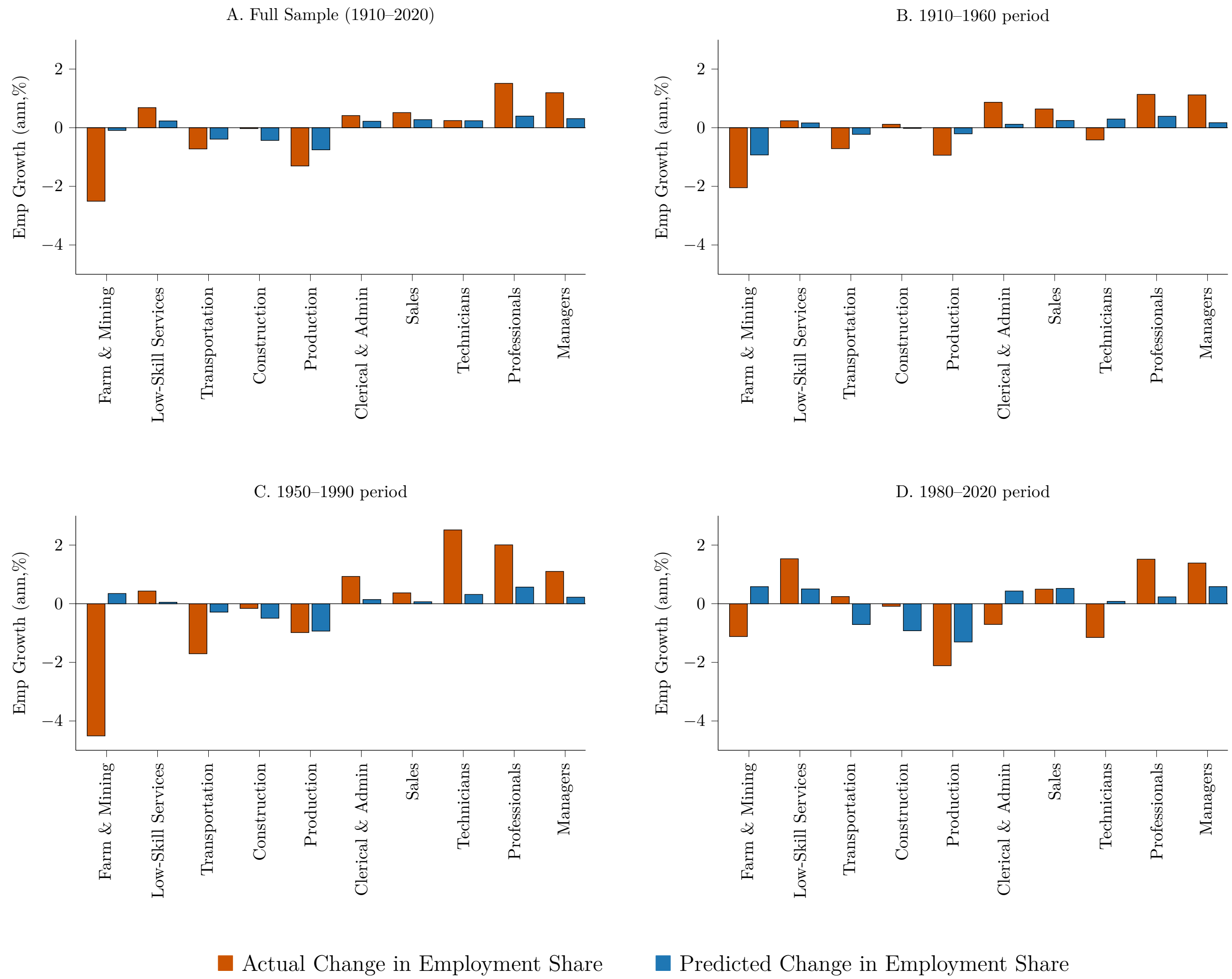
Panels A, B, and C show the coefficients of mean exposure, exposure concentration, and industry spillover, respectively. Controls $\mathbf{\Gamma}_{o,I,t}$ include sector fixed effects and lagged employment share. Coefficients correspond to a one-standard-deviation change in the dependent variable and are multiplied by 100. Standard errors are clustered by occupation and industry. Shaded bands represent 90 percent confidence intervals.

Figure A.4: Technology Exposure and Shifts in Labor Demand Across Occupations: Education



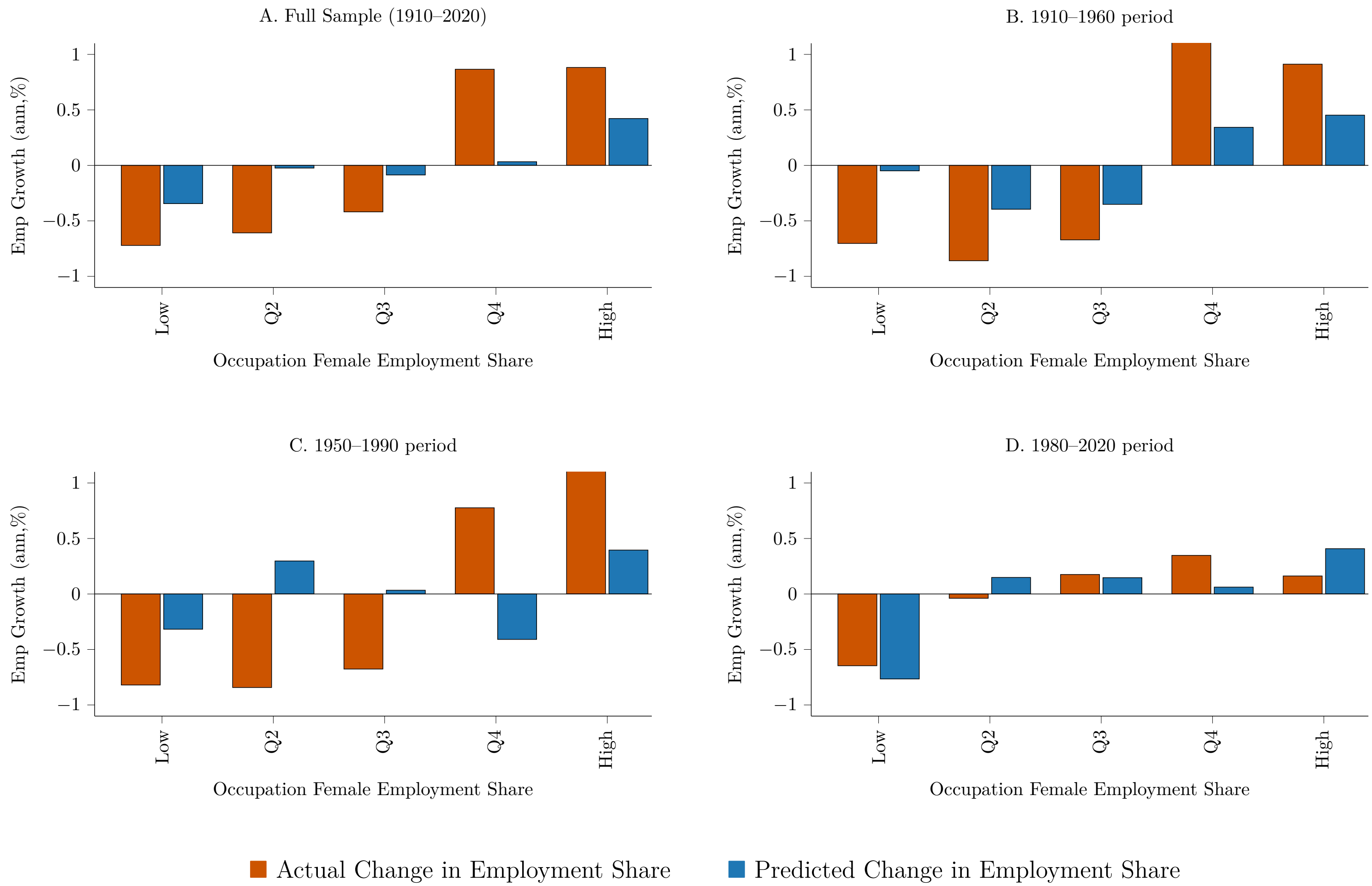
Note: This figure plots actual and technology-predicted average growth rates in employment shares based off estimates of equation (A.3), and also by occupational educational rank, following the procedure described in section 3.3 of the main text. In this figure we sort occupations into yearly employment-weighted quintiles based off educational attainment. We sort using the IPUMs variable "edscor50" for years before 1980 and "edscor90" for 1980 and later.

Figure A.5: Technology Exposure and Shifts in Labor Demand Across Occupations: Polarization



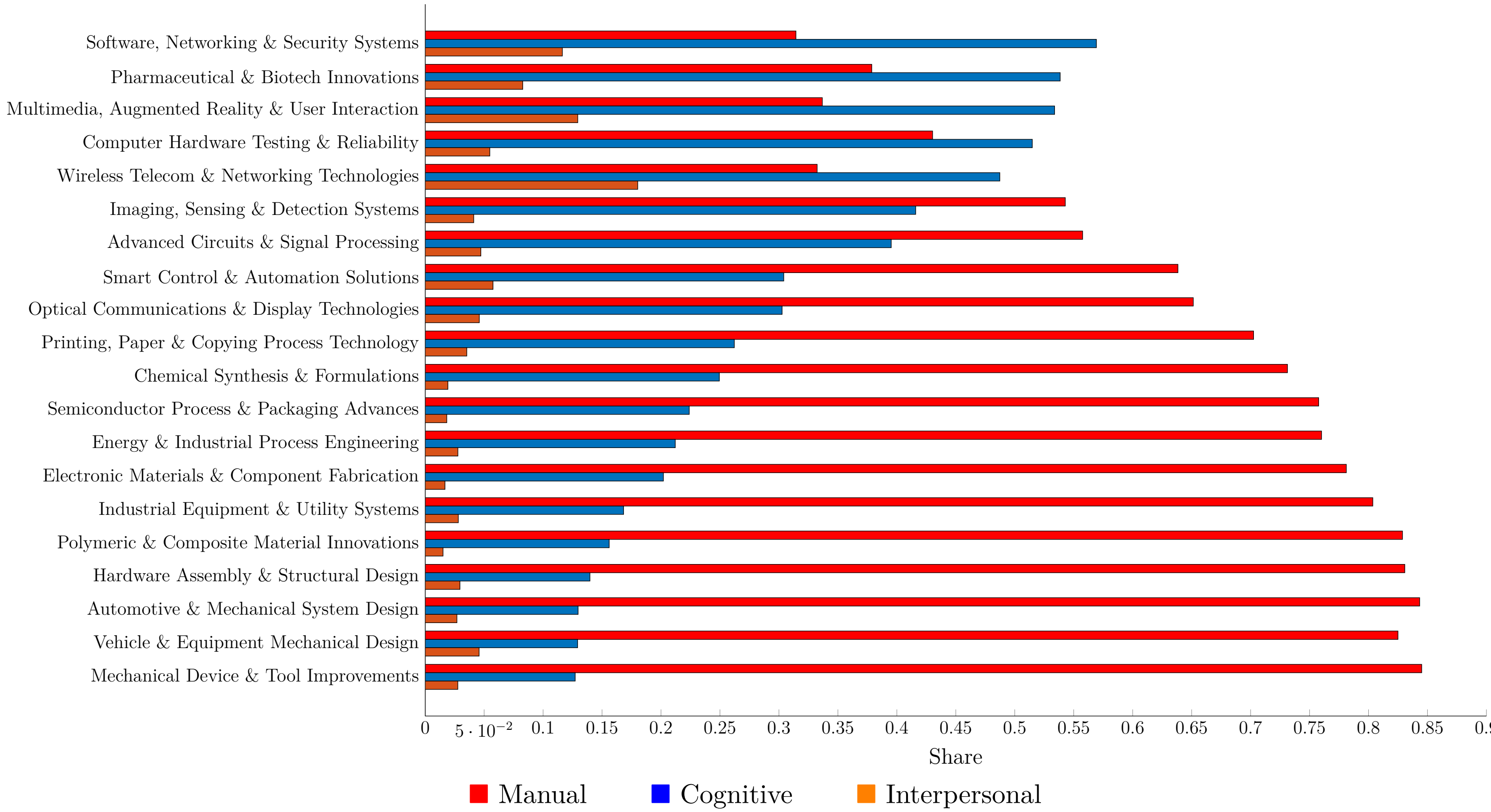
Note: This figure plots actual and technology-predicted average growth rates in employment shares based off estimates of equation (A.3), and also within broad occupation group categories, following the procedure described in section 3.3 of the main text. In this figure we sort occupations into broad time-consistent groups following Autor et al. (2024). Occupation groups are sorted from left to right based off their average wages.

Figure A.6: Technology Exposure and Shifts in Labor Demand Across Occupations: Female Employment Share



Note: This figure plots actual and technology-predicted average annualized growth rates in employment shares based off estimates of equation (A.3), and also by occupation gender composition ranking, following the procedure described in section 3.3 of the main text. In this figure we sort occupations into yearly employment-weighted quintiles based off the share of workers in the occupation who are female.

Figure A.7: Composition of Cluster Technology Exposure, by Task Type



Note: This figure plots the composition of technology exposure by each task type $\tau \in$ {Manual, Cognitive, Interpersonal} in each of the 20 clusters identified in 1980-2025 period. The composition of each type-τ task, c_τ, is defined as the share of all task exposures that are contributed by type-τ tasks in each cluster i. Specifically, we define

$$c_{i,\tau} = \frac{\sum_{T \in \mathcal{T}} \sum_{j \in K(T,\tau)} \text{Exposure}_{i,j,T}}{\sum_{T \in \mathcal{T}} \sum_{j \in K(T)} \text{Exposure}_{i,j,T}},$$

$$\text{Exposure}_{i,j,T} = \frac{1}{|P(i,T)|} \sum_{p \in P(i,T)} \mathbf{1}(\text{similarity}_{p,j} > p95)$$

where $K(T)$ is the set of all tasks in decade T, $K(T,\tau)$ contains all type τ tasks in $K(T)$, the set of decades $\mathcal{T} = \{1980, 1990, 2000\}$, and $P(i,T)$ is the set of all patents in cluster i and issued in decade T.

Table A.1: Examples of Task Sources

Occupation	**Sources**	**Tasks**
1910 blast furnaces and steel rolling mills (includes tin-plate factories)	• Xtek, Inc. - Steel Mill Jobs - A Hardworking History and a Lasting Legacy • MSU Geography - Steel Mill • Wages and Hours of Labor in the Iron and Steel Industry in the United States, 1907 to 1912 - Bulletin of the United States Bureau of Labor Statistics, No. 151 • The Working of Steel - Project Gutenberg	• Unload raw materials such as iron ore, coke, and limestone from railcars and transport them to storage areas or directly to the blast furnace charging system. • Operate skip hoists to charge the blast furnace with measured quantities of raw materials, ensuring proper layering and distribution for efficient smelting. • Monitor and control the heating of blast furnace stoves to produce the necessary hot air blast, adjusting valves and dampers to maintain optimal temperatures. • Regulate the flow of molten iron and slag from the blast furnace during tapping operations, ensuring safe and efficient discharge into ladles or molds. • Charge open-hearth furnaces with pig iron, scrap steel, and fluxes, and control the heating process to produce steel of specified composition and quality.
1940 firemen, fire department	• The Duties of a Fireman, City of Dayton Annual Report for 1940 • Job Bulletin, GovernmentJobs.com • Firefighter Job Description, City of Brentwood • Firefighter, Wikipedia	• Respond to fire alarms and extinguish fires using appropriate equipment and techniques. • Rescue individuals from burning buildings and other hazardous situations. • Operate and maintain firefighting equipment, including hoses, pumps, and ladders. • Perform routine maintenance and cleaning of fire stations, apparatus, and equipment. • Participate in regular training drills to maintain firefighting skills and physical fitness.
1950 airplane-mechanics and repairmen	• U.S. Department of Labor, Dictionary of Occupational Titles, Fourth Edition, Revised 1949 • Civil Aeronautics Administration, Aircraft Maintenance Manual, 1948 • U.S. Air Force, Technical Order 00-20A-1, Maintenance Manual, 1950	• Inspect aircraft frames, engines, and other components for wear, damage, or defects, using hand tools and visual examination. • Perform routine maintenance tasks, such as changing oil, lubricating parts, and replacing worn or defective parts. • Diagnose mechanical or hydraulic system failures to determine necessary repairs. • Repair or replace defective parts, such as wings, brakes, electrical systems, and other aircraft components. • Test aircraft systems to ensure proper functioning using diagnostic instruments.

Note: This table shows example tasks and examples of their corresponding sources as stated by *gpt-4o-search-preview*. The column “Tasks” lists examples of generated task text.

Table A.2: Technology Exposure and Employment Growth, Mean Exposure Only

	A. OLS					
	10-yr Horizon			20-yr Horizon		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Task Exposure	-7.88***	-7.95***	-5.81***	-13.9***	-14.0***	-11.5***
	(1.17)	(1.18)	(1.21)	(2.17)	(2.17)	(2.04)
Obs	3,212	3,212	2,452	3,166	3,166	2,410
R^2 (Within)	0.015	0.022	0.042	0.021	0.032	0.062
	B. IV					
	10-yr Horizon			20-yr Horizon		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Task Exposure	-8.39***	-8.54***	-6.42***	-15.8***	-16.1***	-13.4***
	(1.18)	(1.18)	(1.25)	(2.16)	(2.16)	(2.05)
Obs	3,212	3,212	2,452	3,166	3,166	2,410
R^2 (Within)						
F stat (Exposure)	12,586	12,604	10,846	8,485	8,489	7,194
Year FE	X	X	X	X	X	X
Employment Share, Lag		X	X		X	X
Employment Share, Lag growth			X			X

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,T+H}}{N_{o,T}}\right) = \beta \,\text{Mean Exposure}^H_{o,T} + c\,\mathbf{\Gamma}_{o,t} + \varepsilon_{o,T}.$$

for decades T spanning from 1850–2000, excluding 1890. The main variable of interest is Mean Exposure$^H_{o,T}$, our technology mean exposure measure, normalized to unit standard deviation. Controls $\mathbf{\Gamma}_{o,t}$ include year fixed effects, lagged employment share $N_{o,T}$, and lagged employment growth $\log\left(\frac{N_{o,T}}{N_{o,T-10}}\right)$. The controls included in each regression specification are denoted by X. Coefficients are multiplied by 100. The top panel reports the estimated coefficients using OLS, while the bottom panel reports the IV estimates. Standard errors (in parentheses) are clustered by occupation. Columns 1–3 show regressions with $H = 10$, and columns 4–6 show regressions with $H = 20$.

Table A.3: Technology Exposure and Employment Growth, Comparison across Subsamples

	A. Early Sample (1850–1920)							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-6.34**	-8.79**	-4.12	2.59	-5.44**	-9.91***	-3.85	-2.95
	(2.58)	(3.47)	(4.11)	(6.17)	(2.57)	(3.50)	(4.06)	(6.71)
Concentration in Task Exposure		3.27		-8.64		6.38*		-1.22
		(3.23)		(5.61)		(3.45)		(6.60)
Obs	776	776	791	791	776	776	791	791
R^2 (Within)	0.015	0.016	0.013	0.016				
F stat (Exposure)					3,305	1,993	3,177	1,549
F stat (Concentration)					.	565	.	290
	B. Middle Sample (1910–1970)							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-4.67**	-15.4***	-13.6***	-33.2***	-5.39***	-15.6***	-15.2***	-32.2***
	(2.00)	(3.68)	(3.71)	(6.27)	(2.06)	(4.29)	(3.74)	(6.82)
Concentration in Task Exposure		12.4***		23.1***		12.1**		20.5**
		(4.14)		(7.12)		(5.16)		(8.11)
Obs	1,116	1,116	1,153	1,153	1,116	1,116	1,153	1,153
R^2 (Within)	0.016	0.025	0.037	0.051				
F stat (Exposure)					6,565	3,354	8,080	4,066
F stat (Concentration)					.	898	.	956
	C. Later Sample (1960-2020)							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-11.4***	-15.0***	-21.4***	-29.3***	-12.1***	-15.8***	-23.8***	-32.3***
	(1.41)	(2.71)	(2.95)	(5.00)	(1.44)	(2.80)	(2.92)	(5.21)
Concentration in Task Exposure		4.27*		9.62**		4.48*		10.7**
		(2.26)		(3.78)		(2.40)		(4.29)
Obs	1,320	1,320	1,222	1,222	1,320	1,320	1,222	1,222
R^2 (Within)	0.055	0.058	0.088	0.093				
F stat (Exposure)					14,015	7,698	5,127	2,874
F stat (Concentration)					.	4,175	.	1,790
Year FE	X	X	X	X	X	X	X	X
Employment Share, Lag	X	X	X	X	X	X	X	X

Note: The table above reports results from the following regressions estimated for different subsamples:

$$\log\left(\frac{N_{o,T+H}}{N_{o,T}}\right) = \beta\,\text{Mean Exposure}^H_{o,T} + \gamma\,\text{Exposure Concentration}^H_{o,T} + c\,\mathbf{\Gamma}_{o,t} + \varepsilon_{o,T}.$$

Panels A, B, and C correspond to $T \in [1850, 1900], [1910, 1950], [1960, 2000]$, respectively. Controls $\mathbf{\Gamma}_{o,t}$ include year fixed effects and lagged employment share. The coefficients correspond to a one-standard-deviation change in the dependent variable and are multiplied by 100. Standard errors (in parentheses) are clustered by occupation.

Table A.4: Technology Exposure and Employment Growth, Direct Effects, Employment Weights

	A. OLS					
	10-yr Horizon			20-yr Horizon		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Task Exposure	-11.1***	-11.1***	-11.8***	-21.8***	-21.6***	-21.6***
	(2.46)	(2.58)	(2.98)	(3.89)	(4.09)	(4.22)
Concentration in Task Exposure	5.71***	5.66***	7.14***	11.2***	10.9***	12.3***
	(1.91)	(1.98)	(2.34)	(2.80)	(2.86)	(3.09)
Obs	3,212	3,212	2,452	3,166	3,166	2,410
R^2 (Within)	0.025	0.030	0.029	0.046	0.055	0.064
	B. IV					
	10-yr Horizon			20-yr Horizon		
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Task Exposure	-11.1***	-11.2***	-11.9***	-23.2***	-23.4***	-22.5***
	(2.58)	(2.75)	(3.04)	(4.34)	(4.65)	(4.65)
Concentration in Task Exposure	4.46**	4.53**	5.70**	10.4***	10.4***	10.9***
	(2.13)	(2.24)	(2.51)	(3.27)	(3.42)	(3.51)
Obs	3,212	3,212	2,452	3,166	3,166	2,410
R^2 (Within)						
F stat (Exposure)	2,499	2,474	1,855	1,465	1,462	1,087
F stat (Concentration)	1,407	1,408	1,114	948	955	817
Year FE	X	X	X	X	X	X
Employment Share, Lag		X	X		X	X
Employment Share, Lag growth			X			X

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,T+H}}{N_{o,T}}\right) = \beta\,\text{Mean Exposure}^H_{o,T} + \gamma\,\text{Exposure Concentration}^H_{o,T} + c\,\mathbf{\Gamma}_{o,t} + \varepsilon_{o,T}.$$

for decades T spanning from 1850–2000, excluding 1890. The variables of interest are Mean Exposure$^H_{o,T}$, technology mean exposure, and Exposure Concentration$^H_{o,T}$, technology exposure concentration (both normalized to unit standard deviation). Controls $\mathbf{\Gamma}_{o,t}$ include year fixed effects, lagged employment share $N_{o,T}$, and lagged employment growth $\log\left(\frac{N_{o,T}}{N_{o,T-10}}\right)$. The controls included in each regression specification are denoted by X. Coefficients are multiplied by 100. The top panel reports the estimated coefficients using OLS, while the bottom panel reports the IV estimates. Standard errors (in parentheses) are clustered by occupation. Columns 1–3 show regressions with $H = 10$, and columns 4–6 show regressions with $H = 20$. Regressions are weighted by the employment share of which occupation; weights are right-tail winsorized at the 5% level to reduce the influence of large occupations (for example, agriculture workers).

Table A.5: Technology Exposure and Employment Growth, Subsamples, Men Only

	A. Early Sample (1850–1920)							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-4.32*	-6.55*	-0.89	5.49	-3.78	-7.22**	-0.86	0.48
	(2.61)	(3.49)	(4.09)	(6.10)	(2.58)	(3.58)	(4.02)	(6.62)
Concentration in Task Exposure		2.96		-8.19		4.89		-1.83
		(3.26)		(5.50)		(3.58)		(6.46)
Obs	776	776	791	791	776	776	791	791
R^2 (Within)	0.010	0.011	0.009	0.012				
F stat (Exposure)					3,304	1,993	3,176	1,549
F stat (Concentration)					.	565	.	290
	B. Middle Sample (1910–1970)							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-3.17	-12.9***	-11.1***	-33.8***	-3.96*	-13.1***	-12.7***	-33.8***
	(2.01)	(3.86)	(3.64)	(5.86)	(2.08)	(4.60)	(3.62)	(5.95)
Concentration in Task Exposure		11.2**		26.8***		10.8*		25.7***
		(4.42)		(6.03)		(5.63)		(6.07)
Obs	1,116	1,116	1,151	1,151	1,116	1,116	1,151	1,151
R^2 (Within)	0.011	0.019	0.030	0.048				
F stat (Exposure)					6,529	3,340	8,013	4,093
F stat (Concentration)					.	896	.	1,090
	C. Later Sample (1960-2020)							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-9.79***	-13.2***	-18.4***	-24.1***	-10.4***	-13.8***	-20.7***	-26.8***
	(1.36)	(2.61)	(2.86)	(4.89)	(1.38)	(2.74)	(2.82)	(5.14)
Concentration in Task Exposure		4.10*		6.90*		4.15*		7.70*
		(2.20)		(3.83)		(2.38)		(4.35)
Obs	1,318	1,318	1,222	1,222	1,318	1,318	1,222	1,222
R^2 (Within)	0.044	0.046	0.070	0.073				
F stat (Exposure)					14,024	7,715	5,150	2,884
F stat (Concentration)					.	4,106	.	1,785
Year FE	X	X	X	X	X	X	X	X
Employment Share, Lag	X	X	X	X	X	X	X	X

Note: The table above reports results from the same regressions as Table A.3, but restricts the sample to male labor only. Panels A, B, and C correspond to $T \in [1850, 1900], [1910, 1950], [1960, 2000]$, respectively. Controls $\mathbf{\Gamma}_{o,t}$ include year fixed effects and lagged employment share. Coefficients correspond to a one-standard-deviation change in the dependent variable and are multiplied by 100. Standard errors (in parentheses) are clustered by occupation.

Table A.6: Technology Exposure and Employment Growth, LLM Robustness: 1850–1920

	A. Baseline (GPT4o-Search)							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-6.34**	-8.79**	-4.12	2.59	-5.44**	-9.91***	-3.85	-2.95
	(2.58)	(3.47)	(4.11)	(6.17)	(2.57)	(3.50)	(4.06)	(6.71)
Concentration in Task Exposure		3.27		-8.64		6.38*		-1.22
		(3.23)		(5.61)		(3.45)		(6.60)
Obs	776	776	791	791	776	776	791	791
R^2 (Within)	0.015	0.016	0.013	0.016				
F stat (Exposure)					3,305	1,993	3,177	1,549
F stat (Concentration)					.	565	.	290
	B. Alternative LLM: GPT 4o							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-1.69	-6.57	0.29	-5.43	-1.73	-6.86*	-0.94	-7.41
	(3.39)	(4.17)	(4.84)	(6.07)	(3.26)	(3.89)	(4.60)	(5.65)
Concentration in Task Exposure		4.61***		5.60***		5.06***		6.68***
		(1.45)		(2.12)		(1.38)		(1.90)
Obs	776	776	791	791	776	776	791	791
R^2 (Within)	0.009	0.014	0.011	0.015				
F stat (Exposure)					2,427	7,323	1,989	7,164
F stat (Concentration)					.	61,556	.	30,759
	C. Alternative LLM: Llama							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-1.58	1.46	1.53	5.95	-2.33	-3.53	-0.40	-1.57
	(3.60)	(6.65)	(4.86)	(8.39)	(3.39)	(5.59)	(4.71)	(7.73)
Concentration in Task Exposure		-2.92		-4.50		1.15		1.19
		(4.37)		(5.76)		(3.55)		(5.26)
Obs	776	776	791	791	776	776	791	791
R^2 (Within)	0.009	0.010	0.012	0.013				
F stat (Exposure)					1,994	1,098	1,945	1,114
F stat (Concentration)					.	1,555	.	987
Year FE	X	X	X	X	X	X	X	X
Employment Share, Lag	X	X	X	X	X	X	X	X

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,T+H}}{N_{o,T}}\right) = \beta \,\text{Mean Exposure}^{H}_{o,T} + \gamma \,\text{Exposure Concentration}^{H}_{o,T} + c\,\boldsymbol{\Gamma}_{o,t} + \varepsilon_{o,T}, \quad T \in [1850, 1900].$$

Panels A, B, and C correspond to measures using tasks generated by *gpt-4o-search-preview*, *gpt-4o-2024-11-20.* and *Meta-Llama-3.1-405B-Instruct*, respectively. Controls $\boldsymbol{\Gamma}_{o,t}$ include year fixed effects and lagged employment share. Coefficients correspond to a one-standard-deviation change in the dependent variable and are multiplied by 100. Standard errors (in parentheses) are clustered by occupation.

Table A.7: Technology Exposure and Employment Growth, LLM Robustness: 1910–1970 (Appendix)

	A. Baseline (GPT4o-Search)							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-4.67**	-15.4***	-13.6***	-33.2***	-5.39***	-15.6***	-15.2***	-32.2***
	(2.00)	(3.68)	(3.71)	(6.27)	(2.06)	(4.29)	(3.74)	(6.82)
Concentration in Task Exposure		12.4***		23.1***		12.1**		20.5**
		(4.14)		(7.12)		(5.16)		(8.11)
Obs	1,116	1,116	1,153	1,153	1,116	1,116	1,153	1,153
R^2 (Within)	0.016	0.025	0.037	0.051				
F stat (Exposure)					6,565	3,354	8,080	4,066
F stat (Concentration)					.	898	.	956
	B. Alternative LLM: GPT 4o							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-0.96	-14.7***	-5.49	-29.4***	-2.11	-15.4***	-7.60*	-30.2***
	(2.09)	(3.35)	(4.07)	(5.57)	(2.05)	(3.36)	(3.98)	(5.52)
Concentration in Task Exposure		24.3***		41.7***		24.2***		40.8***
		(5.95)		(8.03)		(5.95)		(8.00)
Obs	1,116	1,116	1,153	1,153	1,116	1,116	1,153	1,153
R^2 (Within)	0.011	0.030	0.023	0.062				
F stat (Exposure)					5,373	3,091	7,268	5,276
F stat (Concentration)					.	1,299	.	2,659
	C. Alternative LLM: Llama							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-2.77	-7.20**	-10.7***	-21.1***	-4.11**	-9.53***	-13.0***	-25.0***
	(2.16)	(2.89)	(3.73)	(4.89)	(2.07)	(3.01)	(3.60)	(5.10)
Concentration in Task Exposure		5.83**		13.7***		7.32**		16.5***
		(2.91)		(4.64)		(3.38)		(5.19)
Obs	1,116	1,116	1,153	1,153	1,116	1,116	1,153	1,153
R^2 (Within)	0.013	0.016	0.031	0.037				
F stat (Exposure)					6,088	3,161	7,322	4,268
F stat (Concentration)					.	1,211	.	1,710
Year FE	X	X	X	X	X	X	X	X
Employment Share, Lag	X	X	X	X	X	X	X	X

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,T+H}}{N_{o,T}}\right) = \beta\,\text{Mean Exposure}^{H}_{o,T} + \gamma\,\text{Exposure Concentration}^{H}_{o,T} + c\,\mathbf{\Gamma}_{o,t} + \varepsilon_{o,T}, \quad T \in [1910, 1950].$$

Panels A, B, and C correspond to measures using tasks generated by *gpt-4o-search-preview*, *gpt-4o-2024-11-20.* and *Meta-Llama-3.1-405B-Instruct*, respectively. Controls $\mathbf{\Gamma}_{o,t}$ include year fixed effects and lagged employment share. Coefficients correspond to a one-standard-deviation change in the dependent variable and are multiplied by 100. Standard errors (in parentheses) are clustered by occupation.

Table A.8: Technology Exposure and Employment Growth, LLM Robustness: 1960–2020

	A. Baseline (GPT4o-Search)							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-11.4***	-15.0***	-21.4***	-29.3***	-12.1***	-15.8***	-23.8***	-32.3***
	(1.41)	(2.71)	(2.95)	(5.00)	(1.44)	(2.80)	(2.92)	(5.21)
Concentration in Task Exposure		4.27*		9.62**		4.48*		10.7**
		(2.26)		(3.78)		(2.40)		(4.29)
Obs	1,320	1,320	1,222	1,222	1,320	1,320	1,222	1,222
R^2 (Within)	0.055	0.058	0.088	0.093				
F stat (Exposure)					14,015	7,698	5,127	2,874
F stat (Concentration)					.	4,175	.	1,790
	B. Alternative LLM: GPT 4o							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-10.3***	-14.6***	-18.6***	-30.8***	-11.1***	-15.6***	-21.7***	-34.3***
	(1.35)	(2.34)	(2.82)	(4.64)	(1.38)	(2.46)	(2.84)	(4.85)
Concentration in Task Exposure		7.05**		20.3***		7.48**		21.8***
		(2.98)		(5.89)		(3.21)		(6.18)
Obs	1,320	1,320	1,222	1,222	1,320	1,320	1,222	1,222
R^2 (Within)	0.046	0.051	0.069	0.087				
F stat (Exposure)					15,201	8,043	6,250	3,329
F stat (Concentration)					.	2,683	.	1,117
	C. Alternative LLM: Llama							
	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-10.3***	-13.8***	-19.2***	-27.8***	-11.4***	-15.1***	-22.9***	-30.2***
	(1.40)	(2.14)	(2.91)	(4.04)	(1.43)	(2.22)	(2.96)	(4.39)
Concentration in Task Exposure		5.21**		12.4***		5.58**		10.8**
		(2.42)		(4.35)		(2.35)		(4.36)
Obs	1,320	1,320	1,222	1,222	1,320	1,320	1,222	1,222
R^2 (Within)	0.047	0.050	0.074	0.082				
F stat (Exposure)					11,247	5,469	4,703	2,263
F stat (Concentration)					.	2,284	.	920
Year FE	X	X	X	X	X	X	X	X
Employment Share, Lag	X	X	X	X	X	X	X	X

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,T+H}}{N_{o,T}}\right) = \beta\,\text{Mean Exposure}^{H}_{o,T} + \gamma\,\text{Exposure Concentration}^{H}_{o,T} + c\,\mathbf{\Gamma}_{o,t} + \varepsilon_{o,T}, \quad T \in [1960, 2000].$$

Panels A, B, and C correspond to measures using tasks generated by *gpt-4o-search-preview*, *gpt-4o-2024-11-20.* and *Meta-Llama-3.1-405B-Instruct*, respectively. Controls $\mathbf{\Gamma}_{o,t}$ include year fixed effects and lagged employment share. Coefficients correspond to a one-standard-deviation change in the dependent variable and are multiplied by 100. Standard errors (in parentheses) are clustered by occupation.

Table A.9: Technology Exposure and Employment Growth, Manual Share Control

	OLS				IV			
	10-yr Horizon		20-yr Horizon		10-yr Horizon		20-yr Horizon	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-3.55**	-3.30	-7.10***	-5.63	-4.32**	-5.86*	-7.35***	-8.64*
	(1.66)	(2.97)	(2.51)	(4.24)	(1.70)	(3.04)	(2.83)	(4.75)
Concentration in Task Exposure			3.76*	2.47			3.32	3.08
			(1.94)	(3.20)			(2.20)	(3.66)
Obs	3,212	3,166	3,212	3,166	3,212	3,166	3,212	3,166
R^2 (Within)	0.032	0.052	0.033	0.052				
F stat (Exposure)					8,251	5,875	4,474	3,038
F stat (Concentration)					.	.	1,820	1,222
Year FE	X	X	X	X	X	X	X	X
Employment Share, Lag	X	X	X	X	X	X	X	X
Share of Manual Tasks	X	X	X	X	X	X	X	X
Share of Cognitive Tasks	X	X	X	X	X	X	X	X

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,T+H}}{N_{o,T}}\right) = \beta\,\text{Mean Exposure}^{H}_{o,T} + \gamma\,\text{Exposure Concentration}^{H}_{o,T} + c\,\boldsymbol{\Gamma}_{o,t} + \varepsilon_{o,T}$$

for decades T spanning from 1850–2000, excluding 1890. The variables of interest are Mean Exposure$^{H}_{o,T}$, technology mean exposure, and Exposure Concentration$^{H}_{o,T}$, technology exposure concentration (both normalized to unit standard deviation). Controls $\boldsymbol{\Gamma}_{o,t}$ include year fixed effects, lagged employment share, and the occupational share of manual and cognitive tasks. Coefficients are multiplied by 100. Standard errors (in parentheses) are clustered by occupation.

Table A.10: Technology Exposure and Employment Growth, Controlling for Industry Trends

	A. OLS							
Employment Growth (%)	10 Years				20 Years			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-4.21***	-4.27***	-8.12***	-8.37***	-7.21***	-7.31***	-14.9***	-15.3***
	(1.24)	(1.24)	(2.06)	(2.07)	(2.27)	(2.32)	(3.41)	(3.48)
Concentration in Task Exposure			4.75**	4.97***			9.42***	9.82***
			(1.88)	(1.87)			(2.80)	(2.83)
N	149,217	149,217	149,217	149,217	138,923	138,923	138,923	138,923
R^2 (Within)	0.002	0.002	0.002	0.003	0.003	0.003	0.005	0.006
	B. IV							
Employment Growth (%)	10 Years				20 Years			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-5.05***	-5.07***	-8.11***	-8.43***	-9.64***	-9.65***	-16.1***	-16.5***
	(1.20)	(1.21)	(2.08)	(2.09)	(2.18)	(2.21)	(3.26)	(3.32)
Concentration in Task Exposure			3.81*	4.17**			8.20***	8.78***
			(1.93)	(1.92)			(2.86)	(2.89)
N	149,217	149,217	149,217	149,217	138,923	138,923	138,923	138,923
R^2 (Within)								
F stat (Exposure)	8,398	8,482	4,390	4,438	4,677	4,745	2,473	2,511
F stat (Concentration)	.	.	1,624	1,636	.	.	1,150	1,162
Year FE	X		X		X		X	
Industry FE	X		X		X		X	
Year × Industry FE		X		X		X		X
Employment Share, Lag	X	X	X	X	X	X	X	X

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,I,T+H}}{N_{o,I,T}}\right) = \beta\,\text{Mean Exposure}^H_{o,T} + \gamma\,\text{Exposure Concentration}^H_{o,T} + c\,\boldsymbol{\Gamma}_{o,I,t} + \varepsilon_{o,I,T}.$$

for decades T spanning from 1910–2000. The variables of interest are Mean Exposure$^H_{o,T}$, technology mean exposure, and Exposure Concentration$^H_{o,T}$, technology exposure concentration (both normalized to unit standard deviation). Controls $\boldsymbol{\Gamma}_{o,I,t}$ include year fixed effects, industry fixed effects (or year × industry fixed effects), and lagged employment share. Specific control specifications are denoted by X. Coefficients are multiplied by 100. Standard errors (in parentheses) are clustered by occupation and industry.

Table A.11: Technology Exposure and Employment Growth, Industry Controls–Innovation Spillovers, Male Only

	A. OLS							
Employment Growth (%)	10 Years				20 Years			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-2.48*	-2.43*	-5.92***	-5.97***	-4.38*	-4.30*	-11.5***	-11.5***
	(1.28)	(1.28)	(2.08)	(2.07)	(2.34)	(2.34)	(3.39)	(3.41)
Concentration in Task Exposure			4.20**	4.33**			8.79***	8.85***
			(1.84)	(1.83)			(2.68)	(2.69)
Industry Spillover	8.18***	10.8***	8.19***	10.8***	17.6***	21.0***	17.6***	21.0***
	(2.46)	(1.74)	(2.46)	(1.74)	(3.38)	(3.43)	(3.38)	(3.43)
N	125,351	125,351	125,351	125,351	115,057	115,057	115,057	115,057
R^2 (Within)	0.004	0.006	0.005	0.006	0.011	0.012	0.013	0.014
	B. IV							
Employment Growth (%)	10 Years				20 Years			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-2.98**	-3.16**	-5.95***	-6.17***	-8.24***	-8.18***	-15.1***	-15.1***
	(1.26)	(1.25)	(2.08)	(2.08)	(2.22)	(2.22)	(3.23)	(3.24)
Concentration in Task Exposure			3.72**	3.77**			8.74***	8.87***
			(1.85)	(1.86)			(2.71)	(2.71)
Industry Spillover	43.9**	42.2***	43.9**	42.2***	15.0*	19.1**	15.0*	19.1**
	(19.30)	(15.72)	(19.29)	(15.71)	(8.09)	(7.41)	(8.10)	(7.41)
N	124,000	124,000	124,000	124,000	100,420	100,420	100,420	100,420
R^2 (Within)								
F stat (Exposure)	4,132	4,163	2,886	2,913	2,287	2,348	1,645	1,699
F stat (Concentration)			1,056	1,355			861	890
F stat (Spillover)	7	3	6	39	25	20	16	54
Year FE	X		X		X		X	
Sector FE	X		X		X		X	
Year × Sector FE		X		X		X		X
Employment Share, Lag	X	X	X	X	X	X	X	X

Note: The table above reports results from the same regressions as Table 2, but restricts the sample to male labor only. Controls $\mathbf{\Gamma}_{o,t}$ include year fixed effects, broad-sector fixed effects (or year × sector fixed effects), and lagged employment share. Specific control specifications are denoted by X. Coefficients correspond to a one-standard-deviation change in the independent variable and are multiplied by 100. Standard errors (in parentheses) are clustered by occupation and industry. Panel A reports the estimated coefficients using OLS, while Panel B reports the IV estimates. Columns 1, 2, 5, and 6 show estimation excluding exposure concentration measure, while the others include it. Columns Columns 1–4 show regressions with $H = 10$, while the others show regressions with $H = 20$.

Table A.12: Technology Exposure and Employment Growth, Industry Controls–Innovation Spillovers (1910-1970 period)

	A. OLS							
Employment Growth (%)	10 Years				20 Years			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-1.78	-1.89	-9.85***	-10.0***	-3.20	-3.25	-16.2***	-16.3***
	(2.33)	(2.33)	(3.43)	(3.44)	(3.86)	(3.86)	(5.57)	(5.59)
Concentration in Task Exposure			9.91***	9.99***			16.2***	16.3***
			(2.90)	(2.91)			(4.50)	(4.54)
Industry Spillover	11.9**	16.8***	11.9**	16.8***	29.3***	32.9***	29.2***	32.8***
	(4.66)	(2.95)	(4.66)	(2.95)	(5.51)	(5.05)	(5.51)	(5.05)
N	61,669	61,669	61,669	61,669	54,821	54,821	54,821	54,821
R^2 (Within)	0.006	0.010	0.008	0.012	0.015	0.018	0.019	0.022
	B. IV							
Employment Growth (%)	10 Years				20 Years			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-2.38	-2.37	-10.7***	-10.8***	-4.26	-4.32	-16.4***	-16.4***
	(2.40)	(2.38)	(3.60)	(3.60)	(3.83)	(3.83)	(5.45)	(5.49)
Concentration in Task Exposure			10.5***	10.6***			15.7***	15.6***
			(3.30)	(3.32)			(4.86)	(4.89)
Industry Spillover	86.4	53.7*	86.5	53.9*	17.7	25.2**	17.5	25.1**
	(77.92)	(28.24)	(78.03)	(28.31)	(13.30)	(12.44)	(13.35)	(12.45)
N	61,319	61,319	61,319	61,319	54,682	54,682	54,682	54,682
R^2 (Within)								
F stat (Exposure)	2,362	2,358	1,919	1,929	2,204	2,593	1,873	2,034
F stat (Concentration)			477	478			540	503
F stat (Spillover)	1	2	1	2	15	20	11	15
Year FE	X		X		X		X	
Sector FE	X		X		X		X	
Year × Sector FE		X		X		X		X
Employment Share, Lag	X	X	X	X	X	X	X	X

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,I,T+H}}{N_{o,I,T}}\right) = \beta\,\text{Mean Exposure}^H_{o,T} + \gamma\,\text{Exposure Concentration}^H_{o,T} + \delta\,\text{Spill}_{I,T} + c\,\mathbf{\Gamma}_{o,I,t} + \varepsilon_{o,I,T}, \quad T \in [1910, 1950]$$

for decades T spanning 1910–1950. Controls $\mathbf{\Gamma}_{o,I,t}$ include year fixed effects, broad-sector fixed effects (or year × sector fixed effects), and lagged employment share. Specific control specifications are denoted by X. Coefficients correspond to a one-standard-deviation change in the independent variable and are multiplied by 100. Standard errors (in parentheses) are clustered by occupation and industry. Panel A reports the estimated coefficients using OLS, while Panel B reports the IV estimates. Columns 1, 2, 5, and 6 show estimates excluding exposure concentration measure, while the others include it. Columns 1–4 show regressions with $H = 10$, while the others show regressions with $H = 20$.

Table A.13: Technology Exposure and Employment Growth, Industry Controls–Innovation Spillovers (1960-2020 period)

	A. OLS							
Employment Growth (%)	10 Years				20 Years			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-5.99***	-6.03***	-7.17***	-7.23***	-10.5***	-10.5***	-14.8***	-14.8***
	(1.31)	(1.31)	(2.29)	(2.28)	(2.42)	(2.42)	(4.03)	(4.04)
Concentration in Task Exposure			1.41	1.45			5.16	5.20
			(1.94)	(1.93)			(3.55)	(3.54)
Industry Spillover	4.07**	5.43***	4.07**	5.43***	10.5***	13.5***	10.5***	13.5***
	(1.86)	(2.02)	(1.86)	(2.02)	(3.55)	(3.94)	(3.55)	(3.94)
N	75,523	75,523	75,523	75,523	71,642	71,642	71,642	71,642
R^2 (Within)	0.005	0.006	0.005	0.006	0.012	0.014	0.013	0.015
	B. IV							
Employment Growth (%)	10 Years				20 Years			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Task Exposure	-6.69***	-6.74***	-7.98***	-8.07***	-13.0***	-13.0***	-17.7***	-17.8***
	(1.29)	(1.28)	(2.18)	(2.18)	(2.36)	(2.36)	(3.80)	(3.81)
Concentration in Task Exposure			1.59	1.64			5.91*	5.96*
			(1.91)	(1.92)			(3.44)	(3.43)
Industry Spillover	36.7	36.1*	36.7	36.1*	15.6*	16.1*	15.6*	16.1*
	(23.62)	(20.80)	(23.61)	(20.79)	(9.17)	(8.17)	(9.17)	(8.17)
N	74,318	74,318	74,318	74,318	71,274	71,274	71,274	71,274
R^2 (Within)								
F stat (Exposure)	5,287	5,495	4,073	4,181	2,212	2,241	1,711	1,748
F stat (Concentration)			2,750	2,496			1,037	1,021
F stat (Spillover)	2	2	2	2	10	10	7	6
Year FE	X		X		X		X	
Sector FE	X		X		X		X	
Year × Sector FE		X		X		X		X
Employment Share, Lag	X	X	X	X	X	X	X	X

Note: The table above reports results from regressions of the form

$$\log\left(\frac{N_{o,I,T+H}}{N_{o,I,T}}\right) = \beta \,\text{Mean Exposure}^{H}_{o,T} + \gamma \,\text{Exposure Concentration}^{H}_{o,T} + \delta \,\text{Spill}_{I,T} + c\,\mathbf{\Gamma}_{o,I,t} + \varepsilon_{o,I,T}, \quad T \in [1960, 2000]$$

for decades T spanning 1910–2000. Controls $\mathbf{\Gamma}_{o,I,t}$ include year fixed effects, broad-sector fixed effects (or year × sector fixed effects), and lagged employment share. Specific control specifications are denoted by X. Coefficients correspond to a one-standard-deviation change in the independent variable and are multiplied by 100. Standard errors (in parentheses) are clustered by occupation and industry. Panel A reports the estimated coefficients using OLS, while Panel B reports the IV estimates. Columns 1, 2, 5, and 6 show estimates excluding exposure concentration measure, while the others include it. Columns 1–4 show regressions with $H = 10$, while the others show regressions with $H = 20$.