AI and Comparative Advantage*

Lukas Althoff† Hugo Reichardt‡

March 20, 2025

Abstract

Despite large estimated productivity gains from AI across a wide range of tasks, the labor market effects are not well understood. This paper develops a methodology to estimate workers' comparative advantage across tasks and the labor market impact of task-biased technical change. We propose and estimate a dynamic general equilibrium model in which workers build multi-dimensional skills and can switch occupations based on their evolving comparative advantage. The model is informed by existing detailed data on tasks in the US labor market and new data on their skill requirements. Using our new methodology, we derive predictions on how AI affects the overall wage distribution and how individual workers are affected (results forthcoming). We also study whether commonly used ad hoc measures of occupational exposure predict well how much a worker stands to gain or lose from AI.

*Pedro Carvalho provided excellent research assistance. This work was supported by the Stanford Institute of Economic Policy Research (SIEPR).

†Stanford University. lalthoff@stanford.edu

‡Centre de Recerca en Economica Internacional (CREI). hreichardt@crei.cat

1 Introduction

Artificial intelligence (AI) has the potential to significantly impact labor markets by increasing productivity in a wide range of tasks and occupations (e.g., Eloundou et al., 2024; Brynjolfsson et al., 2023). AI is being adopted at unprecedented rates, with 1-in-9 American workers reporting the daily use of generative AI as of August 2024 (Bick et al., 2024). Understanding the labor market effects of such dramatic technological change is important for workers, employers, and policymakers in guiding educational and occupational choices, hiring practices, and potential retraining programs.

In this paper, we develop a methodology to structurally estimate a technology's labor market impact given knowledge about its task-level productivity effects. We use this methodology to assess AI's impact on individual workers, the overall wage distribution, and the returns to various skills.

Our model embeds task-based production in a dynamic discrete choice model of the labor market to capture key features necessary to understand the effect of technological change. First, a worker's productivity in each task—and thus their comparative advantage—depends on the match between the worker's multi-dimensional skills and the skill requirements of the task. Capturing this comparative advantage is crucial for identifying whether a technology's impact on a worker's wages is positive or negative; for example, to a doctor, it matters whether AI can take over their administrative tasks vs. diagnoses and treatment. Second, workers dynamically choose occupations that entail different sets of tasks, reflecting their ability to adapt to technological change. Third, workers accumulate skills on the job, inducing forward-looking occupational decisions and potentially different effects on older workers who may have less incentive to retrain.

Importantly, the model is designed to allow for a direct mapping between its theoretical objects and data on tasks, occupations, and skill requirements. Specifically, we use data from O*NET, which contains detailed information on 19,530 tasks across 974 occupations and occupational skill requirements across 35 dimensions of skill.[1] To identify each worker's task-level comparative advantage, we construct a new database of the skills required to complete each task. Closely following O*NET's questionnaires, we use OpenAI's large language model GPT-4o to estimate the skill requirements for each task and skill dimension. We validate our estimates of the task-level skill requirements by appropriately aggregating them to occupational skill requirements and comparing those with O*NET: the average correlation across skill dimensions is 0.9.

To estimate the labor market effects of AI using our model, we use estimates of AI's productivity gains in each task. Such estimates are already available from the literature, using both humans and generative AI (e.g., Brynjolfsson et al., 2018; Eisfeldt et al., 2023;

[1] For computational tractability and interpretability, we reduce the dimensionality of skill to 4.

Eloundou et al., 2024). To align the estimates of the productivity gains with the model, we estimate them for each tasks using GPT-4o. We consider two scenarios: one where technological change is restricted to generative AI, and another that includes AI-powered robots. Our measure for generative AI correlates strongly with that from Eloundou et al. (2024), including their human-rated assessments.[2] Our methodology also allows for various scenarios for AI's development, and predictions can be updated as more information becomes available.

Related literature. This paper contributes to a large literature on task-based production and technological change (e.g., Zeira, 1998; Autor et al., 2003; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018). We view our contribution to this literature as twofold. First, we design our model to leverage detailed data on occupations, tasks, and skill requirements and provide a methodology to estimate each workers' comparative advantage across tasks, the transferability of their skills between tasks, their occupational outside options, and ultimately, how much they stand to gain or lose from task-level technological change. Second, we integrate a model of task-based production into a general equilibrium dynamic discrete occupational choice model (in the spirit of Keane and Wolpin, 1997; Heckman et al., 1998; Lee and Wolpin, 2006; Llull, 2018) where workers accumulate multidimensional skills (Lindenlaub, 2017; Guvenen et al., 2020; Lise and Postel-Vinay, 2020; Baley et al., 2022).[3]

We also contribute to the empirical literature on technologies' labor market effects, including studies on industrial robots (Graetz and Michaels, 2018; Acemoglu and Restrepo, 2020) and artificial intelligence (e.g. Webb, 2020; Acemoglu et al., 2022). Our new structural approach is uniquely suited to study the effects of technologies that are not yet fully adopted (or even invented), complementing existing papers, which rely on reduced form evidence on the effects of past technology adoption.[4] Further, we think our method is useful even when studying the effects of more mature or historical technologies as plausibly exogenous variation in technology adoption is not generally available.

Last, we provide new estimates of AI's productivity impacts across tasks and occupations. Following Eloundou et al. (2024); Eisfeldt et al. (2023), we use GPT-4o to estimate AI-induced time savings across worker tasks, finding our estimates align closely with existing estimates including human expert evaluations and experimental evidence from randomized controlled trials with AI-augmented workers (Brynjolfsson et al., 2018;

[2] We are not aware of any existing productivity estimates related to AI-powered robots.

[3] Autor and Handel (2013) consider a static model with multi-dimensional skills, tasks, and a discrete occupational choice. Our dynamic model allows to estimate the effects of task-level technological change and incorporates skill accumulation. Further, unlike Autor and Handel (2013), we distinguish between "task efficiencies" and "skills", the latter being transferable across tasks.

[4] Webb (2020) extrapolates a reduced form correlation between occupational wages and "technological exposure" to software and robots to predict how AI will affect the wage distribution. This methodology assumes that these technologies affect occupations equally (and that the correlation reflects causation). Furthermore, even if those assumption are satisfied, it only allows to study how occupational outcomes are affected, not individual *workers*, who may switch occupations in response to technological change.

Webb, 2020; Campero et al., 2022; Kanazawa et al., 2022; Vaithilingam et al., 2022; Brynjolfsson et al., 2023; Dell'Acqua et al., 2023; Eloundou et al., 2024; Noy and Zhang, 2023; Clarke and Hanrahan, 2024; Cui et al., 2024; Peng et al., 2024). Importantly, we extend beyond previous work's focus on generative AI by considering scenarios including AI-powered robots and self-driving vehicles. Our key substantive contribution is developing a methodology to translate these task-level productivity estimates into labor market effects.

2 Theoretical Framework

In the model below, we describe how workers choose occupations, perform tasks, and accumulate human capital over time. Workers select occupations to maximize lifetime utility, considering wages, task requirements, and individual preferences. Productivity in each occupation depends on the match between a worker's skills and the tasks performed. Human capital grows through on-the-job learning, where workers improve skills by performing tasks that are challenging but not too difficult. Occupational choices evolve dynamically, shaping skill development and labor market outcomes.

2.1 Model

Occupations and tasks. Each occupation produces a distinct good by combining a unique combination of tasks. These tasks are combined with a constant elasticity of substitution ρ, so that the production function of occupation j is

$$Y_j = \left(\sum_{\tau \in \mathcal{T}_j} \theta_{j,\tau}^{\frac{1}{\rho}} y_\tau^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}} \tag{1}$$

where $\mathcal{T}_j$ is the discrete set of relevant tasks, y_τ is the output of task τ, and $\theta_j(\tau)$ is task τ's importance weight.

Task-level productivity and skills. The production function for task τ is

$$y_\tau(\boldsymbol{h}, l_\tau) = l_\tau \gamma_\tau f(\boldsymbol{h}, \boldsymbol{r}(\tau)) \tag{2}$$

where l_τ represents the units of labor assigned to task τ, γ_τ is a task-specific productivity parameter, $\boldsymbol{h} = (h_1, \cdots, h_S)'$ is the worker's human capital, and $\boldsymbol{r}(\tau) \equiv (r_1(\tau), \cdots, r_S(\tau))'$ is the skill requirement of task τ. We model the technological change brought about by artificial intelligence as an increase in γ_τ, i.e., an increase in each worker's productivity at completing task τ.[5]

[5]Note that this formulation nests the special case of Hicks-neutral technological change: a proportional

The function that governs a worker's productivity in task τ is

$$f(\boldsymbol{h}, \boldsymbol{r}(\tau)) = \prod_{s \in S} h_s^{\omega_s} \exp\left(-\eta_s \left(h_s - r_s(\tau)\right)^2\right). \tag{3}$$

The first term in equation (3) reflects a force that makes workers with higher skills more productive in any task, independent of its skill requirements. The second term captures the "mismatch" between a worker's skills and the task's skill requirements, reflecting that a worker uses their skills less efficiently in a task for which they are over- or under-qualified. The skill requirement can thus be understood as the skill set that maximizes comparative advantage in task τ.

Time allocation across tasks. Having chosen an occupation, each worker allocates their time across tasks to maximize output in equation (1). Formally, the optimal allocation of time is

$$\{l_\tau^*\}_{\tau \in \mathcal{T}_j} = \underset{\{l_\tau\}_{\tau \in \mathcal{T}_j}}{\arg\max} \left\{ \left(\sum_{\tau \in \mathcal{T}_j} \theta_{j,\tau}^{\frac{1}{\rho}} \left(l_\tau \gamma_\tau f(\boldsymbol{h}, \boldsymbol{r}(\tau))\right)^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}} \right\}, \quad \text{s.t.} \sum_{\tau \in \mathcal{T}_j} l_\tau = 1. \tag{4}$$

The solution to (4) for a given task $\tau \in \mathcal{T}_j$ is

$$l_{j,\tau}^*(\boldsymbol{h}) = \frac{\theta_{j,\tau} \left(\gamma_\tau f(\boldsymbol{h}, \boldsymbol{r}(\tau))\right)^{\rho-1}}{\sum_{\kappa \in \mathcal{T}_j} \theta_{j,\kappa} \left(\gamma_\kappa f(\boldsymbol{h}, \boldsymbol{r}(\kappa))\right)^{\rho-1}}. \tag{5}$$

The solution shows that more time is spent on tasks with greater weight $\theta_{j,\tau}$. Also, if tasks are substitutes ($\rho > 1$) a worker focuses on tasks in which they are more productive. If instead tasks are complements, workers compensate lower productivity in a task by allocating more time to it.

Occupational productivity. Combining the optimal time allocation in equation (5) with the production functions in equations (1), (2), and (3) shows how a worker's output in occupation j depends on their human capital $\boldsymbol{h}$:

$$Y_j(\boldsymbol{h}) = \prod_{s \in S} h_s^{\omega} \left(\sum_{\tau \in \mathcal{T}_j} \theta_{j,\tau} \gamma_\tau^{\rho-1} \exp\left((1-\rho) \sum_{s \in S} \eta_s \left(h_s - r_s(\tau)\right)^2 \right) \right)^{\frac{1}{\rho-1}}. \tag{6}$$

In sum, a worker's productivity in a given occupation depends on the match between their skills and the skill requirements of each task in that occupation.

Human capital accumulation. Before entering the labor market at age $a = 1$, each worker draws an initial skill vector $\boldsymbol{h}_1$ after which they accumulate further skills on the job. We assume that a worker's human capital accumulation depends on their current skills, their

increase in γ_τ for all tasks.

"ability to learn" ψ, and the skill requirements of the job j: $h_{s,t+1} = g_{j,s}(h_{s,t}, \psi)$ (similar to Lise and Postel-Vinay, 2020). Learning follows the following process:

$$g_{j,s}(h_s, \psi) = (1-\delta)h_s + \sum_{\tau \in \mathcal{T}_j} \theta_{j,\tau} \max\{r_s(\tau) - h_s, 0\} \exp\left(-\lambda_\psi \max\{r_s(\tau) - h_s, 0\}\right) \quad (7)$$

which has several implications for human capital accumulation. First, absent learning, skills depreciate with rate $\delta \in [0,1]$. Second, workers learn by performing tasks that are "hard" for them, i.e., tasks that have skill requirements above their current skill levels. Third, as tasks become harder relative to the workers' skill, the rate at which the skills catch up decreases; $\lambda_\psi \geq 0$ governs the rate of this slowdown. Taken together, this means workers learns most from tasks that are "hard" but not "too hard".[6] Fourth, workers differ in their ability to learn ψ (similar to Heckman et al., 1998).

Wages and demand for occupations. Workers consume occupational goods and substitute them with constant elasticity σ. That is, $c^{\frac{\sigma-1}{\sigma}} \equiv \sum_{j=1}^{J} \alpha_j^{\frac{1}{\sigma}} c_j^{\frac{\sigma-1}{\sigma}}$ where c_j denotes consumption of good j and $\sum_{j=1}^{J} \alpha_j = 1$. The price of each good is denoted p_j. Constant elasticity of substitution implies that for each pair of goods i and j, the price ratio is

$$\frac{p_i}{p_j} = \left(\frac{\alpha_i}{\alpha_j}\right)^{\frac{1}{\sigma}} \left(\frac{\mathcal{Y}_j}{\mathcal{Y}_i}\right)^{\frac{1}{\sigma}} \quad (8)$$

where $\mathcal{Y}_j$ denotes total supply of good j.

Perfect competition in the labor market ensures that a worker's wage in occupation j equates the value of their output, so that

$$w_j(\boldsymbol{h}) = p_j Y_j(\boldsymbol{h}). \quad (9)$$

Occupational choice. Every period, each worker chooses from a discrete set of occupations to maximize utility. A worker's expected lifetime utility at age $a = 1, \ldots, A$ is represented by the value function

$$V_a(\boldsymbol{h}, \psi) = \mathbb{E}\left[\max_{j \in \{1,\ldots,J\}} w_j(\boldsymbol{h}) + \mu_j + \varepsilon_j + \beta V_{a+1}\left(g_j(\boldsymbol{h}), \psi\right)\right] \quad (10)$$

where μ_j captures job amenities, $\mathbb{E}[\cdot]$ is the expectations operator over the occupation-specific preference shocks ε_j, $g_j(\boldsymbol{h}, \cdot) \equiv \left(g_{j,1}(h_1, \cdot), \cdots, g_{j,S}(h_S, \cdot)\right)'$, and $\boldsymbol{r}_j \equiv \{\boldsymbol{r}(\tau)\}_{\tau \in \mathcal{T}_j}$.

We assume that the preference shocks ε_j follow a type I generalized extreme value

[6] Specifically, learning from task τ is maximized when $r_s(\tau) - h_s = \frac{1}{\lambda_\psi}$, which yields a learning gain of $\frac{1}{e\lambda_\psi}$. Proof: The function $f(x) = x\exp(-\lambda x)$ has derivative $f'(x) = (1-\lambda x)\exp(-\lambda x)$ which is zero for $x = \frac{1}{\lambda}$, strictly positive for $x < \frac{1}{\lambda}$, and strictly negative for $x > \frac{1}{\lambda}$.

(Gumbel) distribution with mean 0 and scale parameter ζ.[7] This assumption implies that the conditional probability of choosing occupation j has the closed-form solution $\mathbb{P}_j(a, \boldsymbol{h}, \psi) \propto \exp\left(\frac{1}{\zeta}\left(w_j(\boldsymbol{h}) + \mu_j + \beta V_{a+1}\left(g_j\left(\boldsymbol{h}, \psi\right)\right)\right)\right)$ so that the value function in (10) can be simplified to

$$V_a(\boldsymbol{h}, \psi) = \begin{cases} \zeta \ln \sum_{j=1}^{J} \exp\left(\frac{1}{\zeta}\left(w_j(\boldsymbol{h}) + \mu_j + \beta V_{a+1}\left(g_j\left(\boldsymbol{h}\right), \psi\right)\right)\right) & \text{if } a < A \\ \zeta \ln \sum_{j=1}^{J} \exp\left(\frac{1}{\zeta}\left(w_j(\boldsymbol{h}) + \mu_j\right)\right) & \text{if } a = A. \end{cases} \tag{11}$$

Since the problem is static at age $a = A$, equations (6), (9), (7), and (11) solve the value function for a given set of prices by backward iteration from age $a = A$ to $a = 1$.

Overlapping generations. A new generation with measure $\frac{1}{A}$ replaces the old each period, such that the overall population size and age distribution remain constant.

2.2 Equilibrium

Definition (Competitive equilibrium)**.** Given an initial distribution of age, human capital, and learning ability, $G_0(a, \boldsymbol{h}, \psi)$ and a distribution of human capital at birth $G_t(1, \boldsymbol{h}, \psi)$, a *competitive equilibrium* is defined as a sequence of prices $\left\{p_{1,t}, \cdots, p_{J,t}\right\}_{t=1}^{\infty}$ such that

- Workers' occupational choices maximize the present value of lifetime utility given the sequence of prices. That is,

$$\mathbb{P}_{j,t}(a, \boldsymbol{h}_t, \psi) = \frac{\exp\left(\frac{1}{\zeta}\left(w_j(\boldsymbol{h}) + \beta V_{a+1}\left(g_j\left(\boldsymbol{h}, \psi\right)\right)\right)\right)}{\sum_{k=1}^{J} \exp\left(\frac{1}{\zeta}\left(w_k(\boldsymbol{h}) + \beta V_{a+1}\left(g_k\left(\boldsymbol{h}, \psi\right)\right)\right)\right)}$$

  where $V_a(\boldsymbol{h}, \psi)$ is defined in equation (11);

- The distribution of human capital is consistent with occupational choices. That is,

$$G_t\left(a, \boldsymbol{h}, \psi\right) = \sum_{j=1}^{J} \int_{g_j(\boldsymbol{h}_{-1}, \psi) \leq \boldsymbol{h}} \mathbb{P}_{j,t-1}\left(a-1, \boldsymbol{h}_{-1}, \psi\right) dG_{t-1}\left(a-1, \boldsymbol{h}_{-1}, \psi\right) \quad \forall a \geq 2;$$

- Prices are consistent with the demand equation in (8) and the supply of good j is

$$\mathcal{Y}_{j,t} = \int Y_j(\boldsymbol{h}) \mathbb{P}_{j,t}(a, \boldsymbol{h}, \psi) dG_t(a, \boldsymbol{h}, \psi).$$

[7] The CDF is $F(\varepsilon) = \exp\left(-\exp\left(-\frac{\varepsilon + \zeta\overline{\gamma}}{\zeta}\right)\right)$ where $\overline{\gamma} \approx 0.577$ is the Euler–Mascheroni constant.

2.3 Solution Methods

Having defined the equilibrium, we now provide algorithms to solve for a stationary competitive equilibrium as well as the transition path after an unexpected one-off change to the parameters of the model.

Stationary equilibrium. To solve for a stationary equilibrium, we use the following algorithm:

1. Guess an initial vector of relative prices $\left(p_1^{(1)}, \cdots, p_J^{(1)}\right)$.

2. For iteration r, given the prices $\left(p_1^{(r)}, \cdots, p_J^{(r)}\right)$, solve the worker's problem and compute the long-run stationary distribution of human capital, occupational choice, and output of each good $\left(\mathcal{Y}_1^{(r)}, \cdots, \mathcal{Y}_J^{(r)}\right)$. Given the supply of each good, update relative prices from the demand equation in (8). That is, for each pair of goods i and j, the updated relative price is

$$\frac{p_i^{(r+1)}}{p_j^{(r+1)}} = \left(\frac{\alpha_i}{\alpha_j}\right)^{\frac{1}{\sigma}} \left(\frac{\mathcal{Y}_j^{(r)}}{\mathcal{Y}_i^{(r)}}\right)^{\frac{1}{\sigma}}.$$

3. Repeat step 2 until $\left\| \boldsymbol{p}^{(r+1)} - \boldsymbol{p}^{(r)} \right\| < \epsilon$ for a threshold $\epsilon > 0$.

Transition path. We also solve for the transition path after a one-off unexpected parameter change. Starting from the initial stationary equilibrium, we solve for the transition path of equilibrium prices $\left\{p_{1,t}, \cdots, p_{J,t}\right\}_{t=1}^{\infty}$ from the moment the shock is realized onward. We numerically approximate this infinite sequence by solving for $\left\{p_{1,t}, \cdots, p_{J,t}\right\}_{t=1}^{T}$ for a large enough T such that prices are constant after period T. The solution algorithm we propose is based on (Boppart et al., 2018):

1. Compute the stationary equilibrium before ($t = 0$) and after the change ($t = T$).

2. Guess a path for the joint distribution of human capital and occupational choices.[8]

3. For iteration r, given the sequence of prices $\left\{p_{1,t}^{(r)}, \cdots, p_{J,t}^{(r)}\right\}_{t=1}^{T}$ and the value function at time $t = T$, solve for the value function at $t = T - 1$ by backward iteration. Then, given the value function and implied prices at $t = T - 1$, do the same for $t = T - 2$, etc., until $t = 1$. Given these value functions, simulate the distribution of occupational choices and human capital forward from $t = 1, \ldots, T$. Update the guess for the implied sequence of prices using equation (8).

4. Repeat step 3 until $\left\| \boldsymbol{p}_t^{(r+1)} - \boldsymbol{p}_t^{(r)} \right\| < \epsilon$ for all $t = 1, \ldots, T$ for a threshold $\epsilon > 0$.

[8] A reasonable guess is the path that would occur were prices to adjust immediately to the new stationary equilibrium.

3 Data

To estimate the labor market impact of technological change, we combine our theoretical framework with granular data on tasks, occupations, and wages in the US economy and new data on task-level skill requirements and AI capabilities.

3.1 Occupations and Tasks (O*NET)

O*NET is the leading database on occupations, tasks, and skills in the US economy (e.g., Autor et al., 2003; Acemoglu and Autor, 2011; Lise and Postel-Vinay, 2020). O*NET contains detailed descriptions of 19,350 tasks linked to 974 occupations.

We rely on O*NET data to define both the occupations and tasks in our model. That is, we set the tasks employed across occupations in our model, $\mathcal{T}_j$, to mirror those in the O*NET data. We set the weights of each task τ in an occupation, $\theta_{j,\tau}$, to the importance measure of that task as reported in O*NET.[9]

3.2 Task-level Skill Requirements

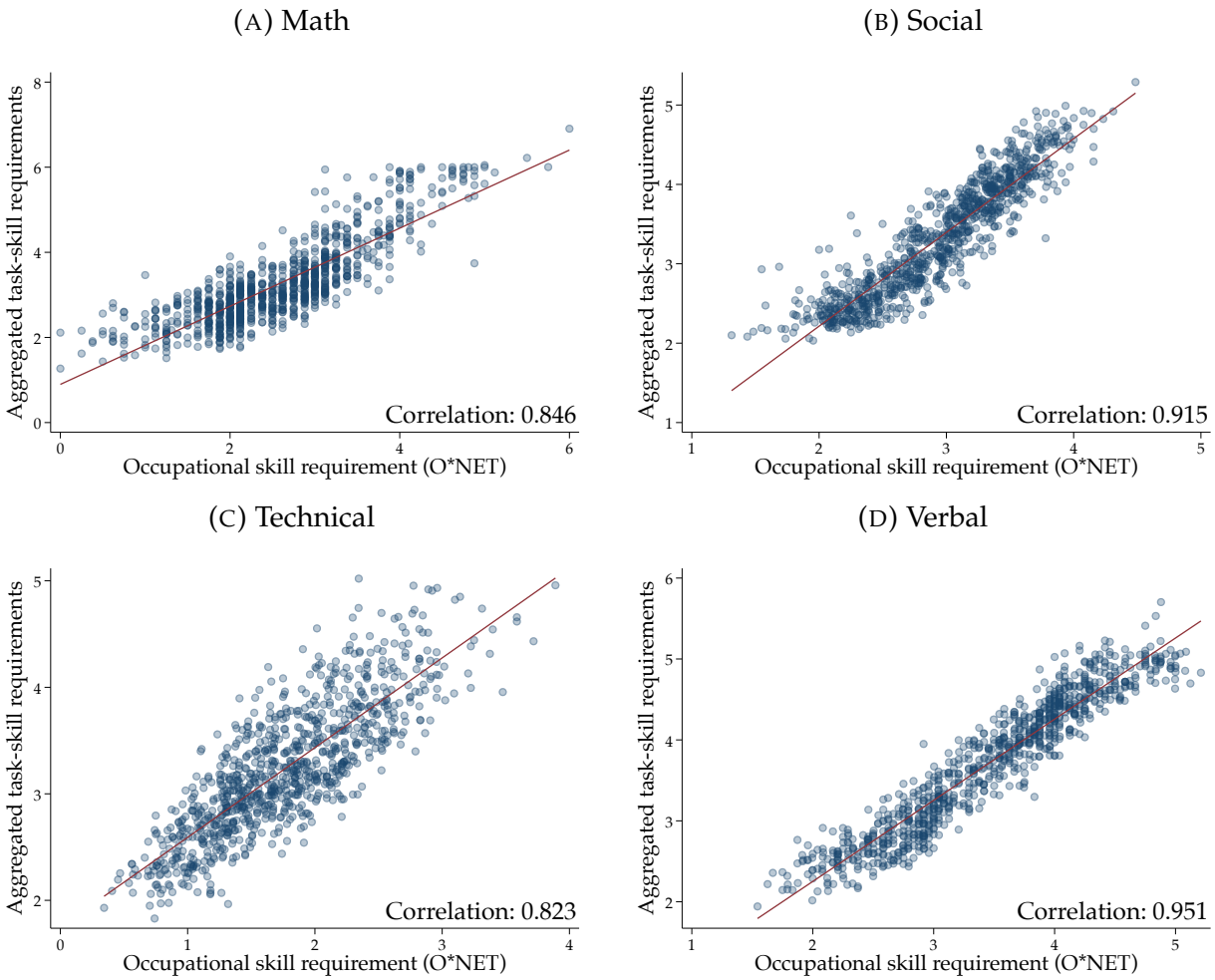
The impact of task-level technical change on workers crucially depends on the worker' comparative advantage in the affected tasks. Our model highlights that workers' comparative advantage across tasks is determined by the match between their skills and the tasks' skill requirements, $\boldsymbol{r}(\boldsymbol{\tau})$. While O*NET provides data on skill requirements at the occupation level, it lacks task-specific data. To fill this gap, we use OpenAI's GPT-4o large language model to estimate skill requirements at the task level. We design our methodology to maximize consistency with O*NET by using their 35 skill dimensions (e.g., "reading comprehension" or "social perceptiveness") and the same questionnaire format, including their 1-7 rating scale and level anchors.[10] This process requires 664,020 GPT-4o queries to cover all 19,530 occupational task descriptions across 35 skill dimensions. Appendix C.5.1 details our prompt design and provides example prompts.

To validate our data on task-level skill requirements, we aggregate them to the occupation level and compare them with O*NET's occupation-level data. Specifically, we compare the importance-weighted average skill requirements, $\sum_{\tau \in \mathcal{T}_j} \theta_{j,\tau} r_s(\tau)$, with O*NET's occupation-level aggregates. The agreement rates are high, with correlations ranging between 0.82 and 0.95 (see Figure 1). Besides being intuitive, this validation method is grounded in theory: a worker's occupational mismatch is minimized when their skills equal the average of task-specific skill requirements, weighted by their importance (to a first order approximation; see Section 4 for further details on this).

[9] The importance weights are normalized to sum to 1 within occupations.

[10] The questionnaire and text for each skill's level descriptions can be found here.

FIGURE 1: Validation of task skill requirement data with O*NET



Notes: This figure shows the correlation between the occupation-level skill requirement in the O*NET database and the GPT-4o generated task-level skill requirements aggregated to the occupation-level for the four skills used in the analysis. Each observation represents an occupation in the O*NET database.

We reduce O*NET's skill dimensions to allow for a mapping between skill requirements and data on worker skills. Following Addison et al. (2020); Baley et al. (2022); DeLoach et al. (2022), we assume one's skill is comprised of four dimensions: mathematics, social, technical, and verbal skills. Table A.1 provides the mapping of the original O*NET skill dimension to the four skills used in the analysis. For each of the four skills, we define the skill requirements to be the average across the relevant O*NET skills.

3.3 Task-level Productivity of AI Technologies

We construct a new dataset on the worker time saved in completing each task with the assistance of AI technologies, γ_τ. While existing work has identified different measures of "AI exposure", none provides a continuous measure of worker time saved—the key measure of interest in our model. To fill this gap, we follow previous work methodologically and elicit estimates of time savings from GPT-4o (Eloundou et al., 2024; Eisfeldt

et al., 2023). Appendix C.5.2 describes our prompt design.

We validate our new data on task-level AI productivity in two ways. First, we correlate them with existing measures of occupations' or tasks exposure to LLMs (see Table A.2 for a list of existing AI exposure measures, most importantly Eloundou et al., 2024). Our estimated productivity gains from LLMs are highly correlated with both coarser human ratings of the share of time saved by LLMs exceeding 50 percent as equivalent ratings based on GPT-4 (see Panels A and B of Figure B.2). Second, we correlate our estimates with experimental estimates of worker time saved from randomized controlled trials (see Table A.5). Again, our estimates closely approximate those experimental estimates. Third, we compare our estimates with detailed data on which tasks the Anthropic's LLM "Claude" is used for in practice, finding that adoption is highest across tasks that we estimate to benefit from the highest productivity gains (see Panel C of Figure B.2).

Relative to previous work, we broaden the set of AI technologies considered beyond large language models to also include technologies such as AI-powered robots. AI-powered robots combine advanced algorithms and sensor technologies to perform tasks autonomously. In total, we consider three different AI technologies to assess all 19,530 occupational task descriptions, resulting in $19,530 \times 3 = 58,590$ requests to GPT-4o.

We also combine the task-level exposure measures to AI-related technologies with the task-level skill requirements. We then correlate the exposure to AI-related technologies with the skill requirements to understand how the exposure to AI-related technologies varies with the skill requirements of tasks (see Figure B.3). We find that tasks that require more analytical and cognitive skills are more exposed to AI technologies, while tasks that require more interpersonal and communication skills are less exposed to AI technologies. This is consistent with the idea that AI technologies are better at automating tasks that require analytical and cognitive skills than tasks that require interpersonal and communication skills.

3.4 Skills, Occupational Choice, and Wages (NLSY79)

We use the National Longitudinal Survey of Youth 1979 (NLSY79) to estimate the task production function and the human capital accumulation function. It contains data on wages, occupations, and skill assessments at the moment of entering the labor market.

We follow the literature in measuring skills in the NLSY79. As Addison et al. (2020); Baley et al. (2022), we measure the three cognitive dimensions of skill by the Armed Services Vocational Aptitude Battery (ASVAB): mathematics skills are measured as the average of the *mathematics knowledge* and *arithmetic reasoning* scores; technical skills are measured using *electronics information*, *general science*, and *mechanical comprehension*; verbal skills are the average of the scores on *paragraph comprehension* and *word knowledge*. We standardize each of the subscores before aggregating. For social skills, we use a compos-

ite measure of self-reported sociability as a young adult, sociability at age 6, the *Rotter Locus of Control Scale*, and the *Rosenberg Self-Esteem Scale* (see also Deming, 2017; Addison et al., 2020; Guvenen et al., 2020).

We also use the NLSY79 cohort to estimate the human capital accumulation function. To do so, we follow the labor market history of this cohort from age 25 to the survey in 2020. We retain information on all jobs held, including their start and end dates, the occupational code, the hourly wage, and the number of hours worked per week. Similar to Lise and Postel-Vinay (2020), we only consider workers for which the maximum gap between observed jobs is no larger than 18 months. We collapse this data to a worker panel of yearly frequency.

3.5 Occupational data

We use labor market data from the Bureau of Labor Statistics (BLS) to estimate the demand for occupational goods. Specifically, we use data on the share of total wages earned in each occupation from the 2018 Occupational Employment and Wage Statistics (OEWS) database from the Bureau of Labor Statistics (BLS). For 2018, the BLS OEWS data records employment, and the average and the variance of wages in 808 distinct occupations.[11] We merge these occupations to the 973 occupations in the O*NET task data. We drop 45 occupations—representing less than 1 percent of total employment—from the BLS data for which O*NET does not provide data on their task. Therefore, the total number of occupations in the model, J, is 763.

4 Estimation

We estimate and calibrate the parameters of the model using the data introduced in the previous section. We provide an overview of all model parameters in Table A.6. In this section, we discuss the procedure for each of the parameters in detail.

Externally set parameters. Some parameters are set externally. In sections 3.1 and 3.2 we explain how we measure the task set $\mathcal{T}_j$, the task weights $\theta_{j,\tau}$, and the skill requirements $\boldsymbol{r}(\tau)$ for each occupation j and for each task $\tau \in \mathcal{T}_j$. We set the discount factor β to 0.95 and the number of periods A to 40 so that each period in the model represents a year between ages 25 and 64. We use an elasticity of substitution between occupations σ of 1.57—in between estimates by Burstein et al. (2019) (1.81) and Caunedo et al. (2023) (1.34)—and an elasticity of substitution ρ between tasks of 0.49, as estimated by Humlum (2019).

[11] We use the year 2018 because the occupational classification can be matched particularly well to the O*NET database in that year.

Human capital accumulation and production. We estimate the parameters that govern production and human capital accumulation using direct inference on NLSY79 data. Specifically, we use a non-linear least squares regression of wages on skills and occupational fixed effects. The main logic of the estimation strategy is to find a set of parameters such that wage variation within occupation is consistent with worker's implied skill accumulation given their occupational history, the returns to skill, and the cost of mismatch.

The NLSY79 includes data on workers' initial skills (around age 20), along with their wages and occupational histories. The NLSY79 does not include direct information on a worker's skills at later ages. However, we use that a worker's skill trajectory can be inferred from their occupational history for a given set of parameters in the human capital accumulation function in (7). We thus solve for the parameters of the human capital accumulation function that best predict worker's wages given their initial skills and occupational history.

Formally, let $\boldsymbol{h}_{i,a}\left(\boldsymbol{h}_{i,1}, \boldsymbol{j}_i^{a-1} \mid \delta, \boldsymbol{\lambda}\right)$ be worker i's implied level of skill at age a given their initial skill $\boldsymbol{h}_{i,1}$, their employment history $\boldsymbol{j}_i^{a-1} \equiv \{j_{i,1}, \ldots, j_{i,a-1}\}$, and the parameters in the human capital accumulation equation (7) to be estimated, δ and $\boldsymbol{\lambda} \equiv \{\lambda_\psi\}$. Also, let $Y_j(\boldsymbol{h} \mid \boldsymbol{\omega}, \boldsymbol{\eta})$ be the production function in (6) given its parameters $\boldsymbol{\omega} = \{\omega_s\}_{s \in S}$ and $\boldsymbol{\eta} = \{\eta_s\}_{s \in S}$. We solve the following non-linear least squares problem:

$$\min_{\delta, \boldsymbol{\lambda}, \boldsymbol{\omega}, \boldsymbol{\eta}} \sum_i \sum_{a=1}^{A} \sum_{j=1}^{J} \mathbb{1}[j_{i,a} = j] \left(\log w_{ia} - \log p_j - \log Y_j\left(\boldsymbol{h}_{i,a}\left(\boldsymbol{h}_{i,1}, \boldsymbol{j}_i^{a-1} \mid \delta, \boldsymbol{\lambda}\right) \mid \boldsymbol{\omega}, \boldsymbol{\eta}\right)\right)^2 \quad (12)$$

where w_{ia} is worker i's wage at age a and $\mathbb{1}[\cdot]$ is the indicator function.

Solving the problem with "brute force" is computationally hard because the parameter space is large and each function evaluation involves iterating over the human capital accumulation function for each worker. Thankfully, the computational complexity of the problem in (12) can be significantly reduced by realizing that $\log Y_j(\cdot)$ is approximately linear. Specifically, we use that a log-linearization of the production function in equation (6) (around no mismatch) implies that

$$\log Y_j(\boldsymbol{h}) \approx \log p_j + \sum_{s \in S} \omega_s \log(h_{is}) - \sum_{s \in S} \eta_s \underbrace{\sum_{\tau \in \mathcal{T}_j} \theta_{j,\tau} \left(h_{is} - r_s(\tau)\right)^2}_{\text{Skill mismatch in skill } s}. \quad (13)$$

Equation (13) suggests a simple and intuitive production function estimation procedure of the production function: a log-linear regression of wages on skills, skill mismatch, and fixed effects for each occupation. This means that the production function can be "concentrated out" of the problem in equation (12): for a given set of human capital accumulation parameters, δ and $\boldsymbol{\lambda}$, we can compute the implied path of skills and estimate

the parameters that govern production, $\hat{\omega}$ and $\hat{\eta}$, in closed form.[12]

While we estimate the parameters in $\boldsymbol{\lambda}$ jointly, the identification of each of them is intuitive: the human capital depreciation rate δ is identified from wages of workers with a history of overqualification, while $\{\lambda_\psi\}$ is identified from the wages of workers with different "ability to learn". Following (Heckman et al., 1998), we proxy the ability to learn by a worker's Armed Forces Qualification Test (AFQT) score before entering the labor market. That is, we estimate the learning parameter λ_ψ separately for each AFQT quartile.

The estimated parameters of the human capital accumulation function are shown in Table 1. The depreciation rate of human capital when doing work for which one is overqualified is 1.4 percent. λ_q denotes the learning parameter of quartile q in the AFQT distribution. The higher λ_q, the less a worker learns in a given occupation for a given set of skills. Therefore, the results in Table 1 show that the ability to learn monotonically decreases with the AFQT score.

TABLE 1: Human Capital Depreciation and Accumulation: Parameter Estimates

δ	λ_1	λ_2	λ_3	λ_4
0.014	14.6	12.6	11.8	10.4

Notes: Estimates are obtained through the non-linear least squares regression in equation (12) using data on the National Longitudinal Survey of Youth 1979.

Table 2 shows the results of the estimation of the corresponding production function in equation (13). The first row presents the estimates of ω, the elasticity of wages to skills, holding mismatch constant. The second row contains the estimates of η_s, the elasticity of wages to skill mismatch. For the purpose of model calibration, we use the estimates in the first column, so that $\hat{\omega}_s = 0.488$, $\hat{\eta}_s = 0.053$. In Table A.7, we show estimates of the same regressions when using the NLSY79 cohort. The parameters are comparable in magnitude to those in the NLSY97 cohort.

[12]The log-linear production function also provides a theoretical underpinning of the validation exercise of our new database of *task-level* skill requirements $r_s(\tau)$ in section 3.2. To a first-order approximation, the level of skill that minimizes mismatch—and thus maximizes comparative advantage—in an occupation is simply the weighted average of skill requirements across tasks in that occupation:

$$h^*_{j,s} \equiv \arg\min_{h_s} \sum_{\tau \in \mathcal{T}_j} \theta_{j,\tau} \left(h_s - r_s(\tau)\right)^2 = \sum_{\tau \in \mathcal{T}_j} \theta_{j,\tau} r_s(\tau).$$

TABLE 2: Production Function Estimation: Wages and Skill Mismatch

Name	Parameter	Dependent variable: Log hourly wage			
Log skill: $\log(h_s)$					
Math	ω_M	0.279***	0.339***	0.384***	0.335***
		(0.010)	(0.014)	(0.015)	(0.021)
Social	ω_S	0.432***	0.455***	0.376***	0.440***
		(0.010)	(0.014)	(0.011)	(0.016)
Technical	ω_T	0.147***	0.137***	0.083***	0.157***
		(0.017)	(0.023)	(0.023)	(0.030)
Verbal	ω_V	0.080***	-0.049**	0.044**	-0.095***
		(0.017)	(0.022)	(0.022)	(0.032)
Mismatch: $\sum_\tau \theta_{j,\tau} \left(h_s - r_s(\tau)\right)^2$					
Math	η_M	0.044	0.302***	-0.119	-0.268**
		(0.065)	(0.085)	(0.080)	(0.110)
Social	η_S	-1.266***	-1.746***	-0.722***	-1.035***
		(0.108)	(0.153)	(0.114)	(0.163)
Technical	η_T	-1.494***	-1.834***	-0.529**	-0.021
		(0.197)	(0.254)	(0.206)	(0.263)
Verbal	η_V	-1.886***	-1.763***	-0.965***	-0.766***
		(0.139)	(0.186)	(0.137)	(0.197)
Cohort		NLSY79	NLSY79	NLSY97	NLSY97
Sample		All	Men	All	Men
Observations		62,413	35,427	29,941	15,208
R^2		0.47	0.47	0.46	0.45

Notes: Data are from the National Longitudinal Survey of Youth 1997. Standard errors are clustered at the individual level. Sample weights are applied. All regressions include age, sex, race, and year fixed effects. The unit of analysis is a worker-job-year combination. We exclude self-employment and only include jobs where the worker is in between 25 and 34 years old and works in between 20 and 80 hours per week in that job. We only include observations where the skill mismatch is below median. Occupational skill requirements refer to $\sum_{\tau \in \mathcal{T}_j} \theta_{j,\tau} r_s(\tau)$, which we include as controls where indicated. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Consistent with the model, we find that the estimated occupational prices $\hat{p}_j$ correlate

strongly with the skill requirements in the respective occupations. In the model, the set of tasks and their skill requirements completely define an occupation. Therefore, the prices of each occupation, must be a (potentially highly non-linear) function of the task-level skill requirements in that occupation. Table A.8 shows that skill requirements are indeed a good predictor of occupational prices ($R^2 = 0.57$), even when prices are restricted to be a simple linear function of the occupational skill requirements.

Preferences for occupations and demand. We estimate the distribution of preferences for occupations in the older population of the NLSY. In the model, old workers (with age $a = A$) only takes into account their wages and preferences when choosing job, not the job's learning potential. Therefore, using this population, we can estimate preferences with observed occupational choices and the counterfactual wages in occupations that the worker did not choose. The counterfactual wages, can be constructed using the parameters of the production and human capital accumulation function estimated above. The Gumbel distribution of the preference shocks means that we can estimate the job amenities and scale parameter of the preference shocks using a conditional logit regression of a dummy indicating whether the worker choose that occupation on the (potentially counterfactual) wage and occupational fixed effects.

Lastly, we estimate the demand shifters $\{\alpha_j\}_{j=1}^{J}$ for each occupation using the demand equation in (8). That is, we estimate the shifters as

$$\frac{\alpha_i}{\alpha_j} = \left(\frac{p_i}{p_j}\right)^{\sigma-1} \times \frac{\text{Wage share of occupation } i}{\text{Wage share of occupation } j}.$$

Since we estimated occupational prices and wage shares can be readily computed, this concludes the estimation.

5 AI's Impact on the Labor Market

5.1 Wages and Wage Inequality

5.2 Winners and Losers

5.3 Returns to Skills

5.4 Is Occupational "Exposure" Informative for Gains from AI?

6 Conclusion

References

ACEMOGLU, D. AND D. AUTOR (2011): "Skills, tasks and technologies: Implications for employment and earnings," in *Handbook of labor economics*, Elsevier, vol. 4, 1043–1171.

ACEMOGLU, D., D. AUTOR, J. HAZELL, AND P. RESTREPO (2022): "Artificial intelligence and jobs: Evidence from online vacancies," *Journal of Labor Economics*, 40, S293–S340.

ACEMOGLU, D. AND P. RESTREPO (2018): "The race between man and machine: Implications of technology for growth, factor shares, and employment," *American economic review*, 108, 1488–1542.

——— (2020): "Robots and jobs: Evidence from US labor markets," *Journal of political economy*, 128, 2188–2244.

ADDISON, J. T., L. CHEN, AND O. D. OZTURK (2020): "Occupational skill mismatch: Differences by gender and cohort," *ILR Review*, 73, 730–767.

AUTOR, D. H. AND M. J. HANDEL (2013): "Putting tasks to the test: Human capital, job tasks, and wages," *Journal of labor Economics*, 31, S59–S96.

AUTOR, D. H., F. LEVY, AND R. J. MURNANE (2003): "The skill content of recent technological change: An empirical exploration," *The Quarterly journal of economics*, 118, 1279–1333.

BALEY, I., A. FIGUEIREDO, AND R. ULBRICHT (2022): "Mismatch cycles," *Journal of Political Economy*, 130, 2943–2984.

BICK, A., A. BLANDIN, AND D. J. DEMING (2024): "The Rapid Adoption of Generative AI," Working Paper 32966, National Bureau of Economic Research.

BOPPART, T., P. KRUSELL, AND K. MITMAN (2018): "Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative," *Journal of Economic Dynamics and Control*, 89, 68–92.

BRYNJOLFSSON, E., D. LI, AND L. R. RAYMOND (2023): "Generative AI at Work," Working Paper 31161, National Bureau of Economic Research.

BRYNJOLFSSON, E., T. MITCHELL, AND D. ROCK (2018): "What Can Machines Learn, and What Does It Mean for Occupations and the Economy?" *AEA Papers and Proceedings*, 108, 43–47.

BURSTEIN, A., E. MORALES, AND J. VOGEL (2019): "Changes in between-group inequality: computers, occupations, and international trade," *American Economic Journal: Macroeconomics*, 11, 348–400.

CAMPERO, A., M. VACCARO, J. SONG, H. WEN, A. ALMAATOUQ, AND T. MALONE (2022): "A Test for Evaluating Performance in Human-Computer Systems," Working paper.

CAUNEDO, J., D. JAUME, AND E. KELLER (2023): "Occupational exposure to capital-embodied technical change," *American Economic Review*, 113, 1642–1685.

CLARKE, S. AND B. HANRAHAN (2024): "Comparing the Effect of Different Task Types on Effective Use of GitHub Copilot," in *Generative AI in Real-World Workplaces: The Second Microsoft Report on AI and Productivity Research*, ed. by S. Jaffe, N. P. Shah, J. Butler, A. Farach, A. Cambon, B. Hecht, M. Schwarz, and J. Teevan.

CUI, Z. K., M. DEMIRER, S. JAFFE, L. MUSOLFF, S. PENG, AND T. SALZ (2024): "The Effects of Generative AI on High Skilled Work: Evidence from Three Field Experiments with Software Developers," Working paper.

DELL'ACQUA, F., E. MCFOWLAND III, E. R. MOLLICK, H. LIFSHITZ-ASSAF, K. KELLOGG, S. RAJENDRAN, L. KRAYER, F. CANDELON, AND K. R. LAKHANI (2023): "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013, The Wharton School Research Paper.

DELOACH, S. B., M. KURT, AND R. SANSALE (2022): "Non-cognitive mismatch and occupational switching," *Journal of Behavioral and Experimental Economics*, 97, 101846.

DEMING, D. J. (2017): "The growing importance of social skills in the labor market," *The quarterly journal of economics*, 132, 1593–1640.

EISFELDT, A. L., G. SCHUBERT, AND M. B. ZHANG (2023): "Generative AI and firm values," Tech. rep., National Bureau of Economic Research.

ELOUNDOU, T., S. MANNING, P. MISHKIN, AND D. ROCK (2024): "GPTs are GPTs: Labor market impact potential of LLMs," *Science*, 384, 1306–1308.

GRAETZ, G. AND G. MICHAELS (2018): "Robots at work," *Review of Economics and Statistics*, 100, 753–768.

GUVENEN, F., B. KURUSCU, S. TANAKA, AND D. WICZER (2020): "Multidimensional skill mismatch," *American Economic Journal: Macroeconomics*, 12, 210–244.

HECKMAN, J. J., L. LOCHNER, AND C. TABER (1998): "Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents," *Review of economic dynamics*, 1, 1–58.

HUMLUM, A. (2019): "Robot adoption and labor market dynamics," *Princeton University*.

KANAZAWA, K., D. KAWAGUCHI, H. SHIGEOKA, AND Y. WATANABE (2022): "AI, Skill, and Productivity: The Case of Taxi Drivers," Working Paper 30612, National Bureau of Economic Research.

KEANE, M. P. AND K. I. WOLPIN (1997): "The career decisions of young men," *Journal of political Economy*, 105, 473–522.

LEE, D. AND K. I. WOLPIN (2006): "Intersectoral labor mobility and the growth of the service sector," *Econometrica*, 74, 1–46.

LINDENLAUB, I. (2017): "Sorting Multidimensional Types: Theory and Application," *The Review of Economic Studies*, 84, 718–789.

LISE, J. AND F. POSTEL-VINAY (2020): "Multidimensional skills, sorting, and human capital accumulation," *American Economic Review*, 110, 2328–2376.

LLULL, J. (2018): "Immigration, wages, and education: A labour market equilibrium structural model," *The Review of Economic Studies*, 85, 1852–1896.

NOY, S. AND W. ZHANG (2023): "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence," *Science*, 381, 187–192.

PENG, S., E. KALLIAMVAKOU, P. CIHON, AND M. DEMIRER (2024): "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot," Working paper.

VAITHILINGAM, P., T. ZHANG, AND E. GLASSMAN (2022): "Expectation vs. Experience: Evaluating the Usability of Code Generation Tools Powered by Large Language Models," Working paper.

WEBB, M. (2020): "The Impact of Artificial Intelligence on the Labor Market," Stanford University.

ZEIRA, J. (1998): "Workers, machines, and economic growth," *The Quarterly Journal of Economics*, 113, 1091–1117.

Appendix

A Tables **21**

B Figures **28**

C Data **30**

C.1 Measures of AI Exposure 30

C.1.1 Task-Based Measures 30

C.1.2 Ability-Based Measures 32

C.1.3 Occupation-Based Measures 33

C.2 Measures of Computer Automation 34

C.2.1 Occupation-Based Measures 34

C.2.2 Task-Based Measures 35

C.3 Measures of Robotic Automation 37

C.4 O*NET Database 38

C.5 Prompts 38

C.5.1 Skill-requirements by task 38

C.5.2 Productivity of AI-related technologies by task 38

D Estimation **40**

D.1 Log-linearization of production function 40

A Tables

TABLE A.1: Aggregation of O*NET's skill requirements to 4 dimensions

Skill	O*NET skill	O*NET skill category
Math	Mathematics	Basic Content
Social	Active Listening	Basic Content
	Coordination	Social
	Instructing	Social
	Management of Personnel Resources	Resource Management
	Negotiation	Social
	Persuasion	Social
	Service Orientation	Social
	Social Perceptiveness	Social
Technical	Complex Problem Solving	Complex Problem Solving
	Equipment Maintenance	Technical
	Equipment Selection	Technical
	Installation	Technical
	Judgment and Decision Making	Systems
	Operation and Control	Technical
	Operations Analysis	Technical
	Operations Monitoring	Technical
	Programming	Technical
	Quality Control Analysis	Technical
	Repairing	Technical
	Science	Content
	Systems Analysis	Systems
	Systems Evaluation	Systems
	Technology Design	Technical
	Troubleshooting	Technical
Verbal	Reading Comprehension	Basic Content
	Speaking	Basic Content
	Writing	Basic Content

Notes: This table shows the mapping of the four skill clusters—Math, Social, Verbal, and Technical—to the relevant O*NET skills and their respective O*NET's skill category. For each of the four skills, we set the requirement to the average across the relevant O*NET skills. We dropped 7 out of 35 O*NET skill dimensions that could not be clearly mapped into the four skills used in the analysis.

TABLE A.2: Overview of Task-Level Exposure Measures Across Technologies

Technology	Measure description	Scale	Source
LLMs	GPT-4 rated exposure to LLMs (% automated)	%-bin	[1]
LLMs	GPT-4 rated exposure to LLMs alone ($\geq$50% time saved)	Yes/No	[1]
LLMs	GPT-4 rated exposure to LLMs + software ($\geq$50% time saved)	Yes/No	[1]
LLMs	Human-rated exposure to LLMs alone ($\geq$50% time saved)	Yes/No	[1]
LLMs	Human-rated exposure to LLMs + software ($\geq$50% time saved)	Yes/No	[1]
LLMs	Exposure based on overlap with AI patents	Percentile	[2]
ML	Suitability for Machine Learning (SML) scores	1–5 scale	[3]
Software	Exposure based on overlap with software patents	Percentile	[2]
Robots	Exposure based on overlap with robotics patents	Percentile	[2]

Notes: [1] Eloundou et al. (2024), [2] Webb (2020), [3] Brynjolfsson et al. (2018). Webb (2020) measures a task's exposure to a given technology via the overlap between the text of patents related to the technology and text on occupational task descriptions. While the estimates are generated at the task-level, the author only provides occupation-level aggregates. Measures by Brynjolfsson et al. (2018) are based on 23-item questionnaire. The coarse %-bins provided by Eloundou et al. (2024) are 0%, 0%–50%, 50%–90%, 90%–100%, 100%.

TABLE A.3: New measures of task-level productivity gains from AI-related technologies

	LLMs	**AI-Powered Robots**	**AI-Powered Robots + LLMs**
Mean	18.5%	44.6%	
Standard Deviation	9.9%	16.4%	
Median	20.0%	40%	
Range	0.0% - 60.0%	0.0% – 100.0%	
Rater	GPT-4o	GPT-4o	GPT-4o
Number of Tasks	19,530	19,530	19,530
Number of Occupations	16,210	16,210	16,210

Notes: This table summarizes our new estimates of workers' time saved by AI-related technologies. We consider (1) state-of-the-art large language models (**LLMs**), (2) **AI-Powered Robots**, and the combination of (1) and (2). Estimates are generated by GPT-4o (September 2024 version) and measured as the time a worker can save in completing the task by getting access to a given technology.

TABLE A.4: Tasks with highest vs. lowest productivity gains from LLMs

Time saved	Task	Occupation
Highest LLM Exposure		
60%	Type acknowledgment letters to persons sending correspondence.	Correspondence Clerks
50%	Compile bibliographies of specialized materials for outside reading assignments.	Forestry & Conservation Science Teachers, Postsecondary
40%	Prepare responses to correspondence containing routine inquiries.	Executive Secretaries & Executive Administrative Assistants
40%	Locate unusual or unique information in response to specific requests.	Librarians
40%	Adjust sales scripts to better target the needs and interests of specific individuals.	Telemarketers
Lowest LLM Exposure		
0%	Stack and sort cut material for packaging, further processing, or shipping, according to types and sizes of material.	Cutting & Slicing Machine Setters, Operators, & Tenders
0%	Press thumbs into centers of revolving clay to form hollows, and press on the inside and outside of emerging clay cylinders with hands and fingers, gradually raising and shaping clay to desired forms and sizes.	Potters, Manufacturing
0%	Remove materials and products from machines and equipment, and place them in boxes, trucks or conveyors, using hand tools and moving devices.	Machine Feeders & Offbearers
0%	Operate controls to open and close transit vehicle doors.	Subway & Streetcar Operators
0%	Maneuver trucks into loading or unloading positions, following signals from loading crew and checking that vehicle and loading equipment are properly positioned.	Heavy & Tractor-Trailer Truck Drivers

Notes: This table shows the 5 tasks with the highest vs. lowest LLM exposure along with the corresponding occupations. When tasks' ranks are tied, we show examples across different occupations (e.g., there are 2,291 tasks with 0% exposure). LLM exposure is GPT-4o generated and measured as the time a worker can save in completing the task by getting access to an LLM.

TABLE A.5: Experimental Estimates vs. Our Estimates of LLMs' Impact

Occupation	Work activity	AI	Productivity change		N	Notes	Source
			Estimate	Our estimate			
Software developer	Coding	GitHub Copilot	26%	26%	4,867	+26.1% number of completed tasks in lab; new developers higher adoption rates & higher productivity gains	[1]
Software developer	Coding	GitHub Copilot	56%	26%	95	55.8% time saved, quality ↑	[2]
Software developer	Coding	GitHub Copilot	36%	26%	23	36% time saved for familiar tasks; no change for unfamiliar tasks; 48% fewer issues	[3]
Programmer	Coding	GPT-3	27%	26%	100	27% time saved among 100 expert programmers; 50 non-programmers can perform tasks similarly well with GPT-3	[4]
Programmer	Coding	GitHub Copilot	0%	26%	24	No time saved; however, most participants still preferred using Copilot	[5]
Management consultant	Consulting	GPT-4	25%	29%	758	25.1% time saved, +12.2% tasks completed, +40% quality (decreased for tasks beyond AI frontier); lower-skilled consultants benefited more	[6]
Customer support agent	Resolution	GPT-4	14%	29%	5,179	+14% productivity (issues resolved per hour), +34% for new & low-skill workers; minimal impact on experienced & high-skill workers	[7]
–	Writing	GPT-3.5	40%	29%	453	40% time saved, +18% output quality; inequality between workers ↓; low-skill workers benefited most; likelihood of using AI after experiment ↑	[8]
Taxi driver	Selecting routes	AI Navi	14%	9%	520	Shorter cruising time; gains only among low-skill drivers	[9]

Notes: Sources correspond to [1] Cui et al. (2024), [2] Peng et al. (2024), [3] Clarke and Hanrahan (2024), [4] Campero et al. (2022), [5] Vaithilingam et al. (2022), [6] Dell'Acqua et al. (2023), [7] Brynjolfsson et al. (2023), [8] Noy and Zhang (2023), [9] Kanazawa et al. (2022). To construct our own estimates of productivity change at the level of work activities, we aggregate our task-level productivity estimates within the relevant occupation as equally weighted averages for tasks we judge to be relevant to the work activity covered in each experiment.

TABLE A.6: Overview of model parameters

Model object	**Symbol**	**Value**	**How it is set**
Elasticity of substitution: goods	σ	1.57	Set in between (Burstein et al., 2019) (1.81) and Caunedo et al. (2023) (1.34).
Elasticity of substitution: tasks	ρ	0.49	Calibrated to Humlum (2019).
Number of occupations	J	763	Set to BLS OEWS data in 2018.
Number of periods	A	40	Years between 25 and 65.
Discount factor	β	0.95	Following Lee and Wolpin (2006).
Number of skill dimensions	S	4	Following Baley et al. (2022).
Occupational task sets	$\mathcal{T}_j$		O*NET task sets by occupation.
Occupational task weights	$\theta_{j,\tau}$		O*NET task importance.
Task-level skill requirements	$\boldsymbol{r(\tau)}$		Using large language model.
Task-level AI productivity gains	γ_τ		"
Returns to skill	ω	0.29	Log-linear regression using NLSY.
Skill mismatch sensitivity	η	0.61	"
Human capital depreciation	δ	0.01	Non-linear regression using NLSY.
Learning: 1st AFQT quartile	λ_1	14.58	"
Learning: 2nd AFQT quartile	λ_2	12.64	"
Learning: 3rd AFQT quartile	λ_3	11.77	"
Learning: 4th AFQT quartile	λ_4	10.44	"
Scale parameter of taste shock	ζ		Internally estimated to match wage distribution and occupational wage shares.
Occupational demand	$\{\alpha_j\}_{j=1}^{J}$		"

Notes: This table provides an overview of the parameters of the model, their mathematical symbols, the value at which they are set, and the procedure with which we arrived at the value.

TABLE A.7: Production Function Estimation: Wages and Skill Mismatch in the NLSY79

		Dependent variable: Log hourly wage			
Name	Symbol				
Log skill	$\log(\bar{h})$	0.395***	0.440***	0.440***	0.483***
		(0.028)	(0.020)	(0.030)	(0.022)
Skill mismatch	$\sum_s \sum_\tau \theta_{j,\tau} (h_{is} - r_{is}(\tau))^2$	-0.534***	-0.510***	-0.653***	-0.552***
		(0.124)	(0.090)	(0.129)	(0.095)
Additional controls					
Occupational fixed effects		✓	✓		
Occupational skill requirements				✓	✓
Sample		Men	All	Men	All
Observations		35,359	66,613	35,369	66,619
R^2		0.34	0.39	0.26	0.31

Notes: Data are from the National Longitudinal Survey of Youth 1979. Standard errors are clustered at the individual level. Sample weights are applied. All regressions include age, sex, race, and year fixed effects. The unit of analysis is a worker-job-year combination. We exclude self-employment and only include jobs where the worker is in between 25 and 34 years old and works in between 20 and 80 hours per week in that job. We only include observations where the skill mismatch is below median. Occupational skill requirements refer to $\sum_{\tau \in \mathcal{T}_j} \theta_{j,\tau} r_s(\tau)$, which we include as controls where indicated. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

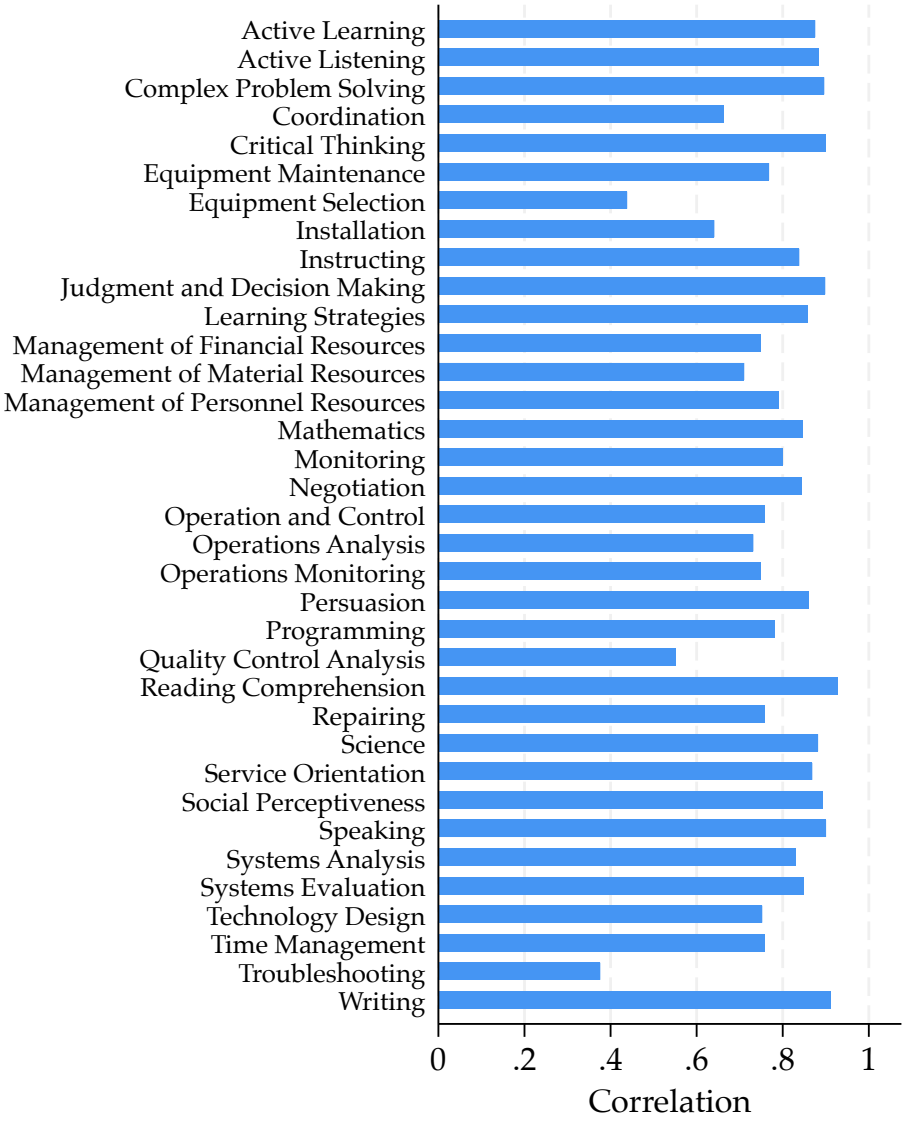
TABLE A.8: Occupational Skill Requirements Predict Occupational Prices

Occupational skill requirements	Dependent variable: Log occupational price $\hat{p}_j$			
Math	0.792***	0.522***	0.764***	0.441***
	(0.217)	(0.183)	(0.166)	(0.162)
Social	0.530	0.411*	0.222	-0.168
	(0.343)	(0.239)	(0.305)	(0.222)
Technical	1.030***	1.139***	0.465**	0.877***
	(0.247)	(0.215)	(0.209)	(0.197)
Verbal	0.021	0.368	0.313	0.893***
	(0.348)	(0.248)	(0.300)	(0.224)
NLSY Cohort	1997	1997	1979	1979
Sample	Men	All	Men	All
Observations	173	246	191	236
R^2	0.63	0.63	0.58	0.61

Notes: This table shows the coefficients and R^2 of a regression of the estimated occupational prices (in logs) on occupational skill requirements. Each observation represents one occupation and is weighted by the number of worker-job-year observations in that occupation. Only occupations with at least 20 observations are included. Occupational prices are estimated as the occupational fixed effects in regression equation (13) with coefficients shown in Table 2 (for NLSY97) and Table A.7 (for NLSY79). Occupational skill requirements refer to $\sum_{\tau \in \mathcal{T}_j} \theta_{j,\tau} r_s(\tau)$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

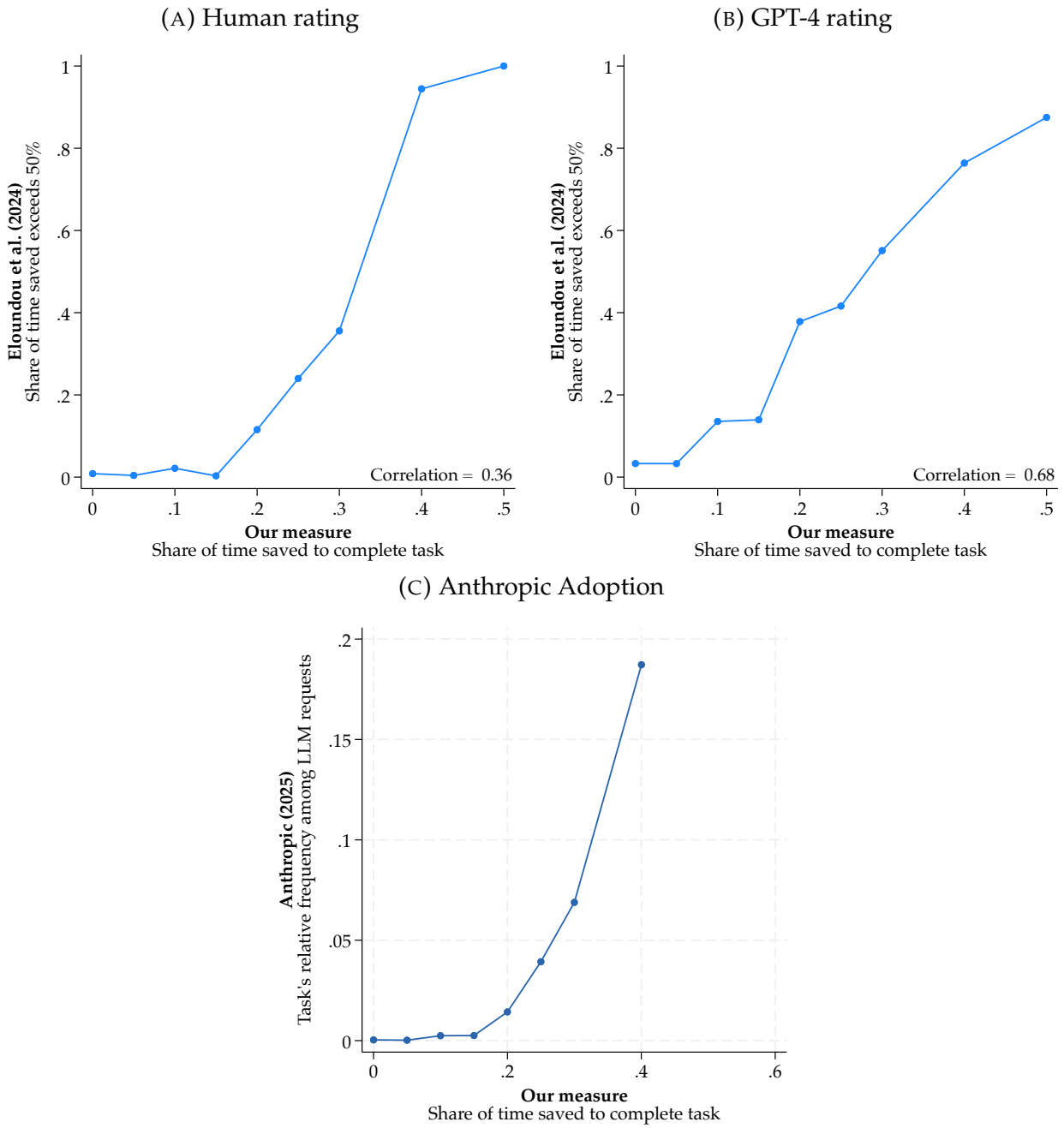
B Figures

FIGURE B.1: Validation of task skill requirement data with O*NET



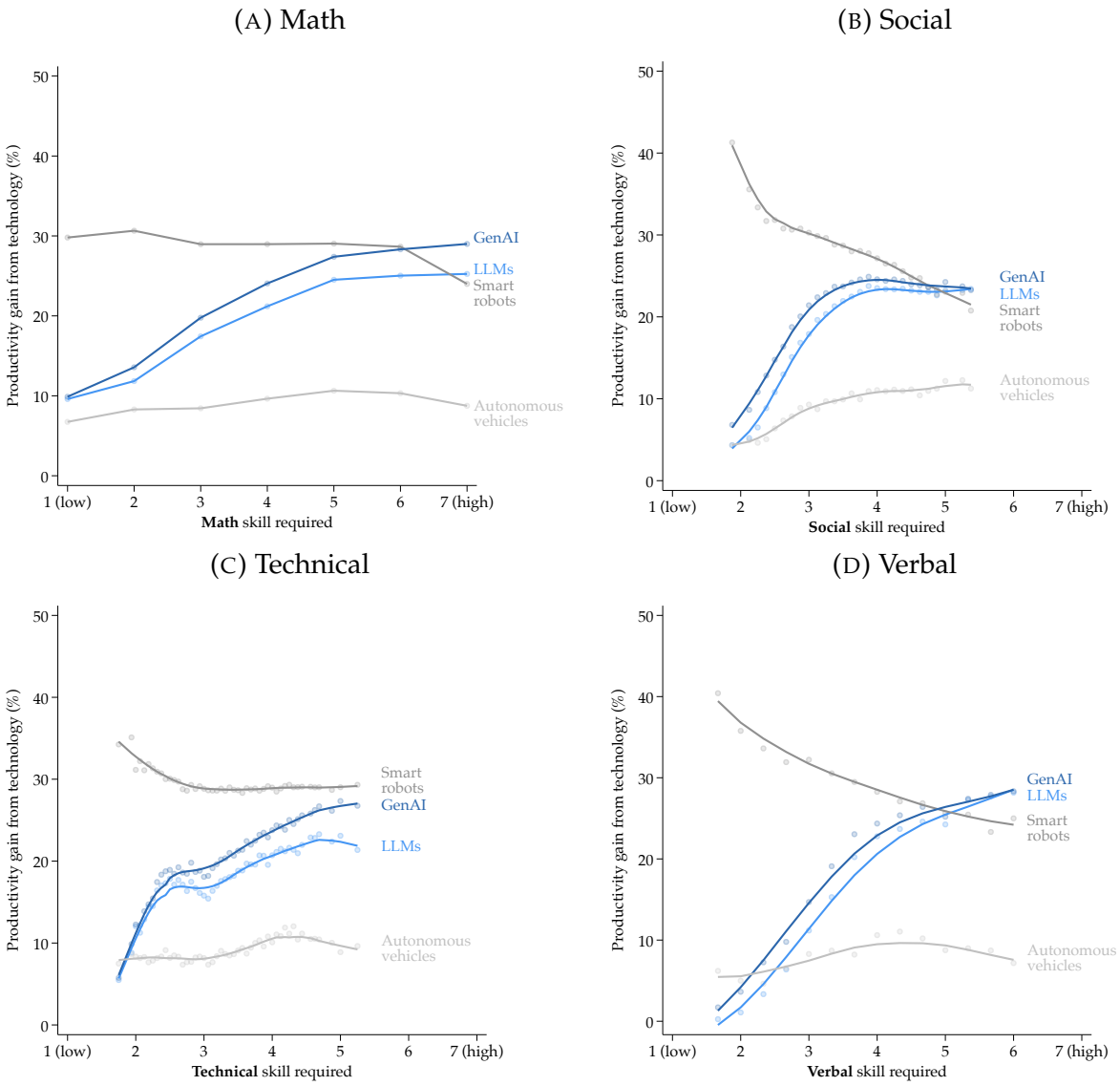
Notes: This figure shows the correlations between the occupation-level skill requirement in the O*NET database and the GPT-4o generated task-level skill requirements aggregated to the occupation-level for each of the 35 skills.

FIGURE B.2: Validation of Our Estimated Productivity Gains from LLMs



Notes: This figure compares our estimates of productivity gains from LLMs to those provided by Eloundou et al. (2024) as well as adoption rates in Anthropic's LLM "Claude" (Handa et al., 2024). Our measure estimates the share of worker time saved through access to LLMs across tasks, which we group into deciles for this figure. Existing data estimates the share of tasks for which LLMs save at least 50 percent of worker time, rated by humans (Panel A) or OpenAI's GPT-4 (Panel B). Panel C compares our data to Anthropic's adoption, measured as each task's prevalence in prompts submitted to Claude, normalized by the employment of tasks' corresponding occupations.

FIGURE B.3: Task-level correlation between skills & productivity gains from technology



Notes: This figure shows the correlation between a task's skill requirements and the productivity boost that different technologies create for workers completing each task. The three technologies that we consider are robots, LLMs, and AI-powered robots. Each dot represents the average productivity boost among tasks with the same requirement in a given skill. Lines show local linear averages.

C Data

C.1 Measures of AI Exposure

C.1.1 Task-Based Measures

1. **Survey responses (Eloundou et al., 2024).** Exposure of a task or Detailed Worker Activity (DWA) to LLMs or LLM-powered systems as judged by humans.

   Details: Authors administered a rubric of exposure to AI on humans. Exposure was defined as whether "access to an LLM or LLM-powered system" could at least

halve the time required for a human to perform each of the 2,087 Detailed Worker Activities (DWAs) plus a subset of the 19,265 tasks in the O*NET database (see Appendix B for an example of the O*NET classification of an economist). Human annotators were familiar with LLMs (having annotated previous OpenAI work) and were not representative of the occupational diversity in the O*NET dataset.

Examples of Occupations, DWAs and Tasks are included below. The Gambling Cage Workers versus Online Merchants example explains the need to include a DWA and task hierarchy:

TABLE C.9: Sample O*NET Occupations, DWAs and Tasks

Occupation	DWA	Task
Computer Systems Engineers/Architects	Monitor computer system performance to ensure proper operation	Monitor system operation to detect potential problems
Gambling Cage Workers	Execute sales or other financial transactions	Cash checks and process credit card advances for patrons
Online Merchants	Execute sales or other financial transactions	Deliver e-mail confirmation of completed transactions and shipment

Scale: Each individual DWA or task was classified into a level of exposure: no exposure (E0 = 0), direct exposure (E1 = 1), or exposure only if additional software is developed on top of the LLM (E2 = 1). These were then converted into one of three weighted $0-1$ values: $\alpha = E1$, $\beta = E1 + 0.5 * E2$, and $\zeta = E1 + E2$. The rest of the paper uses β scores where direct exposure counts twice as much as indirect exposure.

Data URL: N/A

2. **GPT-4 responses (Eloundou et al., 2024).** Exposure of a task to LLMs or LLM-powered systems as judged by OpenAI's large language model "GPT-4."

   Details: Authors administered the same rubric (with modified language to optimize for model response) to evaluate the exposure of occupational tasks to AI by "prompting" the GPT-4 API. Exposure was defined as whether "access to an LLM or LLM-powered system" could at least halve the time required for a human to complete each of 19,265 tasks in the O*NET database. Note that unlike the human survey above, this used tasks only, and not DWAs.

   The prompt defines the exposure taxonomy (E0, E1, E2) and specifies the AI as the "most powerful OpenAI large language model". The taxonomy is then followed by examples of annotation responses the model should provide for occupation tasks or activities:

   "Occupation: Computer and Information Research Scientists Task: Apply theoretical expertise and innovation to create or apply new technology, such as adapting principles for applying computers to new uses. Label (E0/E1/E2/E3): E1 Explanation: The model can learn theoretical expertise during training as part of its general knowledge base, and the principles to adapt can be captured in the text input to the model.

Activity: Schedule dining reservations. Label (E0/E1/E2/E3): E2 Explanation: Automation technology already exists for this (e.g. Resy) and it's unclear what an LLM offers on top of using that technology (no-diff). That said, you could build something that allows you to ask the LLM to make a reservation on Resy for you. (E3)"

Scale: The same scale and weights as the human survey was used. GPT-4 agreed with the human annotation of exposure at the task level 80.8%, 65.6% and 82.1% of the times under α, β and ζ scores respectively.

Data URL: N/A

3. **Survey responses (Brynjolfsson et al., 2018).** Suitability of ML to perform tasks as judged on a Likert-like scale by humans.

   Details: Authors crowd-sourced answers to a rubric (from Brynjolfsson and Mitchell (2017)) that evaluated potential for applying machine learning (ML) to tasks of the 2,069 work activities, 18,156 tasks and 964 occupations in O*NET at the time. They build measures of "suitability for machine learning" (SML) of a given task in an occupation. The rubric focuses on questions of technical feasibility, not economic, organizational, legal or cultural implications of ML adoption.

   Scale: 1-5 scale where 5 represents the highest SML. It is calculated from the mean score from the rubric's 23 statements for each DWA. Each statement was to be evaluated on a 5-point scale that ranges from a "strongly disagree" to "strongly agree" statement. A sample statement taken from Brynjolfsson and Mitchell (2017)'s supplemental materials is:

   "Statement: Information needed to complete the task (inputs) and outputs can be explicitly specified in machine-readable format:

   1: It is very difficult or impossible to identify particular inputs and outputs (e.g. emotions, ideas, impressions)

   3: It is possible to create rankings or partial representability of inputs and outputs

   5: It is easy to quantify results on a machine/computer (e.g. calculations, concrete inputs and outputs)

   Data URL: `https://www.openicpsr.org/openicpsr/project/114436/version/V1/view`

C.1.2 Ability-Based Measures

1. **Mapping AI task performance to human abilities (Felten and Raj, 2018).** Mapping AI task-specific performance metrics to job abilities and aggregating these at the occupation level.

Details: The authors used the Electronic Frontier Foundation (EFF) AI Progress Measurement dataset that tracks the progress of AI performance in categories such as abstract strategy games and image recognition. The EFF AI categories were mapped to 52 O*NET abilities, which are used to describe job requirements. The mapping was done by PhD students. They then aggregate the AI progress score into each occupation based on a weighed prevalence and importance of each ability to a given job.

Scale: An effect score (0-7) is used. This is a relative scale for comparison purposes only. High scores are occupations with a high prevalence of abilities where AI progress is high.

Data URL: N/A

C.1.3 Occupation-Based Measures

1. **Aggregation of task-based measures (Eloundou et al., 2024).** The weighed average exposure of an occupation's tasks to AI.

   Details: The authors' task-based measures described in A.1.1 are aggregated into their corresponding occupations to measure the overall exposure of given occupation.

   Scale: 0-100% exposure of occupation by aggregating tasks associated with that occupation. The authors use the β score for a given task (described in A.1.1), and further weigh the occupation's exposure by giving core tasks twice the weight assigned to supplemental tasks. Core/supplement distinctions are determined by O*NET.

   Data URL: N/A

2. **Aggregation of ability-based measures (Felten et al., 2021).** The ability's prevalence- and importance-weighed average exposure of an occupation's tasks to AI.

   Details: The authors build upon the methodology in Felten and Raj (2018) mapping the AI Progress Measurement dataset from the Electronic Frontier Foundation (EFF) to 52 abilities in the O*NET database. They survey Amazon Mechanical Turk users to generate a measure of relatedness between AI application and job ability.

   Scale: An ability-level exposure score is given by summing the 0-1 measure of AI application i to O*NET job ability j:

$$A_{ij} = \sum_{i=0}^{1} 0 x_{ij}$$

The Artificial Intelligence Occupational Exposure (AOIE) score is the sum of ability-level exposure A_{ij} to AI application i for abilities j in occupation k, weighed by the ability's prevalence (L_{jk}) and importance (I_{jk}) of occupation given by O*NET:

$$\text{AIOE}_k = \frac{\sum_{j=1}^{52} A_{ij} \times L_{jk} \times I_{jk}}{\sum_{j=1}^{52} L_{jk} \times I_{jk}}$$

Data URL: https://github.com/AIOE-Data/AIOE and Felten et al. (2023) applies metric to measure occupational exposure to generative AI, with results in appendix.

3. **Complementarity-adjusted AOIE (Li, 2023).** Adjust the AOIE measure to account for human-AI complementarity that mitigates labor substitution.

   Details: The authors build on top of the AOIE measure described above (Felten et al., 2021) to account for complementarity: cases where AI does not substitute labor. They focus on O*NET work contexts and the occupation's job zones. Work contexts capture "physical and social factors that influence the nature of work" and is used as a proxy for the likelihood that occupational activities would use AI without human supervision (e.g. tasks like piloting an airplane with grave consequences to errors and may be less likely to see full AI substitution despite suitable ability). Job zones show the amount of education or training required for an occupation, and the authors posit that these higher-skilled workers may have a greater capacity to use AI as a tool that complements their tasks, using it to enhance their productivity or to perform tasks in new ways that AI alone cannot manage.

   Scale: The authors adjust the AIOE for a given occupation i with a measure of potential complementarity of AI, θ:

$$\text{C-AIOE}_i = \text{AIOE}_i \times (1 - (\theta_i - \theta_{MIN}))$$

   Where θ_i is a 0-1 value computed as an arithmetic mean, normalized by 100, of six 0-100 scores. The six components are arithmetic means of similarly-grouped O*NET job contexts and job zones. An O*NET job context is scored from 0 to 100 and a job zone is scored from 1-5, but has it values multiplied by 20 to get a 20-100 scale.

   Data URL: N/A

C.2 Measures of Computer Automation

C.2.1 Occupation-Based Measures

1. **Job susceptibility to computerization (Frey and Osborne, 2017).** Estimating probability of computerization of occupations.

Details: The authors use the 2010 version of O*NET with 903 detailed occupations which are linked through the Standard Occupational Classification (SOC) to employment and wage data in the Bureau of Labor Statistics (BLS). They specifically look at the level (low, medium, high) of 9 occupation-specific variables labelled in the O*NET dataset that represent bottlenecks to automation, lowering the probability of automation. The bottleneck variables were grouped into three major categories: perception and manipulation, creative intelligence, and social intelligence. The probability of an occupation being automated given these O*NET variables was then estimated using a Gaussian process classifier.

Scale: In the labelling step, the researchers and students answered 0/1 to the question: "Can the tasks of this job be sufficiently specified, conditional on the availability of big data, to be performed by state of the art computer-controlled equipment?". 1 was only assigned to occupations where 100% of tasks were automatable.

Then, the Gaussian process classifier estimates a probability of automation in a given occupation given the O*NET bottleneck variables.

Data URL: N/A

2. **Classification of Routine and Non-Routine Tasks within Occupations (Lewandowski et al., 2020).** Classifying O*NET task measures into task contents.

   Details: The authors divide O*NET tasks into four task content buckets: non-routine cognitive analytical, non-routine cognitive interpersonal, routine cognitive, and manual. They specify threshold frequencies for a task to fall under each bucket. This is used to calculate an average "routineness" of an occupation, through a Routine Task Intensity (RTI) score.

   Scale: The RTI score is then assigned to occupation i based on the number of tasks that fall into each bucket:

$$RTI_i = \ln(\text{Routine Cognitive}) - \ln\left(\frac{\text{Non-routine analytical} + \text{Non-routine personal}}{2}\right)$$

   Data URL: https://piotr-lewandowski.pl/code_database/

C.2.2 Task-Based Measures

1. **Worker-Task susceptibility to computerization (Arntz et al., 2016).** Percentage of workers within an occupation that perform tasks that cannot be computerized.

   Details: The authors match the occupation automatibility indicator from Frey and Osborne (2017) to US occupation and task data from the Programme for the International Assessment of Adult Competencies (PIAAC) survey. Each datapoint in

the survey is an individual who is classified into an common occupation code, but may perform different tasks to others in the same occupation. They regress the automatibility of a set of tasks (not an occupation) on the O*NET characteristics of the occupations (such as social intelligence, as discussed above in Frey and Osborne (2017)).

Scale: A percentage of workers in an occupation that perform tasks that cannot be automated (i.e. that have the "bottleneck" characteristics described in Frey and Osborne (2017)). This is different from the probability of automating the entire occupation.

Data URL: N/A

2. **Overlap between patents and job task descriptions (Webb, 2020).** Quantifying how many patents contain verb-noun pairs similar to tasks described in job descriptions.

   Details: A natural language processing (NLP) algorithm extracts verb-noun pairs from O*NET task descriptions for a given occupation. Then, Webb quantifies how many patents in a given technology class (software, robots or AI) have similar verb-noun pairs. For instance, a doctor job description might have the "diagnose condition" pair, which matches with "diagnose disease" in a patent.

   Scale: Each verb-noun pair is given an exposure score which corresponds to how much a verb-noun task represents out of all total verb-noun pairs extracted from patents in a given technology class.

   Scores are also given at the occupation level, by taking a weighed average of task-level scores. Weights are assigned by taking the average O*NET frequency, importance and relevance of a task to the occupation.

   Data URL: N/A

3. **Worker-Task susceptibility to computerization (Arntz et al., 2016).** Percentage of workers within an occupation that perform tasks that cannot be computerized.

   Details: The authors match the occupation automatibility indicator from Frey and Osborne (2017) to US occupation and task data from the Programme for the International Assessment of Adult Competencies (PIAAC) survey. Each datapoint in the survey is an individual who is classified into an common occupation code, but may perform different tasks to others in the same occupation. They regress the automatibility of a set of tasks (not an occupation) on the O*NET characteristics of the occupations (such as social intelligence, as discussed above in Frey and Osborne (2017)).

   Scale: A percentage of workers in an occupation that perform tasks that cannot be automated (i.e. that have the "bottleneck" characteristics described in Frey and

Osborne (2017)). This is different from the probability of automating the entire occupation.

Data URL: N/A

4. **Exogenous shock to capital production efficiency (Restrepo, 2023).** Subset of tasks associated with a production function shift from labor to capital.

   Details: In this model, firms assign workers or capital to specific tasks based on their productivities. The authors model the automation of a task by setting the capital productivity of that task initially at zero and increasing it to a sufficiently high, nonzero level.

   Scale: The model gives us the direct task replacement generated, as well as cost-savings gained.

   Data URL: N/A

C.3 Measures of Robotic Automation

1. **Local labor market exposure to Robots via Growth in Robot Stock (Acemoglu and Restrepo, 2020).** Bartik-style measure of robot exposure in specific geographies.

   Details: The authors interact local commuting zones shares of industry with the technological possibility of introducing robots in these industries, derived from data measuring stock of robots in each industry. They use a Bartik-style measure to show which areas will be most affected by robotic exposure.

   Scale: Exposure score is defined as

$$\text{Exposure to Robots}_i = \sum_j \left(\text{Employment Share}_{i,j} \times \text{Penetration of Robots}_j \right)$$

   Where i is a local labor market, and j corresponds to a given industry. Penetration of robots in a given industry is given by:

$$\text{Adjusted Penetration of Robots } (APR)_j = \frac{dM_j}{L_i} - \frac{dY_j}{Y_j} \cdot \frac{M_j}{L_j}$$

   Where for industry j, dM_j is the change in the use of robots, L_j is the labor employed, dY_j is the change in the output of the industry and Yj is its initial output.

   Data URL: N/A

2. **Local labor market exposure to AI and Robots via Patent Analysis (Gathmann and Grimm, 2022).** Bartik-style measure of robot and AI exposure of industries via patent analysis.

   Details: Natural language processing is used to identify Robotics and AI-related patents, matched to industries via keywords. The Robotics patent measure was validated against the robotic stock dataset used in Acemoglu and Restrepo (2020). The impact of robots/AI was measured using a shift-share Bartik measure. The sum of patents in AI or robotics in a given industry since 1990 was taken as a measure of cumulative knowledge created, the shift variable. The share variable was the employment share of each industry each region.

   Scale: Exposure score is defined as

$$\text{Exposure}_r^C = \sum_i \left(\text{Employment Share}_{i,r} \times \sum_{t=1990}^{2018} \log(1 + \text{Patents Filed}_{i,t}^C) \right)$$

   Where i is the industry, r is a region, and C could either be robots or AI.

   Data URL: N/A

C.4 O*NET Database

C.5 Prompts

C.5.1 Skill-requirements by task

The prompt is as follows, where the text in bold are examples for, in order of occurrence, the occupation title, the task description, the skill name, and the skill level anchors:

*The occupation **Chief Executives**, contains the task: **Prepare budgets for approval, including those for funding or implementation of programs.** What level of skill in **reading comprehension** is needed to perform the task in this occupation well? Provide the answers on a scale from 1 to 7, where 2 means **read step-by-step instructions for completing a form**, 4 means **understand an email from management describing new personnel policies**, and 6 means **read a scientific journal article describing surgical procedures**. Output only a single integer, valued between 1 and 7. Do not output anything else.*[13]

C.5.2 Productivity of AI-related technologies by task

The prompt is as follows, where the text in bold are examples for, in order of occurrence, the specific AI-related technology, a description of that technology, the worker's occupa-

[13] The LLM's answer to this specific prompt was 4.

tion the task description, the skill name, and the skill level anchors:

*We are conducting a rigorous assessment of the time a worker can save on specific tasks by using a **AI-powered robot**.*

1. *Description of technology: **AI-powered robot***

   AI-powered robots combine sophisticated algorithms and sensor technologies to perform tasks with increased autonomy and efficiency. They are highly effective at tasks that require precision, endurance, and the ability to process large amounts of data. In manufacturing, industrial robots automate repetitive tasks like assembly and welding using precise actuators and machine learning algorithms, boosting productivity. Self-driving cars employ sophisticated perception systems—such as LiDAR, radar, and cameras—for autonomous navigation, enabling features like lane keeping and adaptive cruise control, though they still struggle in unpredictable urban environments and adverse weather conditions. Medical robots enhance surgical precision and assist in rehabilitation by providing dexterity beyond human capabilities. Drones execute autonomous flight operations for surveillance, delivery, and data collection in sectors like agriculture and infrastructure inspection. However, these technologies face limitations: they often lack adaptability to unstructured or dynamic environments. Additionally, they depend heavily on structured data and lack general intelligence, restricting them to specific tasks within predefined parameters.

2. *Description of the worker and the task:*

   *Worker's role: **Chief Executives** with an average level of expertise.*

   *Worker's access to tools: Has all the standard tools available to someone in this position. In addition, this worker now gains access to an **AI-powered robot**.*

   *Worker's task: **Prepare budgets for approval, including those for funding or implementation of programs.***

3. *Question:*

   *Estimate the percentage of time that the worker can save by using the described **AI-powered robot** to assist with the task.*

4. *Output Format:*

   Provide your answer as a percentage (numeric value between 0 and 100). Do not output an explanation or any additional information. The answer should be a single number representing the estimate.[14]

[14]The LLM's answer to this specific prompt was 30%.

D Estimation

D.1 Log-linearization of production function

In this paragraph, we derive the log-linear wage regression equation in (13). First, from combining equations (6) and (9) and taking logs, we obtain

$$\log w_j(\boldsymbol{h}) = \log p_j + \sum_{s\in S} \omega_s \log(h_s) + \frac{1}{\rho - 1} \log\left(\sum_{\tau\in\mathcal{T}_j} \theta_{j,\tau} \exp\left((1-\rho)\sum_{s\in S} \eta_s \left(h_s - r_s(\tau)\right)^2\right)\right).$$

Note that we have imposed that $\gamma_\tau = 1 \ \forall\tau$: i.e., we assume no productivity gains from AI have accrued yet in the estimation sample (we estimate parameters based on data up to 2020). Then, we take a first-order Taylor approximation of the mismatch term around the perfectly matched worker. That is, we approximate around $m_s(\tau) \equiv \left(h_s - r_s(\tau)\right)^2 = 0$ for all $s \in S$ and $\tau \in \mathcal{T}_j$:

$$\begin{aligned}\log\left(\sum_{\tau\in\mathcal{T}_j} \theta_{j,\tau} \exp\left((1-\rho)\sum_{s\in S} \eta_s m_s(\tau)\right)\right) &\approx \log\left(\sum_{\tau\in\mathcal{T}_j} \theta_{j,\tau}\right) + \frac{(1-\rho)\sum_{\tau\in\mathcal{T}_j}\sum_{s\in S} \theta_{j,\tau}\eta_s m_s(\tau)}{\sum_{\tau\in\mathcal{T}_j}\theta_{j,\tau}} \\ &= (1-\rho)\sum_{s\in S}\eta_s \sum_{\tau\in\mathcal{T}_j}\theta_{j,\tau} m_s(\tau)\end{aligned}$$

where the second line follows from $\sum_{\tau\in\mathcal{T}_j}\theta_{j,\tau} = 1$. Combining the equations above with the definition of $m_s(\tau)$, yields equation (13):

$$\log w_j(\boldsymbol{h}) = \log p_j + \sum_{s\in S}\omega_s \log(h_s) - \sum_{s\in S}\eta_s \sum_{\tau\in\mathcal{T}_j}\theta_{j,\tau}\left(h_s - r_s(\tau)\right)^2.$$

References

ACEMOGLU, D. AND P. RESTREPO (2020): "Robots and jobs: Evidence from US labor markets," *Journal of political economy*, 128, 2188–2244.

ARNTZ, M., T. GREGORY, AND U. ZIERAHN (2016): "The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis," OECD Social, Employment and Migration Working Papers No. 189.

BALEY, I., A. FIGUEIREDO, AND R. ULBRICHT (2022): "Mismatch cycles," *Journal of Political Economy*, 130, 2943–2984.

BRYNJOLFSSON, E., D. LI, AND L. R. RAYMOND (2023): "Generative AI at Work," Working Paper 31161, National Bureau of Economic Research.

BRYNJOLFSSON, E. AND T. MITCHELL (2017): "What Can Machine Learning Do? Workforce Implications," *Science*, 358, 1530–1534.

BRYNJOLFSSON, E., T. MITCHELL, AND D. ROCK (2018): "What Can Machines Learn, and What Does It Mean for Occupations and the Economy?" *AEA Papers and Proceedings*, 108, 43–47.

BURSTEIN, A., E. MORALES, AND J. VOGEL (2019): "Changes in between-group inequality: computers, occupations, and international trade," *American Economic Journal: Macroeconomics*, 11, 348–400.

CAMPERO, A., M. VACCARO, J. SONG, H. WEN, A. ALMAATOUQ, AND T. MALONE (2022): "A Test for Evaluating Performance in Human-Computer Systems," Working paper.

CAUNEDO, J., D. JAUME, AND E. KELLER (2023): "Occupational exposure to capital-embodied technical change," *American Economic Review*, 113, 1642–1685.

CLARKE, S. AND B. HANRAHAN (2024): "Comparing the Effect of Different Task Types on Effective Use of GitHub Copilot," in *Generative AI in Real-World Workplaces: The Second Microsoft Report on AI and Productivity Research*, ed. by S. Jaffe, N. P. Shah, J. Butler, A. Farach, A. Cambon, B. Hecht, M. Schwarz, and J. Teevan.

CUI, Z. K., M. DEMIRER, S. JAFFE, L. MUSOLFF, S. PENG, AND T. SALZ (2024): "The Effects of Generative AI on High Skilled Work: Evidence from Three Field Experiments with Software Developers," Working paper.

DELL'ACQUA, F., E. MCFOWLAND III, E. R. MOLLICK, H. LIFSHITZ-ASSAF, K. KELLOGG, S. RAJENDRAN, L. KRAYER, F. CANDELON, AND K. R. LAKHANI (2023): "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013, The Wharton School Research Paper.

ELOUNDOU, T., S. MANNING, P. MISHKIN, AND D. ROCK (2024): "GPTs are GPTs: Labor market impact potential of LLMs," *Science*, 384, 1306–1308.

FELTEN, E. W. AND M. RAJ (2018): "A Method to Link Advances in Artificial Intelligence to Occupational Abilities," *AEA*.

FELTEN, E. W., M. RAJ, AND R. SEAMANS (2021): "Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses," *Strategic Management Journal*, 42.

——— (2023): "Occupational Heterogeneity in Exposure to Generative AI," *SSRN Electronic Journal*.

FREY, C. B. AND M. A. OSBORNE (2017): "The future of employment: How susceptible are jobs to computerisation?" *Technological Forecasting & Social Change*, 114, 254–280.

GATHMANN, C. AND F. GRIMM (2022): "The Diffusion of Digital Technologies and its Consequences in the Labor Market," Presented in the VfS Annual Conference 2022 (Basel): Big Data in Economics.

HANDA, K., A. TAMKIN, M. MCCAIN, S. HUANG, E. DURMUS, S. HECK, J. MUELLER, J. HONG, S. RITCHIE, T. BELONAX, K. K. TROY, D. AMODEI, J. KAPLAN, J. CLARK, AND D. GANGULI (2024): "Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations," Working paper.

HUMLUM, A. (2019): "Robot adoption and labor market dynamics," *Princeton University*.

KANAZAWA, K., D. KAWAGUCHI, H. SHIGEOKA, AND Y. WATANABE (2022): "AI, Skill, and Productivity: The Case of Taxi Drivers," Working Paper 30612, National Bureau of Economic Research.

LEE, D. AND K. I. WOLPIN (2006): "Intersectoral labor mobility and the growth of the service sector," *Econometrica*, 74, 1–46.

LEWANDOWSKI, P., A. PARK, AND S. SCHOTTE (2020): "The Global Distribution of Routine and Non-Routine Work," IZA Institute of Labor Economics.

LI, C. P. . A. J. P. . M. M. T. . M. C. . L. (2023): "Labor Market Exposure to AI: Cross-country Differences and Distributional Implications," IMF Working Paper No. 2023/216.

NOY, S. AND W. ZHANG (2023): "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence," *Science*, 381, 187–192.

PENG, S., E. KALLIAMVAKOU, P. CIHON, AND M. DEMIRER (2024): "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot," Working paper.

RESTREPO, P. (2023): "Automation: Theory, Evidence, and Outlook," *Annual Review of Economics*.

VAITHILINGAM, P., T. ZHANG, AND E. GLASSMAN (2022): "Expectation vs. Experience: Evaluating the Usability of Code Generation Tools Powered by Large Language Models," Working paper.

WEBB, M. (2020): "The Impact of Artificial Intelligence on the Labor Market," Stanford University.