AI, Scale, and Skills: A Quantitative Task-Based Theory of Automation*

March 20, 2025

Danial Lashkari
FRB of NY

Wensu Li
MIT

Christina Qiu
Yale

Neil Thompson
MIT

Please click here for the most recent draft.

Abstract

This paper develops a quantitative task-based theory of automation that allows a direct mapping to empirical measures of AI exposure and AI automation costs. The theory accounts for the increasing returns to scale property of AI, where fixed training costs often outweigh variable inference costs. To further leverage the efficiency gains from scale, the model allows for the entry of upstream task-as-service providers, enabling automation through outsourcing. Automation decisions across tasks are shaped not only by standard patterns of comparative advantage but also by patterns of scale advantage. We estimate these patterns relying on AI scaling laws, which establish robust empirical relationships between AI performance and its computational costs. We estimate this relationship in a fine-tuning AI experiment and use information about task complexity and error tolerance from a worker survey data to back out the cost structure of AI across tasks. Our model provides a quantitative, general equilibrium approach to studying the labor market impacts of AI, predicting how automation affects different skill groups across firms and industries.

*We thank Haoran Lyu and Kaizhi Qian for tremendous help with the data and Lucy Hampton for excellent research assistance.

1 Introduction

The explosive surge in the capabilities of Artificial Intelligence (AI) systems has sparked urgent questions about how these technologies will shape labor outcomes for different groups of workers. Central to this question is the analysis of AI's impact at the level of production tasks, a focus shared by both empirical studies (e.g., Frey and Osborne, 2017; Brynjolfsson et al., 2018; Felten et al., 2018) and theoretical frameworks (e.g., Acemoglu and Restrepo, 2018a, 2020). In this task-based framework, whether a particular production task is automated depends primarily on which factor–AI or a specific skill group–offers the lowest marginal cost of production, reflecting its comparative advantage in that task.

Yet this standard approach leaves out one key distinguishing feature of the current era of deep learning automation, which is the role of *scale* in driving AI performance and capabilities (Thompson et al., 2020). Over the past decade, the computational resources dedicated to the fixed costs of training frontier AI foundation models have grown dramatically.[1] Even though costly models such as GPT4 are trained as general-purpose technologies that could serve a wide range of applications, businesses aiming to leverage them still have to incur substantial "last mile" costs (often in the form of fine-tuning) to tailor them to their specific production needs (Fleming et al., 2024). Once the AI system is deployed, however, its marginal cost in production is typically orders of magnitude less than the training or fine-tuning costs–that is, these models display enormous economies of scale. With such large economies of scale, the scale of production for each task may become a more important determinant of its potential for AI automation than its marginal cost.

In this paper, we construct a theory that integrates into the task-based framework this increasing returns to scale property of AI to study the implications of AI automation for different skill groups. In our theory, in addition to the patterns of comparative advantage, automation decisions are further shaped by the patterns of *scale advantage* across tasks. As in the standard model, firms tend to automate a task if the marginal cost of AI is lower relative to that of other skill groups. But they also tend to automate a task if the fixed costs of AI training is lower relative to the scale of production. Furthermore, complementarities in task-level production imply that the two dimensions of comparative and skill

[1] AlexNet, an early success example of deep neural networks in computer vision, had a model with around 60 million parameters and was trained using a dataset with just over a million images (Krizhevsky et al., 2012). A mere few years later, GPT-3, a large language model, had computational needs exceeding that of AlexBet by a few orders of magnitude: it featured 175 billion parameters and was trained on a textual dataset with around 500 billion tokens (Brown et al., 2020).

advantage reinforce one another. This task-level increasing returns additionally calls for the possibility of efficiency benefits to outsourcing: our theory allows for the entry of upstream task-as-service providers in each industry that pay the training costs to then sell the task as a service to the downstream producers.

In addition to incorporating increasing returns, our theory is *quantitative* in the sense that we can directly map it to data on the patterns of task-level productive (comparative and scale) advantage. The model predicts the distribution of the employment of different skill groups and the usage of automation (computing) capital across industries as a function of a number of core primitives: (i) the patterns of productive advantage of AI and workers across tasks, and (ii) the production intensities of different tasks across industries.

The quantitative aspect of our theory allows us to evaluate previous measures of "AI exposure" (e.g., Frey and Osborne, 2017; Felten et al., 2018; Brynjolfsson et al., 2018; Webb, 2019; Eloundou et al., 2024; Handa et al., 2025). These exposure measures proxy for the as-yet-unrealized potential for AI automation, often relying on textual information to establish similarity between potential applications and emerging AI capabilities. Our model allows us to examine how similar such predictions are to those found based on our cost-benefit, general equilibrium framework. Specifically, we formalize the logic of these measures through the firm-level substitutability of each skill groups with machines which, in our framework, corresponds to the covariance between the share of time spent working on tasks and the degree of automation in each task. This reveals important limitations of these previous proxies, which are further exacerbated after accounting for equilibrium reallocations of production across firms and industries.

To fully leverage the quantitative aspects of the model, we use various data sources to construct empirical measures for the core model primitives and, in particular, the patterns of productive advantage of AI. For this purpose, we rely on a body of recent work that studies the scaling behavior of AI models, demonstrating robust and predictable relationships between scale and performance (Hestness et al., 2017; Rosenfeld, 2021; Bahri et al., 2024). These so-called *AI scaling laws* have been observed across domains as varied as, e.g., large language models (Kaplan et al., 2020) and computer vision (Zhai et al., 2022). Focusing on the fine-tuning stage, which is relevant to the AI applications in production processes, we rely on data from an experiment on a classification-based computer vision AI system, varying computational inputs, i.e., the sizes of model and the training data, and task complexity (number of classes) to find their relationship with measures of performance such as error rate.[2] The resulting empirical relationship, along with data from a

[2]A follow-up experiment studying the fine-tuning scaling laws in the case of large language models

survey of workers that records the degrees of error tolerance and task complexity across vision-related tasks (based on the standard O*NET classification), allow us to infer the computational needs for automating each vision task in the US economy. A calibration of our theory based on this approach leads to sharp predictions about the future path of AI automation on the demand for different skill groups and for their relative wages.

In Section 2 of the paper, we present our quantitative model, which features firms each producing a differentiated output in a given industry by performing a set of tasks. To perform each task the firm in turn needs to fulfill a continuum of distinct task varieties. Firm output is a CES aggregate of all task varieties with industry-specific intensities across tasks. Firms choose to use machines (AI) or workers from one of the skill groups to perform each task variety. Workers in each skill group have varying levels of productivity in performing different tasks, drawn from a Frechet distribution with skill-and-task specific means. Machines (AI), on the other hand, may be used to automate tasks but subject to both fixed (e.g., training) and variable (e.g., inference) costs. We show that firm automation decisions can be characterized by a task-level cut-off on the level of the lowest labor cost, above which firms automate. This cost cut-off is given as a CES aggregate of the standard comparative advantage (AI marginal cost relative to average labor cost) and a novel scale advantage (a function of the relative fixed to variable production cost given firm scale). Given automation decisions, we characterize the allocations of workers and machines across tasks, firms, and industries, and show how the firm-level scale-dependence in AI demand shapes the relationship between the firm- and aggregate-level outcomes.

We propose two different theoretically consistent measures for the exposure of each skill group to machines. The first measure is the aggregate elasticity of substitution between each skill group and machines, which accounts for the intensive and extensive (automation) margin of substitution across tasks, as well as the substitution across firms and industries. We fully characterize these different margins in moving from the task to the aggregate level. A large and positive elasticity corresponds to a strong automation exposure while a negative elasticity indicates a labor *augmentation* exposure to AI. Our second measure is a general equilibrium object that measures the real wage response of each skill group to a change in the price of computing, accounting for the aggregate patterns of substitutability across different skill groups and the aggregate scale elasticity of skill demand.

Finally, we present the extended model that allows for the entry of AI-based task-as-service providers in each industry subject to additional entry costs. This outsourcing option allows the downstream firms to procure a task variety at a price above the intrinsic

(LLMs) is currently underway.

marginal cost of AI, due to the markups charged by task-as-service providers, but without the need to incur the in-house fixed training costs. We characterize the equilibrium entry and markups across all tasks and industries. The market power of the upstream task-as-service providers is curbed by the option of the marginal downstream firm, which is typically the largest buyer, to instead switch to in-house AI automation to avoid the need to pay the markup. Despite the upstream market power, entry is typically associated with additional expansion in the range of automated tasks.

In Section 3 of the paper, we discuss our approach to uncovering the productive advantages of AI. Here, we rely on *AI scaling laws* (e.g., Hoffmann et al., 2022) that have been shown to not only predict past and current AI performance, but to also extrapolate to new capabilities (Rosenfeld et al., 2019). In particular, we build on the fine-tuning laws uncovered by Thompson et al. (2024) that map task-level data on the degree of error tolerance and task complexity to the system requirements, and hence costs, needed to automate that task. We combine this scaling law structure with detailed survey data on task requirements and AI experiments to calibrate scaling law parameters. Together, these allow us to infer the patterns of comparative and scale advantage of AI across a broad range of productive tasks from the O*NET dataset.

Section 4 presents the results of a quantitative exercise based on our model to study the potential exposures of various worker groups to AI automation and augmentation. We rely on the estimated patterns of productive advantage for AI and different skill groups to calibrate the productivity primitives of the model. We calibrate the remaining structural elasticities based on prior work, and choose values for the price of computing and the time scale over which fixed costs are paid to match the rates of adoption computer vision AI reported in the Business Trends and Outlook Survey from Bonney et al. (2024). The baseline model successfully replicates the observed adoption rates, with an unweighted adoption rate of 0.59% and a cost-weighted adoption rate of 2.56%. The quantitative exercise allows us to further project the future path of AI automation, predicting a significant increase in adoption rates as computing costs decline. By 2026, the unweighted adoption rate is projected to rise to 8.9%, reaching 16.2% by 2030.

Examining the theoretically founded measure of exposure of four different skill groups (high school, associate degree, college, and graduate) to AI, we find that all four skill groups experience some degree of labor replacement. High school and college graduates show the most and the least degree of complementarity with AI in our baseline. However, as AI automation intensifies, the degree of substitutability between all skill groups and AI increases, leading to a convergence in their levels of replacement exposure. In line with these predictions about the degree of substitutability of different skill groups, the model

also predict a relative fall in the wage inequality between college and high school graduates, and points to a particularly strong convergence in wage between associate degree and college graduates.

Prior Work As already mentioned, our paper contributes to and extends the body of work on the task-based theories of automation (Zeira, 1998; Autor et al., 2003; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018b, 2019) in two main ways.[3] First, we incorporate the possibility of task-level fixed costs that might be needed to specialize a general-purpose automation technology to a narrow production task. These tasks may be incurred at the firm level or at the industry level (through automation outsourcing). Second, we construct a quantitative version of these models that could be directly mapped to data on the allocations of workers and machines across industries, firms, and tasks. We use the latter characteristic of the model to perform a quantitative exercise based on the model.

By transforming the task-based theories into a quantitative model, our paper builds a bridge to the recent empirical literature on *AI exposure*. Following early contributions such as Brynjolfsson and Mitchell (2017), Felten et al. (2018), and Felten et al. (2018), papers in this literature commonly develop proxies for the degree to which AI may be effective at fulfilling different production tasks. Combining such proxies with additional information from datasets such as the O*NET, which describes the task content of different occupations, allows researchers to construct measures for the potential exposure of each occupation to AI (see also, Brynjolfsson et al., 2018; Frank et al., 2019; Felten et al., 2021; Lane and Saint-Martin, 2021; Cazzaniga et al., 2024).[4] While these proxies are informative about the degree of relevance of AI to the task content of occupations, they do not inform us about economic trade-offs involved in the automation of such tasks. In contrast, our model expresses the exposure of tasks to AI in terms of the economic trade-offs faced at the firm level, and further accounts for the implications for the allocation of production at the industry and aggregate levels.[5] Moreover, we project the exposure ultimately in terms of the welfare-relevant outcomes for workers, that is, the demand and economic returns to their skills.

[3]For other recent theoretical applications of this framework to the study of the macroeconomic and labor market impacts of AI, see, e.g., Korinek and Suh (2024), and Acemoglu (2025).

[4]More recent contributions to this literature leverage data from large language models, either directly by asking them to generate measures of exposure (e.g., Eloundou et al., 2024), or indirectly by constructing measures of usage (e.g., Handa et al., 2025).

[5]Svanberg et al. (2024) is another recent attempt to account for these economic tradeoffs with a focus on the different dimensions of engineering costs associated by AI adoption by firms in a partial equilibrium environment, but does not account for the industry and general equilibrium considerations.

In our study, we rely on AI scaling laws to estimate the variations in AI training and inference costs across production tasks. An alternative approach could be to rely on observations on the adoption of AI across firms, as such data is slowly becoming available (e.g., McElheran et al., 2024; Bonney et al., 2024).[6] However, as we are in early stages of adoption, we should expect the applications where AI is used to evolve over time, and thus a snapshot of adoption today is unlikely to predict the path of future adoption. In contrast, AI scaling laws have been repeatedly shown to predict future AI capabilities as, e.g., was illustrated with their with highly accurate predictions of GPT4 functionality made before the model was built (Achiam et al., 2023). Furthermore, as Svanberg et al. (2024) have shown, AI cost estimates based on scaling laws are better predictors of real unemployment shifts than a wide range of other measures of automation. Thus, we choose to rely on scaling laws for our estimates both because of the empirical success of measures built upon them and because of the plentiful evidence from state-of-the-art model builders of their importance–e.g., "when we first started OpenAI, we didn't expect scaling to be as important as it has turned out to be" (Altman, 2023).

2 Theory

In this section, we introduce our quantitative task-based model of automation. In Section 2.1, we set up the baseline environment of the model, in which final-good producing firms across different industries decide whether or not to automate each of the tasks required for their production. Section 2.2 focuses on the firm-level automation problem at the core of our model and defines our concept of scale advantage. Section 2.3 presents the equilibrium conditions that determine the allocation activity across firms and industries, while Section 2.4 presents our measures of aggregate substitution between each skill group and machines. Finally, in Section 2.5 we present the extension of model that allows for the outsourcing mode for AI automation.

2.1 Environment

Firms, Production, and Tasks Production in the economy happens across a number of different industries. A collection of firms $j \in \mathcal{J}_i$ operate in each of industry $i \in \mathcal{I}$. Firm j produces a differentiated output y_j by fulfilling a varieties of different tasks needed for

[6] This data has been used for exploring the regional labor market impacts of AI (Huang, 2023). Another fruitful source of information has been analyzing the textual overlap between AI-related terminology and job vacancy or resume announcements (see, e.g., Acemoglu et al., 2022; Babina et al., 2023; Hampole et al., 2025).

production in its industry. More specifically, for each type τ of tasks needed for production, the firm needs to perform a unit continuum ν of varieties of that task according to a production function

$$y_j = a_j \left(\sum_{\tau \in \mathcal{T}} \xi_{i\tau}^{\frac{1}{\lambda}} \int_0^1 y_{j\tau}(\nu)^{\frac{\lambda-1}{\lambda}} d\nu \right)^{\frac{\lambda}{\lambda-1}}, \tag{1}$$

where $y_{j\tau}(\nu)$ is the level of variety ν of task τ performed at the firm, λ is the elasticity of substitution across tasks, $\xi_{i\tau}$ denotes the importance of task-τ in industry-i production defined such that $\sum_\tau \xi_{i\tau} = 1$, and a_j is a firm-specific Hicks neutral productivity. As we will discuss below, the firm has the choice of using machines or a collection of different types of workers to perform the level $y_{i\tau}(\nu)$ of each variety ν of task τ. Task varieties are distinguished because workers perform them with varying degrees of efficiency relative to machines.

Factors of Production: Workers and Machines A continuum of workers populate the economy, each endowed by one of S types of skills. The total mass of workers in skill group s is given by $\overline{L}_s$. Workers belonging to skill group s each inelastically supply one unit of labor in perfectly competitive, skill-specific labor markets and receive wage w_s (we normalize the wage of the lowest skill group to unity, $w_0 \equiv 1$). Firms hire workers and assign to them a collection of different varieties of tasks to perform. Skill-s workers have identical productivity in performing any given variety ν of task τ. This productivity $z_{s\tau}(\nu)$ is drawn from a Fréchet distribution with cumulative density function

$$G_{s\tau}(z) = \exp\left[-\left(\frac{z}{Z_{s\tau}} \right)^{-\theta} \right], \tag{2}$$

where parameters $Z_{s\tau}$ and θ determine the scale and dispersion of draws, respectively. Parameter $Z_{s\tau}$ accounts for the average productivity of skill-s workers in performing different varieties of task $\tau$and determines the patterns of comparative advantage of different skill groups.

In lieu of hiring workers, firms can instead use "machines" to automate performing each variety of tasks. As discussed in the introduction, our leading example for such automation is AI tools that rely on computing inputs. We associate the machines used for automation in this example with such computing inputs, or more simply "compute." We assume that to automate a task variety, in addition to the variable costs that scale with the level of the task performed, the firm may additionally need to pay a fixed cost. In the AI example, these fixed costs correspond to the likely costs of training and fine-tuning an AI

model that is tailored to performing the specific variety of task needed for production in the industry. While these costs do not necessarily grow in the scale of production, there are still inference costs that grow linearly with the level of production (see Section 3 for a more in-depth discussion). We assume that both variable and fixed costs are paid in units of machines (compute): to automate the production of a level $y_\tau(\nu)$ of variety ν of task τ, we need $1/z^f_{k\tau}$ units for fixed and $y_\tau(\nu)/z_{k\tau}$ for variable costs. Finally, we assume an exogenous supply $\overline{K}$ of machines with a corresponding equilibrium price r, which is fully owned by a unit mass of capitalists.

Demand and the Product Market Workers and capitalists consume final output Y, which is created by combining the outputs of different industries according to a nested CES aggregator

$$Y^{\frac{\rho-1}{\rho}} = \sum_{1\leq i\leq I} Y_i^{\frac{\rho-1}{\rho}}, \qquad Y_i^{\frac{\eta-1}{\eta}} = \sum_{j\in\mathcal{I}_i} y_j^{\frac{\eta-1}{\eta}}, \tag{3}$$

where industrial output Y_i is in turn created as a CES aggregator across the outputs of firms y_j within the industry. Here, η is the elasticity of substitution across firms and ρ is the elasticity of substitution across industries. Accordingly, the demand for firm-level products is given by

$$y_j = p_i^{-\eta} P_i^{\eta-\rho} P^{\rho} Y, \tag{4}$$

where p_j, P_i, and P are the firm-, industry-, and aggregate-level CES price indices, respectively.

The number of firms in each industry is pinned down by free entry. We assume that firms in industry i enter the market subject to fixed entry cost $f_{e,i}$ in units of final goods. Upon entry, they draw their productivity a_j from an industry-specific distribution with cumulative distribution function $G_i(a)$. We assume that firms ignore their impact on the industry-level price index and the product market competition is characterized by monopolistic competition.

2.2 Firm-Level Automation Decisions

We begin by studying the decision problem of a given firm j. The firm is choosing the mode of production, machines or workers with a given type of skill, for each variety ν of task τ, as well as the size $y_{j\tau}(\nu)$ of the task to be performed. This choice in turn determines the marginal and fixed costs, respectively, $c_{j\tau}(\nu)$ and $f_{j\tau}(\nu)$. Let $x_{j\tau}(\nu)$ denote

the size of the task-variety-level input chosen for production, which could be skill-s labor group $\ell_{s,j\tau}(\nu)$ or machines $k_{j\tau}(\nu)$.

2.2.1 Task-Level Automation

We can write the cost-minimization problem of such a firm as the optimal collection of task-variety level inputs $\boldsymbol{y}_j \equiv \left(\left[y_{j\tau}(\nu)\right]_0^1\right)_{\tau\in\mathcal{T}}$ and choices of production mode $\boldsymbol{x}_j \equiv \left(\left[x_{j\tau}(\nu)\right]_0^1\right)_{\tau\in\mathcal{T}}$ that solve

$$C_j\left(\boldsymbol{w}, r; y_j\right) = \min_{\boldsymbol{y}_j, \boldsymbol{x}_j} \sum_{\tau\in\mathcal{T}} \int_0^1 \left(c_{j\tau}(\nu)\, y_{j\tau}(\nu) + f_{j\tau}(\nu)\right) d\nu, \tag{5}$$

subject to Equation (1), where $c_{j\tau}(\nu)$ and $f_{j\tau}(\nu)$ are the marginal and fixed costs corresponding to the firm's choice of the mode of production is given by as follows

$$c_{j\tau}(\nu) = \begin{cases} \frac{w_s}{z_{s\tau}(\nu)}, & \text{if } x_{j\tau}(\nu) = \ell_{s,j\tau}(\nu) \\ \frac{r}{z_{k\tau}}, & \text{if } x_{j\tau}(\nu) = k_{j\tau}(\nu) \end{cases} \quad \text{and} \quad f_{j\tau}(\nu) = \begin{cases} 0, & \text{if } x_{j\tau}(\nu) = \ell_{s,j\tau}(\nu), \\ \frac{r}{z^f_{k\tau}}, & \text{if } x_{j\tau}(\nu) = k_{j\tau}(\nu). \end{cases} \tag{6}$$

Lemma (1) characterizes the solution to the firm's cost minimization problem in Equation (5). A firm's optimal automation decision takes the form of a monotonic policy, where a task-variety is automated only if all associated labor productivity draws $z_{s\tau}(\nu)$ remains below a certain cutoff, as discussed below.

Lemma 1. *Suppose task-variety labor productivity draws $z_{s\tau}(j)$ follow the distribution given by Equation (2). Let $\widetilde{c}_\tau(\nu) \equiv \min_s\{w_s/z_{s\tau}(\nu)\}$ denote the minimum labor costs for the production of the task-variety (ν,τ) and c_j denote the firm-level marginal cost of producing output y_j that minimizes the costs in Equation (5) for firm j. Then, the firm's automation decision for task-variety (ν,τ) follows a cut-off rule $x_{j\tau}(\nu) = k_{j\tau}(\nu)$ if $\widetilde{c}_\tau(\nu) \geq c^*_{j\tau}$ and $x_{j\tau}(\nu) = \ell_{s,j\tau}(\nu)$ if $w_s/z_{s\tau}(\nu) = \widetilde{c}_\tau(\nu) < c^*_{j\tau}$ with the cut-off level of costs $c^*_{j\tau}$ given by*

$$c^*_{j\tau} = \frac{r}{z_{k\tau}}\left(1 + (1-\lambda)\frac{z_{k\tau}/z^f_{k\tau}}{y^k_{j\tau}}\right)^{\frac{1}{1-\lambda}}, \tag{7}$$

where $y^k_{j\tau} \equiv \xi_{i\tau}\left(r/c_j z_{k\tau}\right)^{-\lambda}\left(y_j/a_j\right)$ is the scale of the task-variety that needs to be performed conditional on automation. The share of automated varieties, in which machines used for produc-

tion, within the task type τ is given by

$$\varphi_{j\tau} \equiv \exp\left[-\left(\frac{c^*_{j\tau}}{\widetilde{c}_\tau}\right)^{\theta}\right], \tag{8}$$

where we have defined an index $\widetilde{c}_\tau$ of average labor cost of producing task τ as a function of wages and skill-group productivities:

$$\widetilde{c}_\tau \equiv \left[\sum_s \left(\frac{Z_{s\tau}}{w_s}\right)^{\theta}\right]^{-\frac{1}{\theta}}, \tag{9}$$

Moreover, the marginal cost of producing task τ by the firm firm are given by $c_{j\tau}/a_j$ where

$$c_{j\tau} = \left[\widetilde{c}_\tau^{1-\lambda}\,\gamma\left(-\log\varphi_{j\tau};\, 1+\tfrac{1-\lambda}{\theta}\right) + \left(\frac{r}{z_{k\tau}}\right)^{1-\lambda}\varphi_{j\tau}\right]^{\frac{1}{1-\lambda}}, \tag{10}$$

where $\gamma(\vartheta; x) \equiv \int_0^x u^{\vartheta-1}e^{-u}du$ denotes the lower incomplete gamma function.

Proof. See Appendix A.1 (page A1). □

To better under Lemma 1, let us consider two extreme cases. First, we consider the case without fixed costs ($z^f_{k\tau} \to \infty$) in which case the ratio of the automation cut-off $c^*_{j\tau}$ to the average labor costs for a given task is given by the ratio of the marginal costs when automated to when not automated:

$$\lim_{z^f_{k\tau}\to\infty} \frac{c^*_{j\tau}}{\widetilde{c}_\tau} = \frac{r}{\widetilde{c}_\tau z_{k\tau}}. \tag{11}$$

This is the standard trade-off found in the task-based models of automation (Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018b, 2019), where workers machines comparative advantage in tasks for which machine productivity is higher relative to labor $z_{k\tau}\widetilde{c}_\tau$. As such we let the reciprocal of the expression in Equation (11) define our task-specific measure CA^k_τ of the *comparative advantage* of machines relative to labor. Equation (8) shows that, in the absence of fixed costs, the share of task varieties automated is a monotonically increasing function of this measure of comparative advantage.

Second, we consider the case without marginal costs ($z_{k\tau} \to \infty$) in which case the ratio of the automation cut-off $c^*_{j\tau}$ to the average labor costs for a given as a function of the

automation fixed costs $1/z^f_{k\tau}$ and the task's production scale $\xi_{i\tau}\left(y_j/a_j\right)$ following

$$\lim_{z_{k\tau}\to\infty}\frac{c^*_{j\tau}}{\widetilde{c}_\tau}=\left(\frac{(1-\lambda)\,r}{\widetilde{c}_\tau^{1-\lambda}c_j^\lambda}\frac{1}{\xi_{i\tau}\left(y_j/a_j\right)z^f_{k\tau}}\right)^{\frac{1}{1-\lambda}}. \tag{12}$$

In this case, the share of automated varieties rises in the task's scale of production of the task $\xi_{i\tau}\left(y_j/a_j\right)$, firm-level marginal costs c_j, and the labor costs $\widetilde{c}_\tau$, and the productivity of machines in terms of fixed costs. Crucially, a rise in the scale of firm-level output y_j raises the size of the task $y^k_{j\tau}$ that is needed to be performed, lowering the effect of the fixed costs and raising the automation cut-off. We therefore define our measure of the *scale advantage* of machines $SA^k_{j\tau}$ as the reciprocal of the expression in Equation (12).

In the presence of both marginal and fixed costs, Equation (7) shows that the automation cutoff rises relative to the two benchmarks defined by Equations (11) and (12) following

$$\frac{c^*_{j\tau}}{\widetilde{c}_\tau}=\left(CA_\tau^{\lambda-1}+SA_{j\tau}^{\lambda-1}\right)^{\frac{1}{1-\lambda}}, \tag{13}$$

where comparative CA_τ and scale $SA_{j\tau}$ advantages are defined as the reciprocals of the two expressions in Equations (11) and (12).

Equation (10) shows that the average marginal cost of producing task τ for the firm depends on the marginal cost of producing the tasks using labor $\widetilde{c}_\tau$, the marginal cost of production using machines $r/z_{k\tau}$, and the automation share $\varphi_{j\tau}$. There are two ways in which changes in the automation share of the firm impact the marginal cost of production of a given task. First, the automation share determines the share of automated tasks and thus the weight of the automation marginal costs $r/z_{k\tau}$ in the overall marginal costs of task-level production for the firm. Second, the automation share also determines the average productivity of labor in performing the non-automated tasks due to a selection margin. When the automation share rises, the firm assigns the production of task varieties in which labor is currently the least productive to machines. This change reduces the average labor costs over the remaining varieties, as captured by a multiplicative factor $\gamma\left(1+\frac{1-\lambda}{\theta};-\log\varphi_{j\tau}\right)/\left(1-\varphi_{j\tau}\right)$.

Using Lemma 1, it is now straightforward to determine the allocation of different factors to production across tasks, firms, and industries. Let $\omega_{k,j\tau}$ and $\omega_{\ell,j\tau}$ denote the share

of machines and labor in the *variable* costs of firm j for performing task τ, given by[7]

$$\omega_{K,j\tau} \equiv \frac{r\,k_{j\tau}}{c_{j\tau}y_{j\tau}} = \left(\frac{r/z_{k\tau}}{c_{j\tau}}\right)^{1-\lambda} \varphi_{j\tau}, \qquad \omega_{L,j\tau} \equiv \frac{\sum_s w_s \ell_{s,j\tau}}{c_{j\tau}y_{j\tau}} = \left(\frac{\widetilde{c}_\tau}{c_{j\tau}}\right)^{1-\lambda} \gamma\left(1+\tfrac{1-\lambda}{\theta}; -\log\varphi_{j\tau}\right). \tag{14}$$

Next, we let $\omega^s_{\ell,j\tau}$ denote the share of all labor costs in task τ for firm j that go to workers with skill s, which is given by

$$\omega_{s,Lj\tau} \equiv \frac{w_s \ell_{s,j\tau}}{\sum_{s'} w_{s'}\ell_{s',j\tau}} = \frac{\left(Z_{s\tau}/w_s\right)^\theta}{\sum_{s'}\left(Z_{s'\tau}/w_{s'}\right)^\theta} \equiv \omega_{s,L\tau}, \tag{15}$$

where the last equality follows from the fact that there are no variations across firms j in this share. As Equation (15) shows, this share does not vary across firms, and as such we can simplify the notation by referring to it as $\omega_{s,L\tau}$. Together, Equations (14) and (15) characterize the variable component of demand for labor and machines at the task level. We can add to these definitions the ratio of fixed to variable costs given by $\omega_{f,j\tau} \equiv r\,\varphi_{j\tau}/c_{j\tau}y_{j\tau}$ that constitutes the fixed component of machine demand. Without fixed automation costs, we have that this ratio is zero, $\omega_{f,j\tau} = 0$, while without automation marginal costs we have $\omega_{K,j\tau} = 0$ and $\omega_{L,j\tau} = 1$.

2.3 Equilibrium Allocations across Tasks, Firms, and Industries

Firm-Level Allocations across Tasks We can now aggregate factor demand across tasks within firms. Firm-level total costs C_j in Equation (5) is given by

$$C_j\left(\boldsymbol{w}, r\,; y_j\right) = c_j \frac{y_j}{a_j} + c_j^f, \qquad c_j^f \equiv r\sum_\tau \tfrac{1}{z^f_{k\tau}}\varphi_{j\tau}, \tag{16}$$

where c_j and c_j^f are both functions of $\left(\boldsymbol{w}, r\,; y_j\right)$, where the firm-level marginal cost is given by c_j/a_j, with c_j satisfying

$$c_j^{1-\lambda} = \sum_\tau \xi_{i\tau} c_{j\tau}^{1-\lambda}, \tag{17}$$

[7] For the proofs of Equations (14) and (15), see Appendix A.1 (page A3).

where firm-and-task-level marginal cost $c_{j\tau}$ is given by Equation (10). Now, we can define the share of task τ in firm j's *variable* costs as

$$\psi_{\tau,j} \equiv \tfrac{c_{j\tau} y_{j\tau}}{c_j y_j} \equiv \xi_{i\tau} \left(\frac{c_{j\tau}}{c_j} \right)^{1-\lambda}. \tag{18}$$

Accordingly, we can write the share of skill group s and machines in the variable costs of firm j as

$$\omega_{s,j} \equiv \frac{w_s \ell_{s,j}}{c_j y_j / a_j} = \mathbb{E}_{\tau|j} \left[\omega_{L,j\tau}\, \omega_{s,L,j\tau} \right], \qquad \omega_{K,j} \equiv \frac{r\, k_j}{c_j y_j / a_j} = \mathbb{E}_{\tau|j} \left[\omega_{K,j\tau} \right], \tag{19}$$

where $\mathbb{E}_{\tau|j}[\cdot]$ denotes an average across tasks with respect to the distribution of variable costs $\psi_{\tau,j}$ for firm j. Finally, we let $\omega_{f,j} \equiv a_j c_j^f / c_j y_j$ denote the fixed-to-variable cost ratios at the firm level.

Allocations across Firms Since firms are monopolistic in occupational goods markets, they set prices at a constant markup over marginal costs $p_j = \frac{\eta}{\eta-1} c_j / a_j$, where the marginal cost of the firm j is c_j, given by Equation (17). Aggregating across firms within industry i, the price index can be written as $P_i = \frac{\eta}{\eta-1} J_i^{\frac{1}{1-\eta}} \bar{c}_i$ where the average industry-level marginal cost $\bar{c}_i$ is given by

$$\bar{c}_i^{1-\eta} = \frac{1}{J_i} \sum_{j \in \mathcal{J}_i} \left(\frac{c_j}{a_j} \right)^{1-\eta}. \tag{20}$$

From Equation (4), we can now write the scale of firm i's output and its share in the industry-level variable costs as

$$y_j = J_i^{\frac{\rho-\eta}{1-\eta}} I^{\frac{\rho}{1-\rho}} \left(\frac{c_j}{a_j} \right)^{-\eta} \bar{c}_i^{\eta-\rho}\, \bar{c}^{\rho}\, Y, \qquad \Psi_{j,i} \equiv \frac{c_j y_j / a_j}{J_i \bar{c}_i Y_i} = \frac{1}{J_i} \left(\frac{c_j / a_j}{\bar{c}_i} \right)^{1-\eta}, \tag{21}$$

where $J_i \equiv |\mathcal{J}_i|$ and $I \equiv |\mathcal{I}|$ denote denotes the number of firms in industry i and the number of industries, respectively, and where $\bar{c}$ stands for the average industry-level marginal costs satisfying

$$\bar{c}^{1-\rho} = \sum_i J_i^{\frac{1-\rho}{1-\eta}} \bar{c}_i^{1-\rho}. \tag{22}$$

The aggregate price index satisfies $P = \frac{\eta}{\eta-1}\bar{c}$ and we can also define the share of industry i in the aggregate variable production costs as

$$\Psi_i \equiv \frac{J_i^{\frac{1}{1-\eta}}\bar{c}_i Y_i}{\bar{c}\, Y} = \left(\frac{J_i^{\frac{1}{1-\eta}}\bar{c}_i}{\bar{c}}\right)^{1-\rho}. \tag{23}$$

With the above definitions, we can write the profits of a firm j in industry i as $\pi_j = \frac{1}{\eta-1}\Psi_i\Psi_{j,i}\bar{c}Y - c_j^f$, with c_j^f given by Equation (16). Aggregating across firms in each industry give the following expression for industrial profits

$$\overline{\Pi}_i = \frac{1}{\eta-1}\Psi_i\,\bar{c}\,Y - \overline{C}_i^f, \tag{24}$$

where we have defined industry-i demand for machines to expend on fixed automation costs as $\overline{C}_i^f \equiv \sum_{j\in\mathcal{J}_i} c_j^f$. Free entry ensures that $P\,f_{e,i} = \Pi_i$ and pins down the number of firms Aggregate profits are then given by $\Pi \equiv \sum_i \Pi_i$. We can express the share of skill group s and machines in aggregate variable costs as

$$\overline{\omega}_s \equiv \frac{w_s\overline{L}_s}{\bar{c}\,Y} = \sum_{i,j}\Psi_i\Psi_{j,i}\,\omega_{s,j}, \tag{25}$$

$$\overline{\omega}_K \equiv \frac{r\overline{K} - \overline{C}^f}{\bar{c}\,Y} = \sum_{i,j}\Psi_i\Psi_{j,i}\,\omega_{K,j}, \tag{26}$$

where we have defined the aggregate machine demand for fixed automation costs as $\overline{C}^f \equiv \sum_j \overline{C}_j^f$ with c_j^f given by Equation (16). The set of $S+1$ equations implied by (25) and (26) clear the factor markets and determine the equilibrium wage rates $\boldsymbol{w} \equiv (w_s)$, rental price of machines r, and the scale of output Y.

For completion, we also define the shares of machines and different skill groups in total firm-level costs are given by

$$\Omega_{K,j} = \frac{rk_j}{C_j} = \frac{\omega_{K,j}+\omega_{f,j}}{1+\omega_{f,j}}, \qquad \Omega_{s,j} \equiv \frac{w_s\ell_{s,j}}{C_j} = \frac{\omega_{s,j}}{1+\omega_{f,j}}.$$

Correspondingly, we can define the aggregate shares of factors in total factor payments as $\Omega_K = \frac{\overline{\omega}_K+\overline{\omega}_f}{1+\overline{\omega}_f}$ and $\Omega_s = \frac{\overline{\omega}_s}{1+\overline{\omega}_f}$.

2.4 Machine Exposure, Labor Automation, and Labor Augmentation

In this section, we examine the predictions of our model for the heterogeneity in the implications of rising exposure to machines for different types of workers. In Section 2.4.1, we present our theoretically founded measures of exposures for different skill groups to machines. These measures capture to the degree to which the aggregate demand for or the real wages of each skill group responds to further automation. As such, these measures depend on the aggregate degree of substitutability between each skill group and machines and among different skill groups. Moreover, to the extent that in our model aggregate skill demand depends on scale, to account for the responses of real wages, these measures additionally require us to know the scale elasticities of aggregate skill demand.

2.4.1 Skill-Group Machine Exposure and the Aggregate Elasticities of Substitution

To understand the relevant exposure of different skill groups to automation in our environment, we examine the consequences of a fall in the price of machines (computing) r on the labor market outcomes of different skill groups. Our first measure for the short-run exposure of different skill groups to machines is the aggregate elasticity of substitution between each skill group s and machines $\sigma^{s,K}$ defined as

$$\sigma^{s,K} \equiv \frac{1}{\Omega_K} \frac{\partial \log L_s}{\partial \log r}, \tag{27}$$

where L_s is the aggregate demand for skill group s and where Ω_K denotes the share of machines in the aggregate factor payments. The vector $\boldsymbol{\sigma}^K \equiv \left(\sigma^{s,K}\right)_s$ provides a measure that reflects the degree to which a fall in the price of machines lowers the aggregate demand across different skill groups. According to this measure, we can define the following taxonomy for the aggregate technological relationship between each skill group and machines. Machines *automate* skill group s if the aggregate elasticity between machines and that group $\sigma^{s,K}$ is positive, implying that a fall in the price of machines lowers the demand for the skill group. In contrast, machines *augment* this skill group if the elasticity is negative. We call the skill group s *complementary* to machines *relative to* skill group s' if we have $\sigma^{s',K} < \sigma^{s,K}$, implying that a fall in the price of machines raises the demand for skill group s relative to s'. Therefore, machines can both automate a given skill group and be complementary with it relative to another group. In the long-run, we have to compute the cumulative impact of large changes in the price of computing r on the demand for the skill group s.

The aggregate degree of substitutability between each skill group and machines de-

fined in Equation (27) accounts for the reallocations of production across different levels of production, i.e., tasks, firms, and industries, in measuring how the overall skill-group demand responds to a fall in the price of machines. However, workers ultimately care about the returns to their skill and not the demand for their skills. In general equilibrium, changes in demand for all different skill groups leads to a change in the wage of each skill group. Moreover, the wage response should be compared against the changes in the consumption price index to account for the change in *real* wages. To define such a general equilibrium measure of exposure, we need to additionally define the aggregate elasticity of substitution between each two skill groups s and $s' \neq s$ and the scale elasticity of aggregate demand for skill group s as

$$\overline{\sigma}^{s,s'} \equiv \frac{1}{\Omega_{s'}} \frac{\partial \log L_s}{\partial \log w_{s'}}, \qquad \bar{\varepsilon}_s \equiv \frac{\partial \log L_s}{\partial \log Y}. \tag{28}$$

Then, we can introduce a general equilibrium exposure measure as the real-wage response to a fall in the price of machines as $-d \log \boldsymbol{w} / d \log r$. The following proposition characterizes this response as a function of the elasticities of substitution between skill groups and machines defined in Equation (27), the cross-skill-group elasticities of substitution, and the scale elasticities defined in Equation (28).

Proposition 1. *Consider a static economy where the factors of production are S skill groups and machines. The vector of the real wage response of different skill groups to falling machine prices is given by*

$$-\frac{d \log \boldsymbol{w}^R}{d \log r} = \Omega_K \left(1 + \left(\mathbf{1}\boldsymbol{\Omega}_m' - \boldsymbol{I}_{-o}\boldsymbol{M}^{-1}\right)\boldsymbol{M}^{-1}\boldsymbol{\sigma}^K\right), \tag{29}$$

where we have defined the S-dimensional vectors $\mathbf{1} \equiv (1, \cdots, 1)'$, $\boldsymbol{\Omega}_m \equiv \left(\varepsilon_s, (\Omega_s)_{s \neq o}\right)'$, *and* $\boldsymbol{\sigma}^K \equiv \left(\sigma^{s,K}\right)_{s'}'$, $\boldsymbol{I}_{-o}$ *and the* $S \times S$*-dimensional matrix* $\boldsymbol{M} \equiv \left[M_{s,s'}\right]_{s,s'}$ *by* $M_{s,s} \equiv \Omega_K \sigma^{s,K} + \sum_{s' \neq s} \Omega_{s'} \sigma^{s,s'}$ *for* $s \neq o$, $M_{o,o} = 0$, *and* $M_{s,s'} = -\Omega_{s'} \sigma^{s,s'}$,*and the* $S \times S$*-dimensional matrix* $\boldsymbol{I}_{-o} \equiv (\mathbf{0}, \boldsymbol{I})'$ *with* $\mathbf{0}$ *an* S*-dimensional vector of zeros and* $\boldsymbol{I}$ *an* $(S-1) \times (S-1)$*-dimensional identity matrix.*

Proof. See Appendix A.1 (page A11). □

Proposition 1 applies to general models featuring machines and heterogeneous skill types. The remainder of this section shows how we can calculate the aggregate elasticities of substitution between skill groups and machines $\boldsymbol{\sigma}^K$ and the responses of real wages $-d \log \boldsymbol{w}^R / d \log r$ in the context of our model of task-based automation.

As before, Proposition 1 provides us with a measure of the short-run impact of change in the price of computing on the real wage of each skill group. In the long-run, we have to use the model to predict the response of the real wage to large changes in the price of compute.

2.4.2 Firm-Level Scale and Substitution Elasticities

Let us begin by studying the structure of scale and substitution elasticities at the firm level. If we abstract away from the heterogeneity across industries and firms, these firm-level elasticities coincide with the aggregate elasticities that characterize the exposure of different skill groups in Equations (27) and in Proposition 1.

We characterize the scaling and substitution properties of the firm-level production function in two steps. First, Proposition 2 below characterizes the degrees of scale dependence (nonhomotheticity) in the cost and factor demand functions. Second, Proposition 3 characterizes the degrees of substitutability across different factors.

Proposition 2. *The scale elasticities of total and marginal costs satisfy*

$$\varepsilon_j \equiv \frac{\partial \log C_j}{\partial \log y_j} = \frac{c_j y_j}{C_j a_j} = \frac{1}{1+\omega_{f,j}}, \qquad \varepsilon_j^m \equiv \frac{\partial \log c_j}{\partial \log y_j} = \frac{\theta\,\omega_{f,j}\,\mathbb{E}^f_{\tau|j}\left[\log \varphi_{j\tau}\,\zeta_{j\tau}\right]}{1-\lambda-\lambda\theta\,\omega_{f,j}\,\mathbb{E}^f_{\tau|j}\left[\log \varphi_{j\tau}\,\zeta_{j\tau}\right]}, \quad (30)$$

where $\mathbb{E}^f_{\tau|j}[\cdot]$ denotes the average across tasks with respect to the distribution of fixed costs for firm j, given by $\psi_{\tau,j}\,\omega_{f,j\tau}/\omega_{f,j}$, and where the task-level proxy for the importance of fixed costs in automation $\zeta_{j\tau}$ define a task-and-firm-level proxy for the importance of fixed costs in automation decisions as

$$\zeta_{j\tau} \equiv \frac{(1-\lambda)\,z_{k\tau}/z^f_{k\tau}y^k_{j\tau}}{1+(1-\lambda)\,z_{k\tau}/z^f_{k\tau}y^k_{j\tau}}. \quad (31)$$

where, as in Equation (7), $y^k_{j\tau}$ is the scale of the task-variety that needs to be performed conditional on automation. The scale elasticities of demand for machines and for different skill groups are given by

$$\varepsilon_{K,j} \equiv \frac{\partial \log k_j}{\partial \log y_j} = \frac{\omega_{K,j}-\lambda\varepsilon_j^m\,\omega_{L,j}}{\omega_{K,j}+\omega_{f,j}} - \theta\left(1+\lambda\varepsilon_j^m\right)\frac{\omega_{f,j}}{\omega_{K,j}+\omega_{f,j}}\mathbb{E}^f_{\tau|j}\left[\log\varphi_{j\tau}\right], \quad (32)$$

$$\varepsilon_{s,j} \equiv \frac{\partial \log \ell_{s,j}}{\partial \log y_j} = \left(1+\lambda\varepsilon_j^m\right)\left(1+\theta\,\frac{\omega_{f,j}}{\omega_{L,j}}\,\mathbb{E}^f_{\tau|j}\left[\log\varphi_{j\tau}\,\frac{\omega_{s,L\tau}}{\omega_{s,Lj}}\right]\right), \quad (33)$$

where we have defined the share of skill group s in firm labor costs as $\omega_{s,Lj} \equiv \omega_{s,j}/\omega_{L,j}$.

Proof. See Appendix A.1 (page A5). □

Without fixed automation costs ($z^f_{k\tau} = \infty$, $\zeta_{j\tau} = \omega_{f,j\tau} = 0$), the firm-level production function is constant-returns-to-scale ($\varepsilon_j = 1$, $\varepsilon^m_j = 0$) and scale-invariant ($\varepsilon_{K,j} = \varepsilon_{s,j} = 1$). Without marginal automation costs ($z_{k\tau} = \infty$), the firm-level production function, we have $\zeta_{j\tau} = 1$, $\omega_{f,j} > 0$, and Equation (30) shows the key statistic that determines the scaling properties of the cost function is the fixed-cost-weighted mean of the logarithm of the automation share, $\mathbb{E}^f_{\tau|j}[\log \varphi_{j\tau}] < 0$. The scale elasticity of marginal cost ε^m_j is negative and decreasing in this statistic, suggesting that scaling up production lowers the marginal cost of production and that this scalability property become stronger if the automation share is concentrated in those tasks that constitute a lager share of fixed costs. As shown by Equation (32), this statistic also determines the the scale elasticity of machine demand following $\varepsilon_{K,j} = -\varepsilon^m_j / \omega^f_{j,i}$, which is therefore also positive.[8]

More generally, when both marginal and fixed costs exist, Equation (30) shows that the firm-level production function always exhibits *non-decreasing returns to scale* ($\varepsilon_j \leq 1$) and a non-positive scale elasticity of marginal costs ($\varepsilon^m_j \leq 0$), implying that marginal costs fall in firm scale. In addition, we find that, all else equal, the scale elasticity of marginal cost ε_j falls, and thus the degree of returns-to-scale rises, in the ratio of fixed to variable costs $\omega_{f,j}$. The scale elasticity of marginal costs ε^m_j monotonically rises in the correlation between the log automation share $\log \varphi_{j\tau}$ and fixed costs across tasks. Intuitively, when the firm scales up production, it shifts production toward tasks that are more scale-elastic, i.e., have relatively higher fixed costs. To the extent that these tasks are associated with higher automation shares, this shifts leads to larger gains in terms of lower marginal costs.

Moreover, Equations (32) and (33) show that the scale elasticity of the firm-level demand for machines exceeds unity ($\varepsilon_{K,j} \geq 1$) while it is below unity for all skill groups ($\varepsilon_{s,j} \leq 1$). Intuitively, when the firm scales up production, it shifts its production toward tasks that are more scale elastic, i.e., have relatively higher fixed costs. Since these fixed costs are associated with automation, such tasks rely more intensively on machines raising the machine demand. The opposite is the case with labor demand. More specifically, the demand for those skill groups falls most that are more intensively concentrated in the tasks that constitute a higher share of the firm's total fixed costs and are more automated, as measured by the correlation $\mathbb{E}^f_{\tau|j}\left[\log \varphi_{j\tau} \frac{\omega_{s,L\tau}}{\omega_{s,Lj}}\right]$.

Next, the following proposition characterizes the firm-level the elasticities of substitution between each skill group and machines, as well as the elasticity of substitution between different skill groups.

Proposition 3. *The Allen-Uzawa elasticities of substitution between skill groups and machines*

[8] See Appendix A.1 (page A12) for the proof.

$\sigma_j^{s,K}$ and between different skill groups $\sigma_j^{s,s'}$ are given by

$$\sigma_j^{s,K} \equiv \frac{1}{\Omega_{K,j}} \frac{\partial \log \ell_{s,j}}{\partial \log r} = \frac{\lambda}{\varepsilon_j} \varepsilon_{K,j} + \frac{\lambda}{\varepsilon_j} \left(\varepsilon_{K,j} - \frac{1}{\omega_{K,j} + \omega_{f,j}} \right) \left(\frac{\varepsilon_{s,j}}{1 + \lambda \varepsilon_j^{\bar{m}}} - 1 \right) - \frac{\theta}{\varepsilon_j} \frac{\omega_{K,j}/\omega_{L,j}}{\omega_{K,j} + \omega_{f,j}} \mathbb{E}_{\tau|j} \left[\frac{\omega_{s,L\tau}}{\omega_{s,Lj}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \frac{\log \varphi_{j\tau}}{1 - \zeta_{j\tau}} \right], \tag{34}$$

$$\sigma_j^{s,s'} \equiv \frac{1}{\Omega_{s',j}} \frac{\partial \log \ell_{s,j}}{\partial \log w_{s'}} = \frac{1 - \lambda + \theta}{\varepsilon_j} \mathbb{E}_{\tau|j} \left[\omega_{L,j\tau} \frac{\omega_{s,L\tau}}{\omega_{s,j}} \frac{\omega_{s',L\tau}}{\omega_{s',j}} \right] + \frac{\theta}{\varepsilon_j} \mathbb{E}_{\tau|j} \left[\omega_{K,j\tau} \frac{\log \varphi_{j\tau}}{1 - \zeta_{j\tau}} \frac{\omega_{s,L\tau}}{\omega_{s,j}} \frac{\omega_{s',L\tau}}{\omega_{s',j}} \right] + \frac{\lambda}{\varepsilon_j} \frac{\varepsilon_{s,j}\, \varepsilon_{s',j}}{1 + \lambda\, \varepsilon_j^{\bar{m}}}, \qquad \text{for } s \neq s', \tag{35}$$

where, as in Equation (19), $\mathbb{E}_{\tau|j}[\cdot]$ denotes the average across tasks with respect to the distribution of variable costs $\psi_{\tau,j}$ for firm j.

Proof. See Appendix A.1 (page A8). □

To better understand Proposition 3, let us once again consider the two polar cases without fixed or marginal costs. The following corollary of the proposition shows the underlying drivers of the degree of substitutability of each skill group with machines at the task level. In particular, it shows that the task-level elasticities of substitution between different skill groups and machines coincide and exceed from the intensive-margin elasticity λ due to the presence of the extensive margin of automation.

Corollary 1. *The task-level elasticities of substitution between skill demand and machines are given by $\sigma_{j\tau}^{s,K} \equiv \frac{1}{\Omega_{K,j}} \frac{\partial \log \ell_{s,j\tau}}{\partial \log r} = \frac{1}{\Omega_{K,j}} \frac{\partial \log \ell_{j\tau}}{\partial \log r} = \lambda + \lambda_{j\tau}^{auto}$, where the extensive margin elasticity in the two limiting cases without fixed or marginal costs satisfies*

$$\lambda_{j\tau}^{auto} = \begin{cases} \frac{\theta}{\omega_{K,j}} \widehat{\gamma} \left(-\log \varphi_{j\tau}; 1 + \frac{1-\lambda}{\theta} \right), & z_{k\tau}^f \to \infty, \\ \lambda \left(\frac{\varepsilon_{K,j}}{\varepsilon_j} - 1 \right) + \frac{\theta}{1-\lambda} \left(\frac{1}{\varepsilon_j \omega_{f,j}} - \frac{\lambda \varepsilon_{K,j}}{\varepsilon_j} \right) \widehat{\gamma} \left(-\log \varphi_{j\tau}; 1 + \frac{1-\lambda}{\theta} \right), & z_{k\tau} \to \infty, \end{cases}$$

where $\widehat{\gamma}(x; \vartheta) \equiv d \log \gamma(x; \vartheta) / dx$ denotes the elasticity of the lower incomplete gamma function, which is monotonically decreasing in x.

Proof. See Appendix A.1 (page A13). □

Importantly, Corollary 1 shows that in both these polar cases, *the task-level degree of substitutability between skills and machines rises across tasks in the automation share $\varphi_{j\tau}$*. Labor is more substitutable with machines in those tasks that are more exposed to automation in the sense that the combination of the comparative and scale advantage of machines over labor in those tasks leads to relatively lower automation cut-offs and higher automation shares in these tasks.

Moving to the firm level in the case without fixed costs, Equation (34) can be rewritten as

$$\sigma_j^{s,K} = \lambda + \mathbb{E}_{\tau|j}\left[\lambda_{j\tau}^{auto}\right] + \mathbb{C}_{\tau|j}\left(\frac{\omega_{s,j\tau}}{\omega_{s,j}}, \lambda_{j\tau}^{auto}\right),$$

suggesting the elasticity of substitution is *higher for those skill groups with a larger share of their work is in tasks in which machines have stronger comparative or scale advantages*. This result provides the theoretical foundation behind the ad hoc measures of exposure commonly used in the literature, which construct proxies for the degree of automation across tasks performed by different occupations and skill groups. In other words, a skill group is more exposed to machines, in the sense that it is more substitutable with them, to the extent that its patterns of comparative advantage are correlated with the the combination of the comparative and scale advantages of machines.

2.4.3 Aggregate Elasticities of Substitution

In order to characterize the aggregate elasticities of scale and substitution, we need to account for the reallocations of production across firms and industries in response to changes in scale and factor prices. To this end, we follow Lashkari et al. (2024) and define $\bar{\varepsilon}_i^m$ and $\bar{\varepsilon}^m$ as the scale elasticities of marginal costs for industry i and for the aggregate economy, defined by the following two conditions

$$\bar{\varepsilon}_i^m \equiv \frac{1}{\eta}\left(\mathbb{E}_{j|i}\left[\left(1+\eta\,\varepsilon_j^m\right)^{-1}\right]\right)^{-1} - 1, \tag{36}$$

$$\bar{\varepsilon}^m \equiv \frac{1}{\rho}\left(\mathbb{E}_i\left[(1+\rho\,\bar{\varepsilon}_i^m)^{-1}\right]\right)^{-1} - 1, \tag{37}$$

where $\mathbb{E}_{j|i}[\cdot]$ and $\mathbb{E}_i[\cdot]$ denote the revenue-cost-weighted averages across firms (under distribution $\Psi_{j,i}$) and across industries (under the distribution Ψ_i). We then define the a distribution of pass-throughs $\Psi_{j|i}^p$ across firms within a given industry i and across industries Ψ_i^p as

$$\Psi_{j,i}^p \equiv \Psi_{j,i}\frac{1+\eta\,\bar{\varepsilon}_i^m}{1+\eta\,\varepsilon_j^m}, \qquad \Psi_i^p \equiv \Psi_i\frac{1+\rho\,\bar{\varepsilon}^m}{1+\rho\,\bar{\varepsilon}_i^m}. \tag{38}$$

Using these distributions, the following proposition characterizes the scale and substitution elasticities at the aggregate level.[9]

[9]Proposition 4 generalizes the aggregation results Berlingieri et al. (2024) to the cases involving scale dependence (nonhomotheticity) and the aggregation results of Lashkari et al. (2024) to a multi-factor setting.

Proposition 4. *The scale elasticities of machine and skill demand at the industry and aggregate levels are given by*

$$\bar{\varepsilon}_{K,i} \equiv \frac{1}{\overline{\omega}_{K,i}} \mathbb{E}^{p}_{j|i} \left[\omega_{K,j} \varepsilon_{K,j} \right], \qquad \bar{\varepsilon}_{s,i} \equiv \frac{1}{\overline{\omega}_{s,i}} \mathbb{E}^{p}_{j|i} \left[\omega_{s,j} \varepsilon_{s,j} \right], \tag{39}$$

$$\bar{\varepsilon}_{K} \equiv \frac{1}{\overline{\omega}_{K}} \mathbb{E}^{p}_{i} \left[\overline{\omega}_{K,i} \bar{\varepsilon}_{K,i} \right], \qquad \bar{\varepsilon}_{s} \equiv \frac{1}{\overline{\omega}_{s}} \mathbb{E}^{p}_{j|i} \left[\omega_{s,j} \varepsilon_{s,j} \right], \tag{40}$$

where $\mathbb{E}^{p}_{j|i}[\cdot]$ *and* $\mathbb{E}^{p}_{i}[\cdot]$ *denote the pass-through-weighted averages across firms within industry* i *(under the distribution* $\Psi^{p}_{j,i}$*) and across industries (under distribution* Ψ^{p}_{i}*). Aggregate Elasticities of substitution between skill group* s *and machines and between this skill group and another skill group* $s' \neq s$ *are then given by*

$$\sigma^{s,K} = \mathbb{E}^{c}_{j} \left[\frac{\Omega_{s,j}}{\Omega_{s}} \frac{\Omega_{K,j}}{\Omega_{K}} \sigma^{s,K}_{j} \right] - \mathbb{E}_{i} \left[\frac{\eta \bar{\varepsilon}}{1+\eta \bar{\varepsilon}^{m}_{i}} \mathbb{C}^{p}_{j|i} \left(\frac{\omega_{s,j} \varepsilon_{s,j}}{\Omega_{s}}, \frac{\omega_{K,j} \varepsilon_{K,j}}{\Omega_{K}} \right) \right] - \frac{\rho \bar{\varepsilon}}{1+\rho \bar{\varepsilon}^{m}} \mathbb{C}^{p}_{i} \left(\frac{\overline{\omega}_{s,i} \bar{\varepsilon}_{s,i}}{\Omega_{s}}, \frac{\overline{\omega}_{K,i} \bar{\varepsilon}_{K,i}}{\Omega_{K}} \right), \tag{41}$$

$$\sigma^{s,s'} = \mathbb{E}^{c}_{j} \left[\frac{\Omega_{s,j}}{\Omega_{s}} \frac{\Omega_{s',j}}{\Omega_{s'}} \sigma^{s,s'}_{j} \right] - \mathbb{E}^{p}_{i} \left[\frac{\eta \bar{\varepsilon}}{1+\eta \bar{\varepsilon}^{m}_{i}} \mathbb{C}^{p}_{j|i} \left(\frac{\omega_{s,j} \varepsilon_{s,j}}{\Omega_{s}}, \frac{\omega_{s',j} \varepsilon_{s',j}}{\Omega_{s'}} \right) \right] - \frac{\rho \bar{\varepsilon}}{1+\rho \bar{\varepsilon}^{m}} \mathbb{C}^{p}_{i} \left(\frac{\overline{\omega}_{s,i} \bar{\varepsilon}_{s,i}}{\Omega_{s}}, \frac{\overline{\omega}_{s',i} \bar{\varepsilon}_{s',i}}{\Omega_{s'}} \right), \tag{42}$$

where $\mathbb{E}^{c}_{j}[\cdot]$ *denotes the total-cost-weighted average across all firms under the distribution* $\frac{\bar{\varepsilon}}{\varepsilon_{j}} \Psi_{i} \Psi_{ji}$*,* $\mathbb{C}^{p}_{j|i}(\cdot,\cdot)$ *and* $\mathbb{C}^{p}_{i}(\cdot,\cdot)$*denote the within-industry and across-industry covariances under the pass-through distributions* $\Psi^{p}_{j,i}$ *and* Ψ^{p}_{i}*, respectively, and where* $\bar{\varepsilon} \equiv 1/(1+\overline{\omega}_{f})$ *denotes the aggregate scale elasticity of costs.*

Proof. See Appendix A.1 (page A14). □

Proposition 4, together with the firm-level elasticities characterized by Propositions 2 and 3, allow us to derive all the aggregate elasticities needed to construct our measures of exposure.

2.5 Outsourcing Automation: Task-As-Service Providers

In the environment set up in Section 2.1, each firm pays the fixed cost for deploying the machines in each task-variety ν. However, to the extent that the production of task-level output is tradable across the boundaries of firms, this firm-specific approach to the deployment of machines may be wasteful due to the replication of the fixed costs. Therefore, there is room for intermediaries to enter the market to provide automation for a given task variety as a service in each industry to the downstream firms producing final goods firms . In this section, we extend our model to allow for this possibility.

We assume that a finite and discrete number $n_{i\tau}(\nu)$ of identical task-as-service providers can enter the market in each task-variety market (τ, ν) in industry i subject to a fixed entry cost f_e^m in units of final output. Once in the market, they engage in a Cournot competition to provide task automation as a service to firms within the industry. To produce $Y_{i\tau}^m(\nu)$ units of task input, each firm has to pay total costs $C_{i\tau}(\nu) = (Y_{i\nu}^m / z_{k\tau} + 1/z_{k\tau}^f)r$. Since the task-as-service firms are identical, they charge the downstream firms within the industry an identical price $q_{i\tau}(\nu)$ for each unit of the task input purchased. Thus, the downstream firm faces the cost minimization problem of Equation (5) subject to the following cost structure

$$c_{j\tau}(\nu) = \begin{cases} \frac{w_s}{z_{s\tau}(\nu)}, & \text{if } x_{j\tau}(\nu) = \ell_{s,j\tau}(\nu), \\ q_{i\tau}(\nu), & \text{if } x_{j\tau}(\nu) = y_{j\tau}^m(\nu), \\ \frac{r}{z_{k\tau}}, & \text{if } x_{j\tau}(\nu) = k_{j\tau}(\nu), \end{cases} \quad \text{and} \quad f_{j\tau}(\nu) = \begin{cases} 0, & \text{if } x_{j\tau}(\nu) = \ell_{s,j\tau}(\nu), \\ 0, & \text{if } x_{j\tau}(\nu) = y_{j\tau}^m(\nu), \\ \frac{r}{z_{k\tau}^f}, & \text{if } x_{j\tau}(\nu) = k_{j\tau}(\nu). \end{cases} \tag{43}$$

where the case of $x_{j\tau}(\nu) = y_{j\tau}^m(\nu)$ corresponds to outsourcing the production of task-variety (τ, ν) to an AI platform.

The key observation is to that the downstream firms treat the choice of automation through outsourcing just like the choice of hiring workers form one of the skill groups. As a result, their automation decision follows

$$x_{j\tau}(\nu) = \begin{cases} \ell_{j\tau}(\nu), & \text{if } \widetilde{c}_\tau(\nu) < q_{i\tau}(\nu) \leq c_{j\tau}^*, \\ y_{j\tau}^m(\nu), & \text{if } q_{i\tau}(\nu) < \widetilde{c}_\tau(\nu) \leq c_{j\tau}^*, \\ k_{j\tau}(\nu), & \text{if } c_{j\tau}^* \leq q_{i\tau}(\nu), \widetilde{c}_\tau(\nu), \end{cases} \tag{44}$$

where $\widetilde{c}_\tau(\nu) \equiv \min_{s'}\{w_{s'}/z_{s'\tau}(\nu)\}$ is the lowest labor cost, and where, as before, the self-automation cutoff $c_{j\tau}^*$ is given by Equation (7). Therefore, the task-as-service demand of firm j, $y_{j\tau}^m(\nu)$, is given by

$$y_{j\tau}^m(\nu) = \xi_{i\tau} \frac{Y_i}{I_i} \overline{c}_i^{\eta} q_{i\tau}(\nu)^{-\lambda} c_j^{-(\eta-\lambda)} a_j^{\eta}, \qquad \text{if } q_{i\tau}(\nu) < \widetilde{c}_\tau(\nu) \leq c_{j\tau}^*. \tag{45}$$

This observation implies that, conditional on the entry of upstream task-as-service providers in a given task variety (τ, ν), the number of entrants and the prices offered should be uniform across all varieties of tasks in a given industry-task-type. For a platform to generate any revenue in a given variety ν, its price $q_{i\tau}(\nu)$ has to be below the lowest labor costs $\widetilde{c}_{i\tau}(\nu)$. But conditional on this restriction, the platform demand does not directly depend on the variety ν. Using this observation, the following lemma characterizes the solution

to the task-as-service provider entry and pricing problem.

Lemma 2. *Suppose task-variety labor productivity draws $z_{s\tau}(j)$ follow the distribution given by Equation (2) and let the (self)-automation cutoff $c^*_{j\tau}$ be given by Equation (7). If there is platform entry in an industry i and task group τ, the same number $n_{i\tau}$ of platforms enter and operate in all varieties ν for which the corresponding value of lowest labor costs $\widetilde{c}_\tau(\nu)$ exceeds a industry-and-task-type-level input price $q_{i\tau}$ that satisfies*

$$q_{i\nu} = \frac{\lambda^m_{i\tau}}{\lambda^m_{i\tau} - 1/n_{i\tau}} \frac{r}{z_{k\tau}}, \tag{46}$$

where $\lambda^m_{i\tau}$ is the elasticity of a platform-level demand function $Y^m_{i\tau}(q)$ given by

$$Y^m_{i\tau}(q) \equiv \xi_{i\tau} \tfrac{Y_i}{J_i} \overline{c}_i^{\eta} \, q^{-\lambda} \left(\sum_{j:\, q < c^*_{j\tau}} c_j^{-(\eta-\lambda)} a_j^{\eta} \right). \tag{47}$$

The summation is over the set of firms for which the cutoff exceeds the platform price q. If there is entry, the equilibrium number of operating platforms $n_{i\tau} \in \mathbb{N}$ is the largest value, along with the corresponding price $q_{i\nu}$ from Equations (46) and (47), for which platform-level profits given by

$$\Pi^m_{i\tau} \equiv \tfrac{1}{n_{i\nu}\lambda^m_{i\tau} - 1} \tfrac{1}{n_{i\nu}} Y^m_{i\tau} \left(\tfrac{n_{i\nu}\lambda^m_{i\tau}}{n_{i\nu}\lambda^m_{i\tau} - 1} \tfrac{r}{z_{k\tau}} \right) \tfrac{r}{z_{k\tau}} - \tfrac{r}{z^f_{k\tau}} - f^m_e, \tag{48}$$

is nonnegative. There is no platform entry if all values $n_{i\nu} \in \mathbb{N}$ and corresponding prices $q_{i\nu}$ lead to negative values for the profits in Equation 48.

Proof. See the proof in Appendix A.1 (on page A20). □

To get further intuitions about the equilibrium, let us consider the special case without firm heterogeneity, in which case $a_j \equiv a_i$ for all firms in industry i. In this case, there is a single threshold $c^*_{i\tau}$ for all firms within the industry. Firms all either fall above or below this threshold. Let us assume they fall below this threshold, so there is no automation without platforms. If the upstream firm offers any price below this value, downstream firms will all purchase the task from the platform, and the total task-variety-industry-level demand in Equation (47) simplifies to $Y^m_{i\tau}(q) = \xi_{i\tau} \tfrac{Y_i}{J_i} \overline{c}_i^{\eta} \, c_i^{-(\eta-\lambda)} a_i^{\eta} \, q^{-\lambda}$, where c_i is the uniform marginal cost of all firms within the industry. Equation (46) then implies that the platform demand elasticity and the platform price, conditional on entry, is then given by $\lambda^m_{i\tau} = \lambda$ and $q_{i\nu} = \lambda/(\lambda - 1/n_{i\nu})\, r/z_{k\tau}$. In this case, the solution is the largest $n_{i\nu}$

$$B_{i\tau} \left(\tfrac{r}{z_{k\tau}} \right)^{1-\lambda} \tfrac{1}{n_{i\nu}} \tfrac{1}{n_{i\nu}\lambda - 1} \left(\tfrac{n_{i\nu}\lambda}{n_{i\nu}\lambda - 1} \right)^{-\lambda} \geq \tfrac{r}{z^f_{k\tau}} + f^m_e,$$

where $B_{i\nu} \equiv \xi_{i\tau} \frac{Y_i}{J_i} \bar{c}_i^{\eta} c_i^{-(\eta-\lambda)} a_i^{\eta}$ is an industry-and-task-type-level demand shifter. We can show that the left-hand-side of the above equation is monotonically decreasing in the number of platforms $n_{i\nu}$, implying that a unique equilibrium exists in this case.

Lemma 2 implies that the firm-level automation share at the level of task types depends on whether the equilibrium platform price $q_{i\nu}$ exceeds the the (self)-automation cutoff $c_{j\tau}^*$ be given by Equation (7). This share is now determined by

$$\varphi_{j\tau} = \exp\left[-\left(\frac{1}{\widetilde{c}_\tau} \min\left\{q_{i\tau}, c_{j\tau}^*\right\}\right)^{\theta}\right], \tag{49}$$

and the marginal cost of producing the task at the firm $c_{j\tau}$ is given by

$$c_{j\tau} = \left[\widetilde{c}_\tau^{1-\lambda} \gamma\left(-\log \varphi_{j\tau}; 1 + \tfrac{1-\lambda}{\theta}\right)\left(1 - \varphi_{j\tau}\right) + \widetilde{c}_{K,j\tau}^{1-\lambda} \varphi_{j\tau}\right]^{\frac{1}{1-\lambda}},$$

where the marginal cost under automation for the firm is given by $\widetilde{c}_{K,j\tau} = r/z_{k\tau}$ under self-automation (when $c_{j\tau}^* \leq q_{i\tau}$) and $\widetilde{c}_{K,j\tau} = q_{i\tau}$ under outsourcing automation (when $c_{j\tau}^* \geq q_{i\tau}$). The shares of different skill groups in firm-level variable costs are then given by Equations (14) and (15), while the share of machines in variable costs is given by $\omega_{K,j\tau} = (\widetilde{c}_{K,j\tau}/c_{j\tau})^{1-\lambda}\varphi_{j\tau}$.

Aggregating across tasks, the firm-level marginal cost c_j, the shares of tasks $\psi_{\tau,j}$ in variable costs, and the firm-level shares in variable costs are still given by Equations (17), (18), and (19) while the fixed costs in Equation (16) is now given by $c_j^f = r\sum_\tau \frac{1}{z_{k\tau}^f}\varphi_{j\tau}\mathbb{I}\left\{c_{j\tau}^* \leq q_{i\tau}\right\}$.The remaining aggregation results across firms and then across industries in Equations (20)–(26) follow the same as before.

3 Uncovering the Patterns of Productive Advantage

As discussed in the introduction, we aim to apply the model set up in Section 2 to study the automation due to AI technologies. This requires us to characterize the patterns of comparative and scale advantage for AI, the measures of productivity for marginal $z_{k\tau}$ and fixed cost $z_{k\tau}^f$ across all tasks, as well as the patterns of comparative advantage across workers, the mean productivities $Z_{s\tau}$ for different skill groups.

Since these technologies are still in their early phases of the introduction in production processes, while the immense automation potential of AI is widely recognized, detailed and comprehensible data on how this potential varies across tasks and industries remains unavailable. To solve this problem, we rely on the so-called AI scaling laws, the empirical

regularities widely documented on the relationship between the measures of error in the performance of AI systems and their computational requirements in terms of model size and data (Section 3.1). Combined with additional survey information that we separately collect from workers about the degrees of error tolerance across (Section 3.2), these scaling laws allow us to infer the computation costs of tasks that can be automated by AI systems. To perform this inference, we conduct experiments that allow us to estimate the parameters of the scaling laws in two different settings: computer vision (CV) and large language models (LLMs) (Section 3.3).

Finally, we use data on skill-and-industry-level employment to infer the patterns of comparative advantage based on the predictions of our model (Section 3.4).

3.1 AI Scaling Laws

As mentioned in the introduction, along with the rapid expansion in the scale of AI models, an empirical literature has emerged that studies their scaling behavior, demonstrating robust and predictable relationships between scale and performance (Hestness et al., 2017; Rosenfeld, 2021; Bahri et al., 2024), across domains as varied as, e.g., large language models (Kaplan et al., 2020) and computer vision (Zhai et al., 2022).

In principle, there are two main stages of AI deployment: (i) *model training*, which corresponds to the fixed cost incurred when developing an AI model of a given size m (in terms of the number of model parameters) with a dataset of size d (in terms of the number of observations), and (ii) *inference*, which corresponds to the marginal cost paid when making one instance of inference using the AI model. Costs at the inference stage scale with the model size m, which is chosen at the training stage. Let h be a measure of the performance of the AI system, e.g., the cross-entropy .loss in the out-of-sample test data. The standard AI scaling law takes the form of the following relationship between performance h, model size m, and data size d (Hoffmann et al., 2022):

$$h = \mathcal{H}\left(m, d; \boldsymbol{\alpha}, \boldsymbol{\psi}, \underline{h}\right) \equiv \psi_m \, m^{-\alpha_m} + \psi_d \, d^{-\alpha_d} + \underline{h}, \tag{50}$$

where $\boldsymbol{\alpha} \equiv (\alpha_m, \alpha_d)$ denote the key power-law parameters indicating the respective returns to model and data sizes in terms of loss reduction, where $\boldsymbol{\psi} \equiv (\psi_m, \psi_d)$ denote the model- and data-augmenting efficiency shifters that may vary depending on factors such as the complexity of the task and the algorithmic efficiency of the model architecture,. Finally, $\underline{h}$ is the so-called irreducible loss, a lower bound on the achievable measure of performance.

Typically, the training costs constitute the majority of costs of AI development and

deployment. As such, the state-of-the-art approach to the design of AI systems, the so-called "Chinchilla" scaling laws (Hoffmann et al., 2022), select the combination of model and data sizes that minimize the training costs in terms of the overall scale of computation required, measured commonly in terms of the number of fixed point operations (FLOPs) for achieving a given level of performance loss h:[10]

$$FLOP\left(h;\boldsymbol{\alpha},\boldsymbol{\psi},\underline{h}\right)\equiv\min_{m,d}\ 6\cdot m\cdot d\qquad\text{subject to}\ \ h=\mathcal{H}\left(m,d;\boldsymbol{\alpha},\boldsymbol{\psi},\underline{h}\right), \tag{51}$$

where the scaling function $\mathcal{H}\left(m,d;\boldsymbol{\alpha},\boldsymbol{\psi}\right)$ is given by Equation (50) and where the multiplicative factor 6 accounts for the average number of FLOPs for each training step and for each token of data (Kaplan et al., 2020).

To use this framework for uncovering the comparative and scale advantages of AI, we associate each task τ in a given industry i with a tolerance on performance loss $h_{i\tau}$, the model and data scale parameters $\boldsymbol{\alpha}_{i\tau}$, and the model and data efficiency shifters $\boldsymbol{\psi}_{i\tau}$. The solution to the problem in Equation (51) then gives us the fixed cost in units of compute (FLOP), while the marginal cost is proportional to the corresponding model size $m_{i\tau}$ that achieves the minimum. We provide a detailed discussion of how we construct the task-specific variables $\left(h_{i\tau},\boldsymbol{\psi}_{i\tau},\boldsymbol{\alpha}_{i\tau}\right)$ in Section 3.3 below.

Lemma 3. *Consider the compute-minimization of Equation 51 for task τ in industry i where AI scaling behavior is characterized by the tuple $\left(h_{i\tau},\boldsymbol{\psi}_{i\tau},\boldsymbol{\alpha}_{i\tau}\right)$. Then, the marginal $z_{k,i\tau}$ and fixed cost $z^{f}_{k,i\tau}$ productivity of AI for the task and industry combination are given by*

$$z^{f}_{k,i\tau}=\frac{1}{6}\left(\frac{1}{\psi_{d,i\tau}}\frac{\alpha_{m,i\tau}}{\alpha_{d,i\tau}+\alpha_{m,i\tau}}\right)^{\frac{1}{\alpha_{d,i\tau}}}\left(\frac{1}{\psi_{d,i\tau}}\frac{\alpha_{d,i\tau}}{\alpha_{d,i\tau}+\alpha_{m,i\tau}}\right)^{\frac{1}{\alpha_{m,i\tau}}}\left(h_{i\tau}-\underline{h}\right)^{\left(\frac{1}{\alpha_{m,i\tau}}+\frac{1}{\alpha_{d,i\tau}}\right)}, \tag{52}$$

$$z_{k,i\tau}=\left(\frac{1}{\psi_{m,i\tau}}\frac{\alpha_{d,i\tau}}{\alpha_{d,i\tau}+\alpha_{m,i\tau}}\right)^{\frac{1}{\alpha_{m,i\tau}}}\left(h_{i\tau}-\underline{h}\right)^{\frac{1}{\alpha_{m,i\tau}}}. \tag{53}$$

Proof. See Appendix A.1 (on page A22). □

Equations (52) and (53) show that the productivity of AI rises, both in terms of training (fixed) and inference (marginal), as the required tolerance in performance loss $h_{i\tau}$ rises. All else equal, across task-industries, those that have higher tolerances for error are those in which AI requires lower training and inference costs. Moreover, the elasticity of the productivity is higher in training terms compared to the elasticity in inference terms. The

[10] To be more precise, the Chinchilla scaling laws typically solve the dual problem of minimizing the loss given a constraint on the total number of computation FLOPs available.

size of this elasticity gap, $1/\alpha_{d,i\tau}$ determines the rate at which fixed relative to marginal costs fall as the tolerance for error rises.

In addition, both components of AI productivity rise inversely in the efficiency shifters $\psi_{i\tau}$. The effect is multiplicative in the case of training, meaning that AI productivity rises in the inverse of the product of model- and data-augmenting efficiency shifters while only the former is relevant to AI productivity for inference. Therefore, the data-augmenting efficiency term, $\psi_{d,i\tau}$, determines another driver of the gap between fixed and marginal costs.

Lemma 3 fully characterizes the comparative and scale advantages of AI across task-industries based on information on the AI scaling behavior $(h_{i\tau}, \psi_{i\tau}, \boldsymbol{\alpha}_{i\tau})$. In the next two sections, we discuss how we use information from a survey of workers across task/industries to construct these measures.

3.2 Survey of Task Error Tolerance and Complexity

To construct our measures of AI scaling behavior across tasks, we use data from a detailed worker-level surveys of the degree of tolerance for errors and the complexity of production tasks that are amenable to AI automation.

Survey We aim to use the results for two surveys, one covering computer vision (CV) tasks involving classification and the second one covering tasks related to large language models (LLM). The second survey is currently being conducted/analyzed; thus, in this draft we report the results based on the initial survey focusing on the vision tasks. The survey includes 3,778 respondents from 272 occupations and covers a total of 445 tasks. Survey responses were collected in two rounds: Wave 1 occurred from July 2 to July 16, 2023; Wave 2 from August 22 to November 22, 2023. Relative to Wave 1, Wave 2 respondents were presented with additional questions on task accuracy, task type, and familiarity with listed occupations. 94.42% of respondents report understanding listed tasks mostly or completely when answering survey questions. The survey also collects the respondents' demographic information, such as age, gender, and income bracket, as well as response characteristics, such as duration, interview date, and user language. Respondents were compensated with one dollar per each task for which they provided answers.

We use the data to construct three main measures at the task and occupation levels: the error tolerance $e_{o\tau}$, the task complexity $\chi_{o\tau}$, and the task type $t_{o\tau}$. The error tolerance parameter $e_{o\tau} \in (0,1)$ relates to the maximum proportion of mistakes workers in occu-

pation o may commit when performing a task τ to produce acceptable output for their jobs. Higher accuracy requirements correspond to low error tolerance $e_{o\tau}$. Task complexity corresponds to the number of potential outcomes a task can yield. In general, more complex tasks are more costly to automate. Among vision tasks, task types correspond to classification, segmentation, and generative tasks. Classification tasks categorize entire images into classes, while segmentation tasks categorizes parts, i.e. segments, of images. Generative tasks require the creation of new images based on prompts. We consider classification tasks to be less costly to automate than segmentation tasks, and generative tasks to be the most costly to automate.

To better understand the interpretation of each variable, let us consider two practical examples where a computer vision model is implemented to automate the process of 1) defect detection in a manufacturing plant, and 2) tumor detection in a radiology application. In the first case, the model may categorizes components into a number of multiple classes based on different defect types, while in the latter case, the classification task may only involve the separation of tumorous from healthy tissue (2 classes). In this case the task complexity might be higher in the first application. However, the error tolerance is likely to be substantially lower in the second application.

We use the responses to the question "What's the worst error rate that a worker doing this task could have and still be considered qualified to do it?" as our measure $e_{o\tau}$ of tolerance for errors in a given task τ for a worker in occupation o. Survey respondents were given 24 categorical options as potential answers to this question. We convert this answer to a numerical level by taking the midpoint error rate of the range listed in the answer chosen. We further use additional information in the survey on tolerance for error including typical error rate ("What's a typical error rate for employees currently doing this task?") and random error rate ("What would the error rate on this task be for a worker who had to guess their answer without any information?").

Given that the CV survey focuses on classification tasks, we use the responses to the question "How many possible choices/categories are in the outcome?" as our measure $\chi_{o\tau}$ of task complexity for task τ in occupation o. Survey respondents were given 14 categorical options as potential answers to this question ranging in value from 2 to 1000 or above. Similarly to the task accuracy measure, we convert this value to a numerical level. Additionally, respondents characterized tasks into (i) those with binary conclusions, (ii) those whose conclusions involveing a choice of *only one* item from a set of choices, (iii) those whose conclusion involves more than one items from a set of choices, and (iv) those that require the creation of visual representations. These responses were used to characterize a variable we will refer to as the task type $t_{o\tau}$. Tasks for which re-

spondents answered (i) or (ii) were described as "classification", while (iii) corresponds to "segmentation," and (iv) to "generative." We consider only classification tasks. Further task-level information collected in the survey and analyzed in this paper includes variables on monthly frequency ("How often does a worker have to make a judgment for the visual part of this task?") and time use proportion ("If the task is automated by digital technology for image analysis, what percentage of time can be saved for workers in this position?").

Estimating Task Characteristics We use the survey data to estimate the task characteristics first at the occupation level and then at the industry level. For continuous variables, we eliminate the impact of irrelevant respondent or response characteristics on estimates by projecting responses onto a vector of respondent characteristics and occupation-task fixed effects. We consider the following regression equation for respondent r in occupation o regarding task τ,

$$outcome^{r}_{o\tau} = FE_{o\tau} + \Gamma' \boldsymbol{X}_r + residual^{r}_{o\tau}, \tag{54}$$

where the outcomes are continuous responses associated with the respondent, $FE_{o\tau}$ are the fixed effect, and $\boldsymbol{X}_r$ are respondent characteristics. Respondent characteristics include response duration and confidence level, as well as respondent language, age, gender, experience, and income-group fixed effects. We use the estimated occupation-task fixed effect estimates as our measures for the task characteristics. For the tolerance error measures, we use the logarithm of the error as the outcome in Equation (54) and then transform the fixed effects to find the level of the error tolerance corresponding to each task-occupation. We normalize fixed effects measures to have the same mean as responses. Variables include the minimum task accuracy requirement, task complexity, and task type.

As a last step, we transform the results from the task-occupation level to the task-industry level. In general, the mapping from occupations to industries is not straightforward: industries are composed of many occupations, and workers in the same occupation can be employed across many industries. To address this issue, we construct a crosswalk of occupations to industry using the CPS, where we aggregate to the industry level by weighing occupation-task characteristics by an occupation's share in an industry's wage-bill. More information on this concordance is found in Appendix B.1.

3.3 Estimating Scaling Laws for Fine-Tuning

We next use data from an AI experiment to estimate the scaling law parameters that allow us, when combined with information obtained from the survey data, to infer our measures of comparative and scale advantage using Lemma 3.

In the current draft, we report the results based on an experiment that focuses the computer vision applications–another experiment focusing on the large language models is currently underway. In our experiment, we study the scaling behavior in AI *fine-tuning* stage rather than the initial training of foundation models (Hernandez et al., 2021; Zhang et al., 2024). Fine-tuning is considered the "last mile" in deploying AI systems to specific applications: it corresponds to the process of tailoring a pre-trained AI model to a business task with the aim of specializing the system for performance for the specific and narrow context of the business need. This step bridges the gap between the general-purpose models and the economic needs of the firm.

In our experiment, we fine-tune a Swin Transformer CV model (Liu et al., 2021) under 80 distinct experimental settings. The CV foundation model was pre-trained on 500 randomly selected classes from the ImageNet dataset.[11] These 80 settings capture variations across three fundamental dimensions: task complexity, data size, and model size.

We define task complexity as the number of classes in the classification tasks and consider four cases: 2 classes, 10 classes, 100 classes, and 500 classes. This dimension reflects the increasing difficulty of classification tasks, allowing us to analyze how task complexity influences the production function of computer vision systems. We consider five different data sizes: 13 samples per class, 65 samples per class, 130 samples per class, 650 samples per class, and 1,300 samples per class. We also consider models with four different size configurations: 7.3 thousand, 0.4 million, 28.3 million, and 87.8 million parameters. These combinations of task complexity, data, and model size variations together formed the 80 unique experimental settings in our study. For each of these 80 settings, we recorded model performance in terms of both cross-entropy loss and the error rates for all parameter updating iterations, allowing us to capture the trajectory of performance improvements as training progressed.

We consider the following parameterization of how the scaling-law parameters $(\alpha_{i\tau}, \psi_{i\tau})$ vary with task complexity $\chi_{i\tau}$:

$$\alpha_{f,i\tau} = \beta^{\alpha}_{f,0} + \beta^{\alpha}_{f,1} \log \chi_{i\tau}, \qquad \log \psi_{f,i\tau} = \beta^{\psi}_{f,0} + \beta^{\psi}_{f,1} \log \chi_{i\tau}, \qquad f \in \{m, d\}, \tag{55}$$

[11] The ImageNet database contains 1,000 classes. For this study, we randomly selected 500 classes for pre-training and used the remaining 500 classes in our test cases to vary task complexity, ranging from 2-class to 500-class classification tasks.

and the following relationship between the standard cross-entropy measure of performance loss in the scaling-law function in Equation (50) and the task classification error $e_{i\tau}$ and complexity $\chi_{i\tau}$:

$$h_{i\tau} = \beta_0^h + \beta_1^h \log e_{i\tau} + \beta_2^h \log \chi_{i\tau} + \beta_3^h \log \chi_{i\tau} \log e_{i\tau} + \beta_4^h \chi_{i\tau}^{-1} + \beta_5^h e_{i\tau} \log e_{i\tau}. \tag{56}$$

In each setting s of the 80 settings in our experiment, we observe a collection of the error rate, cross-entropy loss measure, task complexity, model size, and data size $\left(e_{i\tau}^s, h_{i\tau}^s, \chi_{i\tau}^s, m_{i\tau}^s, d_{i\tau}^s\right)$. Together this data allows us to estimate 10 parameters $\left(\beta^\alpha, \beta^\psi, \beta^h\right)$ along with the level of irreducible cross-entropy loss $\underline{h}$.[12]

Appendix B.4 presents our estimated parameters from the AI experiment. Armed with the estimates from the experiment and the survey estimates of task-and-industry-level error tolerance and complexity, we can first state the tolerance in terms of cross-entropy loss using Equation (56) and then apply then use Equation (55) to infer our measures of AI comparative and scale advantage using Lemma 3.

3.4 Inferring the Patterns of Worker Productivity

To infer the patterns of comparative advantage across different skill groups, we consider the environment before the adoption of AI in which computational costs are prohibitively high ($r \to \infty$, $\varphi \approx 0$). In this case, as we show in Appendix B.6.1, the model predicts the following relationship between average productivities $Z_{s\tau}$ and $Z_{s'\tau}$ in task $\tau$for two skill groups s and s', and their respective relative wages and employment in each industry i as

$$\frac{L_{is}}{L_{is'}} = \left(\frac{w_s}{w_{s'}}\right)^{-(1+\theta)} \frac{\sum_\tau \xi_{i\tau} \widetilde{c}_\tau^{1+\theta-\lambda} Z_{s\tau}^\theta}{\sum_{\tau'} \xi_{i\tau'} \widetilde{c}_{\tau'}^{1+\theta-\lambda} Z_{s'\tau}^\theta},$$

where the average labor cost $\widetilde{c}_\tau$ is given, as before, by Equation (9). Appendix B.6.2 discusses how we use this relationship to infer the patterns of worker comparative advantage from data on wages and industrial employment of different labor groups.

For this exercise, we use the following data sources: (i) the Current Population Survey Merged Outgoing Rotation Group Earnings Data (CPS MORG) for the years of 2016 to 2019 and 2023, (ii) the Bureau of Labor Statistics' O*NET Database (version 27.0). We obtain wages w_s and skill allocations over industries L_{si} from the 2023 CPS MORG. Following the procedure of Burstein et al. (2019), we restrict the sample to workers who

[12] As we discuss in Appendix B.3.1, our specification has yet another degree of freedom to account for potential re-use of the same data in the fine-tuning process.

Table 1: List of Calibrated Parameters

Parameter		Value	Source	Moment	Literature	Model
From Literature						
Task EoS	λ	0.5	Humlum (2019)			
Firm EoS	η	4.0				
Industry EoS	ρ	0.8				
Labor productivity	θ	1.3	Burstein, et al. (2019)			
Calibration						
Firm productivity	ϑ	3.2	BDS			
Capital price	$\log r$	53.98	Bonney, et al. (2024)	ComputerVision adoption rate (unweighted)	0.50%	0.59%
Fixed cost shifter	$\log \widehat{z}_k^f$	34.70	Bonney, et al. (2024)	ComputerVision adoption rate (weighted)	2.40%	2.56%

are between 17 and 64 years of age, who report positive hours worked, and are not self-employed. We multiply top-coded weekly and hourly earnings by a factor of 1.5 and drop observations of workers who report earnings below \$1.675 or above \$123.6 an hour. Proxying skill levels, we separate workers into education groups of (i) high school or lower, (ii) some college, (iii) college, and (iv) graduate school.

4 Quantitative Exercise

In this section, we use our estimates of the productive advantage of AI and different skill groups in Section 3 to quantify the theory developed in Section 2 and study the potential exposures of different skill groups to AI automation and augmentation. We present our approach to the calibration in Section 4.1. We then present the results based on the benchmark model with firm-level automation in Section 4.3.

4.1 Model Calibration

In addition to the primitives that characterize the patterns of productive advantage of AI and workers, there are a few remaining parameters that we calibrate as follows.

Structural Elasticities Crucial elasticities of substitution in the model include the elasticity between tasks λ, firms η, and industries ρ. The Fréchet shape parameter θ dictates cross-task-variety dispersion in labor productivities. We calibrate these parameters to

match values common in the prior literature.[13] Table 1 shows these calibrated parameters.

Firm Productivity Distribution We use the 2022 Business Dynamics Statistics (BDS) dataset to infer the the empirical firm distribution in each industry. We assume that productivity draws for a firm j in industry i follow a Pareto distribution with shape parameter ϑ and scale parameter $\underline{a}_i$ in each industry i. Observing that, in the absence of AI adoption, firm employment is proportional within industry to an is-oelastic function of Hicks-neutral productivity $\ell_j \propto a_j^{\eta-1}$, we use the BDS data to uncover the distributional parameters through the regression specification

$$-\log G_{j,i} = -\frac{\vartheta}{\eta-1}\log \ell_j + \vartheta \log \underline{a}_i + \nu_j, \qquad j \in \mathcal{I}_i, \tag{57}$$

where $G_{j,i}$ is the proportion of employees in industry i employed in firms smaller in size than those in size bin j, ℓ_j is the number of employees in firms of size bin j of industry i, and $\vartheta \log \underline{a}_i$ corresponds to an industry fixed effect. This approach yields the value $\vartheta = 3.2$ and a vector of industry scale parameters $(\underline{a}_i)$.

Patterns of Productive Advantage Based on the task-level information on error tolerance and complexity from our survey in Section 3.2, we use estimated relationships between task characteristics and AI development costs in Equations (50) to infer the productive advantage of AI from Lemma 3.[14]

Inferring the Computing Costs With these values in hand, we have two remaining degrees of freedom. First, we need to set the price of computing r in our baseline. Second, we need to calibrate the time scale over which the fixed costs is paid. The first might be clear but let us expand on the second. To map the variations in the fixed automation costs across tasks in our static model to the data, we need to take a stance on how often these fixed costs have to be paid within one year of the data. We account for this degree of freedom as a constant shifter $\widehat{z}_k^f$ across all training productivities $z_{k\tau}^f$ relative to the fixed costs.

[13] Note that, as shown by Lashkari et al. (2024), we have a condition for the existence of equilibrium, given by $\min_j \varepsilon_j^m > -1/\eta$ and $\min_i \bar{\varepsilon}_i^m > -1/\rho$.

[14] Figure C.1 in Appendix C shows the full distribution of estimated AI fine-tuning and inference cost estimates. In the figure, cost estimates are given in units of FLOPS. The cost estimate distributions display a considerable amount of variation and are skewed left. The distribution of fine-tuning costs stochastically dominates the distribution of inference costs.

We calibrate these two parameters (r and $\widehat{z}_k^f$) in our baseline economy to the levels of AI adoption in September 2023 reported in the Business Trends and Outlook Survey from Bonney et al. (2024). During this time period, 0.50% of firms reported using computer vision AI tools. When weighted by firm-level employment, the average adoption rate increases to 2.40%, suggesting that larger firms are more likely to adopt. Intuitively, the first statistic informs us about the relative costs of automation, thereby pinning down r, while the second statistic informs us about the importance of scale and fixed costs to the adoption patterns, thereby pinning down $\widehat{z}_k^f$. The calibrated parameter values are listed in Table 1. The baseline model generates an unweighted adoption rate of 0.59% and a cost-weighted adoption rate of 2.56%.

We further use the survey data to to project the trends in AI automation forward in time and to examine workers' economic outcomes as the cost of compute r declines. To map out AI's automation path over time, we use the data reported by Bonney et al. (2024) on the firms' expectations about future automation rates.[15] The data suggests that the share of firms expecting to adopt computer vision AI systems by March 2024 rises to 2.28% (from the baseline rate of 0.50% in September 2023).[16] We project the unweighted ComputerVision adoption rate to increase from 0.50% to 2.28% in six months. According to our calibration, such a rise in adoption rates requires a decline of around 89.1% in compute costs r on an annual basis. Below, we use this inferred rate of decline in compute costs to predict the labor market outcomes in 2026 and 2030.

4.2 The Inferred Patterns of Productive Advantage

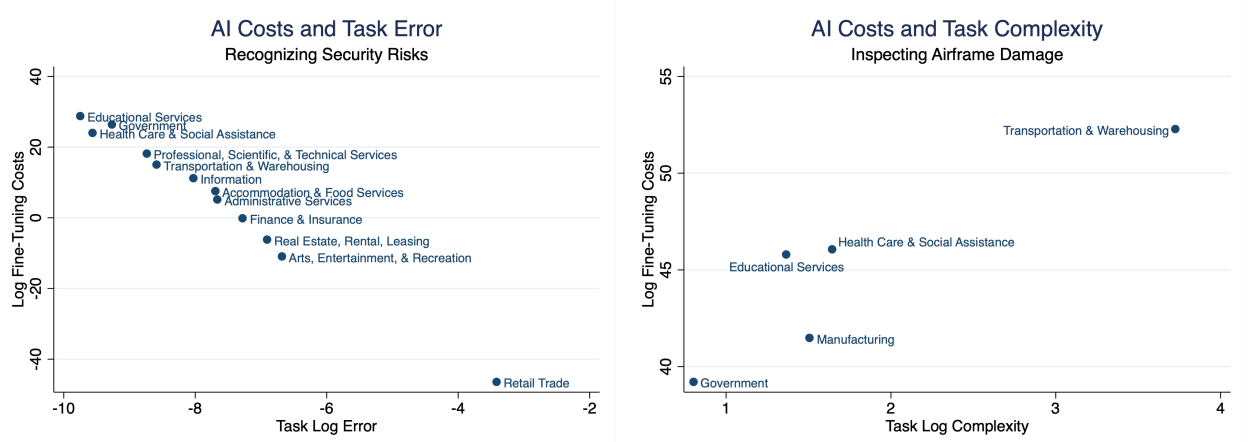
Let us begin by exploring the patterns of productive advantage. From the scaling law of Equation (50), we compute greater AI fine-tuning costs for more complex tasks: within each task complexity group, estimated costs decrease in maximum error tolerance. Figure 1 illustrates the relationship between task error tolerance and estimated AI costs using a number of specific task examples.

In the left panel of Figure 1, we consider the estimated fine-tuning costs associated with the task of "watch[ing] for and recogniz[ing] security risks and thefts and know[ing]

[15] We assign adoption status to a firm if it has a an automated variety share φ in any task above the threshold of 10^{-5}.We find that our calculated adoption rates within the model vary by less than 0.1 percentage point if we adjust the value this threshold from by three orders of magnitude around our chosen threshold.

[16] In September 2023, 67.4% of surveyed firms reported that they had not adopted AI and were not planning on using AI in the next six months. Out of the remaining 32.5% of firms that reported future plans for AI adoption, 7% of those firms planned on adopting computer vision AI tools by March 2024 (Bonney et al., 2024).

Figure 1: AI Costs and Task Characteristics

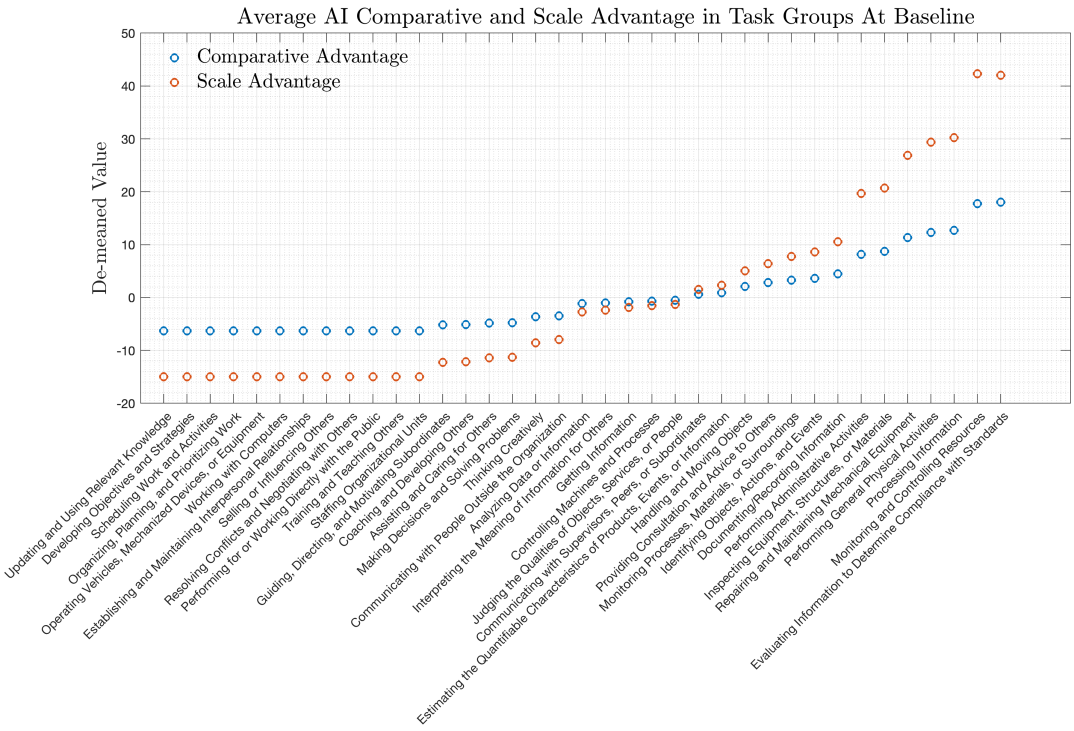


Note: Estimated AI costs increase with more complex tasks and lower maximum error tolerances. We average estimated fine-tuning costs across industries in 2-digit NAICS sectors. In the left figure, we examine estimated costs for the task "watch[ing] for and recogniz[ing] security risks and thefts and know[ing] how to prevent or handle these situations." This task displays variation in maximum error tolerance but not task complexity. We exclude some sectors in the visualization for readability. In the right figure, we examine estimated costs for the task of "inspect[ing] airframes for wear or other defects." This task displays variation in task complexity but not maximum error tolerance. In the visualization, we exclude sectors with no industries whose task complexity exceeds two outcomes for readability.

how to prevent or handle these situations". This commonly-cited task acts as an input in the production of 153 industries and encompasses a large amount of variation in automating costs. However, the task demonstrates very few differences in complexity across industries, with the entire distribution ranging from two to three potential outcomes. The variation in fine-tuning costs are, instead, entirely accounted for by differences in task error tolerance requirements across industries. While this task in the industry "Shoe Retailers" (NAICS 458200) has a maximum error tolerance of 11.5%, its error tolerance in the industry "Elementary and Secondary Schools" (NAICS 611100) is 6.5×10^{-6} and 4.6×10^{-6} in the industry "Nursing Care Facilities" (NAICS 623100). In this figure, the estimated fine-tuning cost of AI declines with the maximum error tolerance. Averaging within more aggregated NAICS 2-digit sectors, estimated fine-tuning costs for recognizing and addressing security risks are highest for Educational Services and lowest for Retail Trade.

In the right panel of Figure 1, we illustrate the relationship between task complexity and estimated AI costs through the task "inspect airframes for wear or other defects." This task is used in 25 industries primarily involving transportation and trade. Across industries, this task displays relatively little variation in maximum error tolerances: the maximum value reported is 1.6%. However, there is substantial variation in reported task complexity. The number of potential outcomes ranges from two outcomes in industry "Electronic and Precision Equipment Repairs and Maintenance" (NAICS 811200) to

Figure 2: AI Comparative Advantage in Task Categories



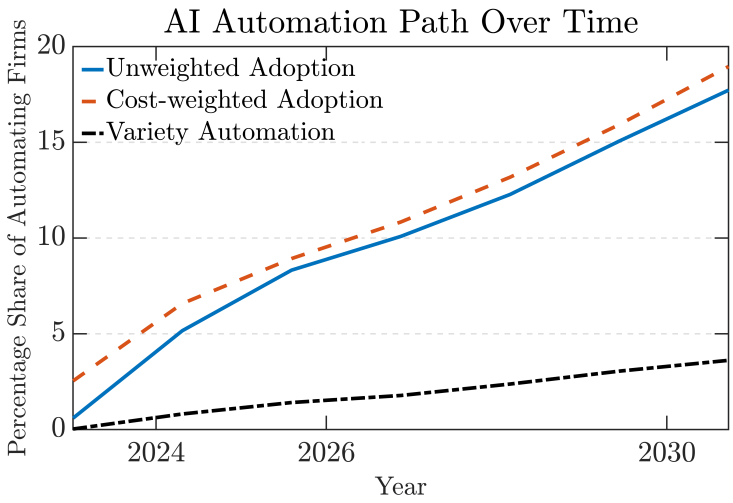
Note: The figure shows the unweighted average of AI's task-level comparative advantage and scale advantage in each category of tasks. Task-level comparative advantage is computed as the cost-weighted average of firm-task-level comparative advantage estimates across firms. All 445 vision tasks in the study are mapped into 198 unique 7-digit DWA categories, 94 unique 6-digit IWA categories, and 26 unique 5-digit task element categories. We include non-vision aggregates of 38 total task element categories. Twelve task element categories do not include a vision task.

130.4 in industry "Support Activities for Air Transportation" (NAICS 488100). In this figure, fine-tuning costs increase with task complexity. Averaging within more aggregated groups, estimated fine-tuning costs are highest for Transportation & Warehousing sectors.

Figure 2 displays the average of AI comparative advantage CA_t and scale advantage SA_t across aggregate task groups at baseline. Here, scale and comparative advantage have the same ranks across tasks categories, though scale advantage magnitudes show more dispersion. This correspondence occurs because the scaling law generates a tight correlation between AI's training and inference costs.

At baseline, AI demonstrates the greatest comparative advantage in tasks involving the evaluation of information to determine compliance with standards, monitoring and controlling resources, and processing information. ComputerVision displays the least comparative advantage in task categories with limited composition of vision tasks, such as establishing and maintaining interpersonal relationships, resolving conflicts, or staffing organizational units. Among categories with some vision tasks, AI displays the

Figure 3: The Automation Path of AI



Note: The figure shows the evolution of aggregate firm adoption rates as the cost of compute declines. We calibrate the cost of compute to match 2023 firm AI adoption in the Business Trends and Outlook Survey as the baseline. We use baseline costs to calculate the weighted estimates and variety share.

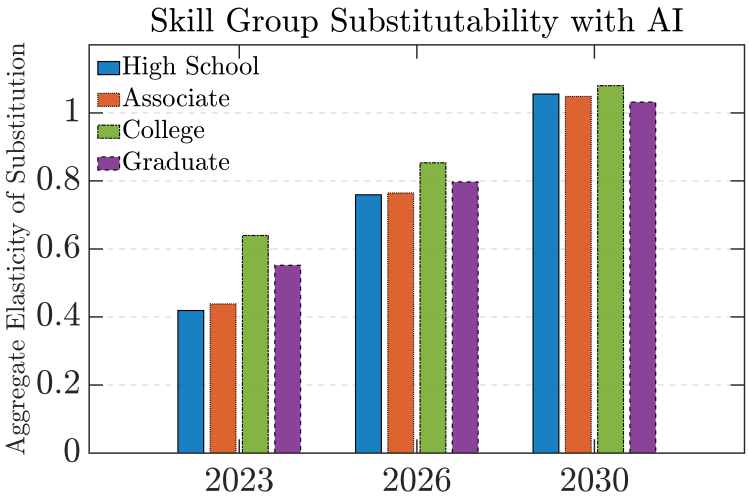
least comparative advantage in tasks that involve coaching others, caring for others, and making decisions.

We note that as the cost of compute r declines,, and as wages adjust in general equilibrium, scale advantage evolves along the automation path. We discuss the evolution of both scale and comparative advantage further in Section 4.3.

4.3 Worker Exposure in the Benchmark Model

As discussed in Section 4.1, we use our calibration to project the AI automation path from 2023 to 2030 as the cost of compute declines. Our projected path, shown in Figure 3, exhibits a dramatic growth in the course of just seven year leading to 2030. In 2026, the unweighted adoption rate rises to 8.9%, further increasing to 16.2% of firms by 2030. The cost-weighted adoption rate remains larger than the unweighted rate, demonstrating that larger firms are more likely to adopt along the automation path. This adoption measure corresponds to the extensive margin of automation at the *firm* level, thereby inflating the intensity with which firms use AI. To explore this margin, Figure 3 also shows the cost-weighted average of the share of automated varieties across firms and tasks, $\mathbb{E}_j[\mathbb{E}_{\tau|j}[\varphi_{ij\tau}]]$, which corresponds to the extensive margin of automation at the *task* level in our model. As vision tasks compose a relatively small amount of all production tasks, the share of

Figure 4: Substitutability of Skill Groups and AI on Automation Path



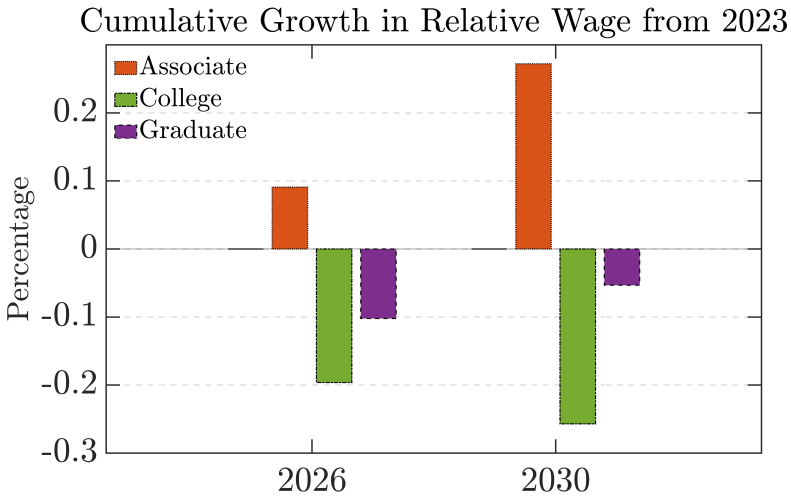
Note: This figure examines the substitutability of skill groups to AI as the price of compute declines, where substitutability is defined as the aggregate elasticity of substitution between capital and labor. Higher values indicate greater substitutability. Elasticities are calculated at the projected price of compute in years 2023, 2026, and 2030.

automated task varieties remains less than 4% throughout the automation path.

Figure 4, leftmost panel, shows the elasticity of substitution between each of the four skill groups and computer vision AI in our baseline calibration (2023), presented in Equation 27. This is our first short-run measure of skill group exposure. First, we note that all measures of positive, suggesting that all skill groups experience some degree of labor replacement, rather than augmentation, based on our model. However, all groups are gross complements with AI in our baseline. In contrast to earlier patterns of automation, the skill group exhibiting the most complementarity with computer vision AI appears to be the high school graduates, followed by associate degree holders. College graduates experience the most substitutability with AI in our baseline.

The figure further explores how this measure of the short-run exposure of skill groups to AI evolves over time. Our first observation is that, as AI automation intensifies, the degree of substitutability between each skill group and computer vision AI rises. By 2030, all skill groups are gross substitutes with AI. A second observation is that the heterogeneity among skill groups in their short-run exposure to AI, e.g., the degree of relative complementarity between low-skilled workers an AI, also falls over time. In a sense, labor appears to become more homogenous in terms of its patterns of substitutability with AI.

Figure 5: Substitutability of Skill Groups and AI on Automation Path

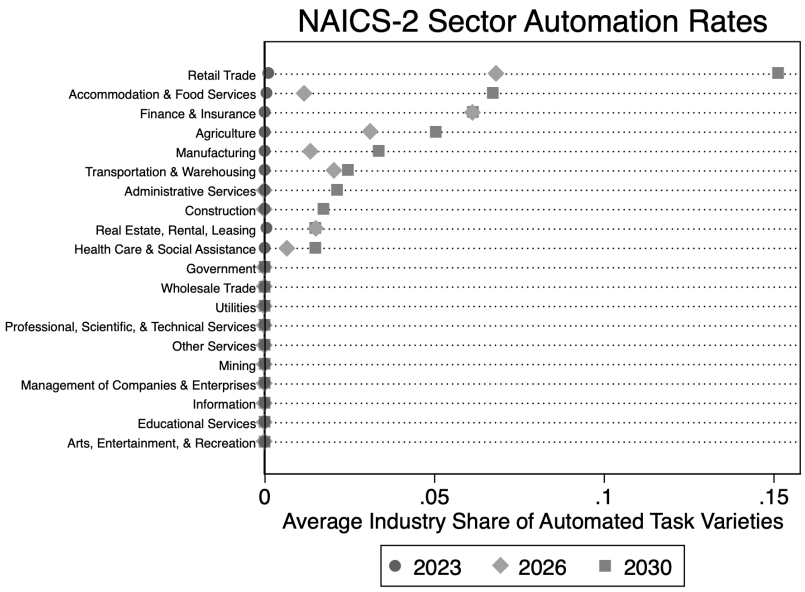


Note: The figure illustrates nominal wage growth relative to the baseline year of 2023. Nominal wage growth for a skill group s at time t relative to baseline year 0 is $(\frac{w_{st}}{w_{s0}} - 1)$. Wage growth is calculated at the projected price of compute in years 2026 and 2030.

Next, we explore the medium-run measures of exposure between worker groups and AI. Figure 5 shows the changes implied by the falling computing prices in the wages of different skill groups relative to that of the high school graduates. The patterns of substitutability with AI between skill groups are reflected in relative wage growth over time. In 2026, the wages of both college and graduate school workers decline relative to high school workers, while the wages of associate degree workers increase. The relative wage of college workers declines further in 2030, as other skill groups remain more complementary with AI than college workers. However, the relative wage of graduate school workers increases between 2026 and 2030 as the relative complementary of low-skilled groups with AI becomes less pronounced. The evolution of real wages are shown in Figure C.2 of the Appendix. These account for the general equilibrium effects on aggregate scale as well as the responses in the waves of other skill groups. Growth in real wages as a result of falling computing prices and AI automation is substantial and positive during this period.

Finally, we consider heterogeneity in automated variety share by industry in Figure 6. The top three industries with the greatest projected automated variety rates by 2030 are Retail Trade, Accommodation & Food Services, and Finance & Insurance with estimated variety automation shares of 15.1%, 6.7%, and 6.1%, respectively. In the Business Trends and Outlook Survey from Bonney et al. (2024), these three sectors are the only ones aside

Figure 6: Vulnerable Sectors to AI Automation



Note: The figure displays average cost-weighted automated variety shares in 6-digit NAICS industries by 2-digit NAICS sectors. The automated variety share are calculated at the projected price of compute in years 2023, 2026, and 2030.

from sector "Information" (NAICS 51) to report employment-weighted firm shares of expected AI use exceeding 15.0%. Our model points to the same sectors as being vulnerable to automation, despite these sectors having limited AI takeup at baseline.

5 Conclusion

This paper develops a quantitative task-based theory of automation that integrates the increasing returns to scale property of AI, offering new insights into how automation decisions are shaped by both comparative and scale advantages across tasks. By formalizing the economic trade-offs involved in AI adoption, our model bridges the gap between task-based theories of automation and the empirical literature on AI exposure. We show that automation decisions are not only driven by the relative marginal costs of labor and machines but also by the scale of production and the fixed costs of AI training. Our framework allows for the entry of upstream task-as-service providers, highlighting the role of outsourcing in expanding the range of automated tasks. The model's quantitative nature enables us to map it directly to data, providing a robust foundation for understanding the labor market impacts of AI across different skill groups.

Our findings reveal that AI automation has heterogeneous effects on workers, with all skill groups experiencing some degree of labor replacement, but the extent of exposure varies over time. Initially, low-skilled workers exhibit the most complementarity with AI, but as automation intensifies, the substitutability between all skill groups and AI increases, leading to a convergence in their exposure levels. Despite these challenges, the decline in computing prices drives substantial growth in real wages across all skill groups, as firms reduce marginal production costs through automation. These results underscore the importance of considering both firm-level and aggregate-level reallocations of production when assessing the impacts of AI on labor markets. Future research could extend this framework to explore the implications of AI for occupational mobility and the broader macroeconomic effects of automation.

References

Acemoglu, Daron, "The simple macroeconomics of AI," *Economic Policy*, 2025, *40* (121), 13–58. 6

_ **and David Autor**, "Skills, tasks and technologies: Implications for employment and earnings," in "Handbook of labor economics," Vol. 4, Elsevier, 2011, pp. 1043–1171. 6, 11

_ **and Pascual Restrepo**, "Artificial intelligence, automation, and work," in "The economics of artificial intelligence: An agenda," University of Chicago Press, 2018, pp. 197–236. 2

_ **and** _ , "The race between man and machine: Implications of technology for growth, factor shares, and employment," *American Economic Review*, 2018, *108* (6), 1488–1542. 6, 11

_ **and** _ , "Automation and new tasks: How technology displaces and reinstates labor," *Journal of Economic Perspectives*, 2019, *33* (2), 3–30. 6, 11

_ **and** _ , "The wrong kind of AI? Artificial intelligence and the future of labour demand," *Cambridge Journal of Regions, Economy and Society*, 2020, *13* (1), 25–35. 2

_ **, David Autor, Jonathon Hazell, and Pascual Restrepo**, "Artificial intelligence and jobs: Evidence from online vacancies," *Journal of Labor Economics*, 2022, *40* (S1), S293–S340. 7

Achiam, Josh, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat et al., "GPT-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023. 7

Altman, Sam, "Planning for AGI and beyond," 2023. 7

Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly journal of economics*, 2003, *118* (4), 1279–1333. 6

Babina, Tania, Anastassia Fedyk, Alex X He, and James Hodson, *Firm investments in artificial intelligence technologies and changes in workforce composition*, Vol. 31325, National Bureau of Economic Research, 2023. 7

Bahri, Yasaman, Ethan Dyer, Jared Kaplan, Jaehoon Lee, and Utkarsh Sharma, "Explaining neural scaling laws," *Proceedings of the National Academy of Sciences*, 2024, *121* (27), e2311878121. 3, 26

Berlingieri, Giuseppe, Filippo Boeri, Danial Lashkari, and Jonathan Vogel, "Capital-skill complementarity in firms and in the aggregate economy," Technical Report, National Bureau of Economic Research 2024. 21

Bonney, Kathryn, Cory Breaux, Cathy Buffington, Emin Dinlersoz, Lucia S Foster, Nathan Goldschlag, John C Haltiwanger, Zachary Kroff, and Keith Savage, "Tracking firm use of AI in real time: A snapshot from the Business Trends and Outlook Survey," Technical Report, National Bureau of Economic Research 2024. 5, 7, 35, 40

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell et al., "Language models are few-shot learners," *Advances in neural information processing systems*, 2020, *33*, 1877–1901. 2

Brynjolfsson, Erik and Tom Mitchell, "What can machine learning do? Workforce implications," *Science*, 2017, *358* (6370), 1530–1534. 6

_ , _ , and Daniel Rock, "What can machines learn and what does it mean for occupations and the economy?," in "AEA papers and proceedings," Vol. 108 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2018, pp. 43–47. 2, 3, 6

Burstein, Ariel, Eduardo Morales, and Jonathan Vogel, "Changes in Between-Group Inequality: Computers, Occupations, and International Trade," *American Economic Journal: Macroeconomics*, April 2019, *11* (2), 348–400. 32

Cazzaniga, Mauro, Ms Florence Jaumotte, Longji Li, Mr Giovanni Melina, Augustus J Panton, Carlo Pizzinelli, Emma J Rockall, and Ms Marina Mendes Tavares, *Gen-AI: Artificial intelligence and the future of work*, International Monetary Fund, 2024. 6

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs are GPTs: Labor market impact potential of LLMs," *Science*, 2024, *384* (6702), 1306–1308. 3, 6

Felten, Edward, Manav Raj, and Robert Seamans, "Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses," *Strategic Management Journal*, 2021, *42* (12), 2195–2217. 6

Felten, Edward W, Manav Raj, and Robert Seamans, "A method to link advances in artificial intelligence to occupational abilities," in "AEA Papers and proceedings," Vol. 108 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2018, pp. 54–57. 2, 3, 6

Fleming, Martin, Neil Thompson, and Wensu Li, "The Last Mile Problem in AI," Report, Brookings Institution, United States of America 2024. Retrieved from https://coilink.org/20.500.12592/vg1bd71 on 21 Mar 2025. COI: 20.500.12592/vg1bd71. 2

Frank, Morgan R, David Autor, James E Bessen, Erik Brynjolfsson, Manuel Cebrian, David J Deming, Maryann Feldman, Matthew Groh, José Lobo, Esteban Moro et al., "Toward understanding the impact of artificial intelligence on labor," *Proceedings of the National Academy of Sciences*, 2019, *116* (14), 6531–6539. 6

Frey, Carl Benedikt and Michael A Osborne, "The future of employment: How susceptible are jobs to computerisation?," *Technological forecasting and social change*, 2017, *114*, 254–280. 2, 3

Hampole, Menaka, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller, "Artificial intelligence and the labor market," Technical Report, National Bureau of Economic Research 2025. 7

Handa, Kunal, Alex Tamkin, Miles McCain, Saffron Huang, Esin Durmus, Sarah Heck, Jared Mueller, Jerry Hong, Stuart Ritchie, Tim Belonax et al., "Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations," *arXiv preprint arXiv:2503.04761*, 2025. 3, 6

Hernandez, Danny, Jared Kaplan, Tom Henighan, and Sam McCandlish, "Scaling laws for transfer," *arXiv preprint arXiv:2102.01293*, 2021. 31

Hestness, Joel, Sharan Narang, Newsha Ardalani, Gregory Diamos, Heewoo Jun, Hassan Kianinejad, Md Mostofa Ali Patwary, Yang Yang, and Yanqi Zhou, "Deep learning scaling is predictable, empirically," *arXiv preprint arXiv:1712.00409*, 2017. 3, 26

Hoffmann, Jordan, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark et al., "Training compute-optimal large language models," *arXiv preprint arXiv:2203.15556*, 2022. 5, 26, 27

Huang, Yueling, "The Labor Market Impact of Artificial Intelligence: Local vs. Aggregate Effect," 2023. Working Paper. 7

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei, "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020. 3, 26, 27

Korinek, Anton and Donghyun Suh, "Scenarios for the Transition to AGI," Technical Report, National Bureau of Economic Research 2024. 6

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, 2012, *25*. 2

Lane, Marguerita and Anne Saint-Martin, "The impact of Artificial Intelligence on the labour market: What do we know so far?," *OECD Social, Employment, and Migration Working Papers*, 2021, (256), 0_1–60. 6

Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard, "Information technology and returns to scale," *American Economic Review*, 2024, *114* (6), 1769–1815. 21, 34, A14

Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in "Proceedings of the IEEE/CVF international conference on computer vision" 2021, pp. 10012–10022. 31

McElheran, Kristina, J Frank Li, Erik Brynjolfsson, Zachary Kroff, Emin Dinlersoz, Lucia Foster, and Nikolas Zolas, "AI adoption in America: Who, what, and where," *Journal of Economics & Management Strategy*, 2024, *33* (2), 375–415. 7

Muennighoff, Niklas, Alexander Rush, Boaz Barak, Teven Le Scao, Nouamane Tazi, Aleksandra Piktus, Sampo Pyysalo, Thomas Wolf, and Colin A Raffel, "Scaling data-constrained language models," *Advances in Neural Information Processing Systems*, 2023, *36*, 50358–50376. A25

Rosenfeld, Jonathan S, "Scaling laws for deep learning," *arXiv preprint arXiv:2108.07686*, 2021. 3, 26

Rosenfeld, Jonathan S., Amir Rosenfeld, Yonatan Belinkov, and Nir Shavit, "A Constructive Prediction of the Generalization Error Across Scales," 2019. 5

Svanberg, Maja, Wensu Li, Martin Fleming, Brian Goehring, and Neil Thompson, "Beyond AI exposure: which tasks are cost-effective to automate with computer vision?," *Available at SSRN 4700751*, 2024. 6, 7

Thompson, Neil C, Kristjan Greenewald, Keeheon Lee, Gabriel F Manso et al., "The computational limits of deep learning," *arXiv preprint arXiv:2007.05558*, 2020, *10*. 2

Thompson, Neil, Martin Fleming, Benny J Tang, Anna M Pastwa, Nicholas Borge, Brian C Goehring, and Subhro Das, "A model for estimating the economic costs of computer vision systems that use deep learning," in "Proceedings of the AAAI Conference on Artificial Intelligence," Vol. 38 2024, pp. 23012–23018. 5

Webb, Michael, "The impact of artificial intelligence on the labor market," *Available at SSRN 3482150*, 2019. 3

Zeira, Joseph, "Workers, machines, and economic growth," *The Quarterly Journal of Economics*, 1998, *113* (4), 1091–1117. 6

Zhai, Xiaohua, Alexander Kolesnikov, Neil Houlsby, and Lucas Beyer, "Scaling vision transformers," in "Proceedings of the IEEE/CVF conference on computer vision and pattern recognition" 2022, pp. 12104–12113. 3, 26

Zhang, Biao, Zhongtao Liu, Colin Cherry, and Orhan Firat, "When scaling meets LLM fine-tuning: The effect of data, model and fine-tuning method," *arXiv preprint arXiv:2402.17193*, 2024. 31

Appendix to "AI, Scale, and Skills: A Quantitative Task-Based Theory of Automation"

Danial Lashkari, FRB NY
Wensu Li, MIT
Christina Qiu, Yale
Neil Thompson, MIT

March 2025

Contents

A Theoretical Appendix **A1**
A.1 Proofs and Derivations A1

B Appendix for Uncovering the Patterns of Productive Advantage **A23**
B.1 Task Survey A23
B.2 Concordance of Occupations and Industries A24
B.3 AI Scaling Law Experiments A24
B.3.1 Effective Data Size A25
B.4 Scaling Laws Estimation Results A25
B.5 Mapping Error Rates to Cross-entropy Loss A27
B.6 Inferring Task-Level Labor Productivities A27
B.6.1 Patterns of Skill Group Comparative Advantage and Industrial Employment A27
B.6.2 Estimation of the Patterns of Comparative Advantage of Skill Groups A29

C Additional Tables and Figures **A30**

A Theoretical Appendix

A.1 Proofs and Derivations

Proof for Lemma 1. We consider the following two step solution for the firm problem and prove later that this solution is optimal. The firm first considers using labor for producing each task. Without fixed cost, the problem reduces to the choice of the skill group that can produce each input variety ν in task type τ at the cheapest prices $c_{s\tau}(\nu) = w_s / z_{s\tau}(\nu)$. Given our assumption about the task-level distribution of worker productivities, the distribution of the marginal cost $c_{s\tau}(\nu)$ of producing the task-variety (ν, τ) by each skill group s is distributed according to $\mathbb{P}\left(c_{s\tau}(\nu) = \frac{w_s}{z_{j\tau}(\nu)} \geq c\right) = \exp\left[-\left(\frac{Z_{s\tau}}{w_s}\right)^{\theta} c^{\theta}\right]$. Then, we find that the distribution for the minimum cost of producing the task-variety using labor $\widetilde{c}_{j\tau}(\nu) \equiv \min_s c_{s\tau}(\nu)$ is given by

$$\mathbb{P}\left(\widetilde{c}_{j\tau}(\nu) \geq c\right) = \exp\left[-\left(\sum_s \left(\frac{Z_{s\tau}}{w_s}\right)^{\theta}\right) c^{\theta}\right]. \tag{A.1}$$

Suppose that the automation rule for each task is monotonic, such that the labor cost draws $\widetilde{c}_{j\tau}(\nu)$ exceeding a task-specific threshold $c^*_{j\tau}$ are automated. Since firms are identical, this threshold is the same for all firms within the same industry, so we refer to it as $\bar{c}_{i\tau}$. Given the corresponding marginal cost for each task-variety ν, the optimal choice of task-variety input satisfies

$$y_{j\tau}(\nu) = \xi_{i\tau} c_{j\tau}(\nu)^{-\lambda} c_j^{\lambda} y_j,$$

for a firm that belongs to industry i, where c_j is an index of marginal costs given by $c_j^{1-\lambda} = \sum_\tau \xi_{i\tau} \int_0^1 c_{j\tau}(\nu)^{1-\lambda} d\nu$.

Substituting the above expression in Equation (5), the collection of thresholds $\overline{z}_j \equiv \left(\bar{z}_{j\tau}\right)_{\tau \in \mathcal{T}_i}$ is chosen to minimize costs at the occupation-skill level, where cost are given by

$$\begin{aligned} C_j(y) &= \min_{\overline{z}_j} \sum_{\tau \in \mathcal{T}} \int_0^1 \left(\xi_{i\tau} c_{j\tau}(\nu)^{1-\lambda} c_j^{\lambda} \frac{y_j}{a_j} + f_{j\tau}(\nu)\right) d\nu, \\ &= \min_{\overline{z}_j} \sum_{\tau \in \mathcal{T}} \int_0^1 \xi_{i\tau} c_{j\tau}(\nu)^{1-\lambda} c_j^{\lambda} \frac{y_j}{a_j} d\nu + \sum_{\tau \in \mathcal{T}} \int_{c^*_{i\tau}}^{\infty} \left(\frac{r}{z^f_{k\tau}}\right) dG_\tau(z), \end{aligned} \tag{A.2}$$

$$= \min_{\overline{z}_j} c_j \frac{y_j}{a_j} + \sum_{\tau \in \mathscr{T}} \exp\left[- \left(\sum_s \left(\frac{Z_{s\tau}}{w_s} \right)^\theta \right) \left(c^*_{j\tau} \right)^\theta \right] \frac{r}{z^f_{k\tau}}, \tag{A.3}$$

where $G_\tau(\cdot)$ denotes the cumulative function of the distribution of marginal costs in Equation (A.1) and where can further rewrite the firm-level marginal cost c_j as $c_j^{1-\lambda} = \sum_\tau \xi_{i\tau} c_{j\tau}^{1-\lambda}$ where the firm-task-type-level marginal costs satisfy

$$\begin{aligned}
c_{j\tau}^{1-\lambda} &= \int_0^1 c_{j\tau}(\nu)^{1-\lambda} d\nu, \\
&= \int_0^{c^*_{j\tau}} c^{1-\lambda} dG_\tau(c) + \int_{c^*_{j\tau}}^\infty \left(\frac{r}{z_{k\tau}} \right)^{1-\lambda} dG_\tau(z), \\
&= \theta \widetilde{c}_\tau^{-\theta} \int_0^{c^*_{j\tau}} c^{\theta-\lambda} \exp\left(- \left(\left(\frac{c}{\widetilde{c}_\tau} \right)^\theta \right) \right) dc + \left(\frac{r}{z_{k\tau}} \right)^{1-\lambda} \exp\left[- \left(\frac{c^*_{j\tau}}{\widetilde{c}_\tau} \right)^\theta \right], \\
&= \widetilde{c}_\tau^{1-\lambda} \int_0^{\left(\frac{c^*_{j\tau}}{\widetilde{c}_\tau} \right)^\theta} x^{\frac{\theta+1-\lambda}{\theta} - 1} \exp(-x)\, dx + \left(\frac{r}{z_{k\tau}} \right)^{1-\lambda} \exp\left[- \left(\frac{c^*_{j\tau}}{\widetilde{c}_\tau} \right)^\theta \right], \\
&= \widetilde{c}_\tau^{1-\lambda} \gamma\left(\left(\frac{c^*_{j\tau}}{\widetilde{c}_\tau} \right)^\theta ; 1 + \tfrac{1-\lambda}{\theta} \right) + \left(\frac{r}{z_{k\tau}} \right)^{1-\lambda} \exp\left[- \left(\frac{c^*_{j\tau}}{\widetilde{c}_\tau} \right)^\theta \right],
\end{aligned} \tag{A.4}$$

where we have defined $\widetilde{c}_\tau^{-\theta} \equiv \sum_s \left(\frac{Z_{s\tau}}{w_s} \right)^\theta$ and where in the last equality $\gamma(\vartheta; x) \equiv \int_0^x u^{\vartheta-1} e^{-u} du$ denotes the lower incomplete gamma function.

The first order condition of Equation (A.3) with respect to the threshold $c^*_{j\tau}$ is given by

$$\begin{aligned}
0 = \frac{\partial C_j}{\partial c^*_{j\tau}} &= \frac{\partial c_j}{\partial c^*_{j\tau}} \frac{y_j}{a_j} - g_\tau\left(c^*_{j\tau} \right) \frac{r}{z^f_{k\tau}}, \\
&= \frac{1}{1-\lambda} c_j^\lambda \xi_{i\tau} \frac{\partial c_{j\tau}^{1-\lambda}}{\partial c^*_{j\tau}} \frac{y_j}{a_j} - g_\tau\left(c^*_{j\tau} \right) \frac{r}{z^f_{k\tau}},
\end{aligned}$$

where $g_\tau(\cdot)$ is pdf corresponding to the distribution $G_\tau(\cdot)$ and where in the second equality we have used $c_j^{1-\lambda} = \sum_\tau \xi_{i\tau} c_{j\tau}^{1-\lambda}$. Next, we can use Equation (A.4) to compute the derivative term

$$\frac{\partial c_{j\tau}^{1-\lambda}}{\partial c^*_{j\tau}} = g_\tau\left(c^*_{j\tau} \right) \left[\left(c^*_{j\tau} \right)^{1-\lambda} - \left(\frac{r}{z_{k\tau}} \right)^{1-\lambda} \right].$$

Rearranging the terms in the first order condition, we find the condition

$$\frac{1}{1-\lambda} c_j^{\lambda} \xi_{i\tau} \frac{y_j}{a_j} \left[\left(c_{j\tau}^* \right)^{1-\lambda} - \left(\frac{r}{z_{k\tau}} \right)^{1-\lambda} \right] - \frac{r}{z_{k\tau}^f} = 0, \tag{A.5}$$

which we can simplify as

$$\begin{aligned} c_{j\tau}^* &= \frac{r}{z_{k\tau}} \left[1 + (1-\lambda) \frac{1}{\xi_{i\tau} c_j^{\lambda} y_j} \frac{r}{z_{k\tau}^f} \left(\frac{r}{z_{k\tau}} \right)^{\lambda-1} \right]^{\frac{1}{1-\lambda}}, \\ &= \frac{r}{z_{k\tau}} \left[1 + (1-\lambda) \frac{1}{\xi_{i\tau} \left(r / c_j z_{k\tau} \right)^{-\lambda} \left(y_j / a_j \right)} \frac{z_{k\tau}}{z_{k\tau}^f} \right]^{\frac{1}{1-\lambda}}. \end{aligned}$$

Let us consider the limit of $c_{j\tau}^*$ when $z_{k\tau} \to \infty$:

$$\begin{aligned} \lim_{z_{k\tau} \to \infty} c_{j\tau}^* &= \lim_{z_{k\tau} \to \infty} \frac{r}{z_{k\tau}} \left[1 + (1-\lambda) \frac{1}{\xi_{i\tau} \left(r / c_j \right)^{-\lambda} \left(y_j / a_j \right)} \frac{z_{k\tau}^{1-\lambda}}{z_{k\tau}^f} \right]^{\frac{1}{1-\lambda}}, \\ &= r \left[(1-\lambda) \frac{1}{\xi_{i\tau} \left(r / c_j \right)^{-\lambda} \left(y_j / a_j \right)} \frac{1}{z_{k\tau}^f} \right]^{\frac{1}{1-\lambda}}, \\ &= \left(\frac{(1-\lambda) r}{\xi_{i\tau} c_j^{\lambda} \left(y_j / a_j \right) z_{k\tau}^f} \right)^{\frac{1}{1-\lambda}}. \end{aligned}$$

□

Proof for Equations (14) and (15). First, we compute the share of task varieties for which skill group s is hired for work in any firm in industry i:

$$\begin{aligned} \mathbb{P}\left(x_{j\tau} = \ell_{js\tau} \right) &= \int_0^1 \mathbb{I}\left\{ \widetilde{c}_{j\tau}(\nu) = c_{s\tau}(\nu) \right\} \cdot \mathbb{I}\left\{ \widetilde{c}_{j\tau}(\nu) = c_{j\tau}(\nu) \right\} d\nu, \\ &= \mathbb{P}\left\{ \widetilde{c}_{j\tau}(\nu) = c_{s\tau}(\nu), \widetilde{c}_{j\tau}(\nu) = c_{j\tau}(\nu) \right\}, \\ &= \int_0^{c_{j\tau}^*} \prod_{s' \neq s} \mathbb{P}\left(c_{s'\tau}(\nu) \geq c \right) dG_{s\tau}(c), \\ &= \int_0^{c_{j\tau}^*} \prod_{s' \neq s} \exp\left(- \left(\frac{Z_{s'\tau}}{w_{s'}} \right)^{\theta} c^{\theta} \right) dG_{s\tau}(c), \\ &= \left(\frac{Z_{s\tau}}{w_s} \right)^{\theta} \int_0^{c_{j\tau}^*} \exp\left(- \sum_{s' \neq s} \left(\frac{Z_{s'\tau}}{w_{s'}} \right)^{\theta} c^{\theta} \right) \exp\left(- \left(\frac{Z_{s\tau}}{w_s} \right)^{\theta} c^{\theta} \right) c^{\theta-1} d\theta, \end{aligned}$$

$$= \left(\frac{Z_{s\tau}}{w_s}\right)^{\theta} \int_0^{c^*_{j\tau}} \exp\left(-\sum_{s'}\left(\frac{Z_{s'\tau}}{w_{s'}}\right)^{\theta} c^{\theta}\right) c^{\theta-1} d\theta,$$

$$= \frac{\left(\frac{Z_{s\tau}}{w_s}\right)^{\theta}}{\sum_{s'}\left(\frac{Z_{s'\tau}}{w_{s'}}\right)^{\theta}} \left[1 - \exp\left(-\left(\frac{c^*_{j\tau}}{\widetilde{c}_{\tau}}\right)^{\theta}\right)\right],$$

where the term in the square bracket on the right hand side of the last equation captures the share of nonautomated tasks. Thus, conditional on the task variety not getting automated, the share of task varieties produced by different skill groups does *not* vary by firm.

Next, we can show that conditional on choosing each skill group, the average marginal cost of producing the task-variety input is the same:

$$\begin{aligned}
\mathbb{P}\left(\widetilde{c}_{j\tau}(\nu) \leq c \,\middle|\, \widetilde{c}_{j\tau}(\nu) = c_{s\tau}(\nu)\right) &= \frac{\int_0^c \prod_{s'\neq s} \mathbb{P}\left(c_{s'\tau}(\nu) \geq c\right) dG_{s\tau}(c)}{\omega_{s,i\tau}}, \\
&= \frac{1 - \exp\left(-\left(\frac{c}{\widetilde{c}_{\tau}}\right)^{\theta}\right)}{1 - \exp\left(-\left(\frac{c^*_{j\tau}}{\widetilde{c}_{\tau}}\right)^{\theta}\right)}, \\
&= \mathbb{P}\left(\widetilde{c}_{j\tau}(\nu) \leq c \,\middle|\, \widetilde{c}_{j\tau}(\nu) \leq c^*_{j\tau}\right),
\end{aligned}$$

which does not depend on the skill type. Thus, all the variations across skill groups are due to the extensive margin of task varieties. The average payment per nonautomated variety is the same across all skill groups. From Equation (A.4), the share of all labor payments in the costs of a firm j in the production of task type τ is given by

$$\omega_{L,j\tau} = \left(\frac{\widetilde{c}_{\tau}}{c_{j\tau}}\right)^{\lambda-1} \left[1 - \exp\left(-\left(\frac{c}{\widetilde{c}_{\tau}}\right)^{\theta}\right)\right] \frac{\gamma\left(\left(\frac{c^*_{j\tau}}{\widetilde{c}_{\tau}}\right)^{\theta}; \frac{\theta+1-\lambda}{\theta}\right)}{1 - \exp\left(-\left(\frac{c^*_{j\tau}}{\widetilde{c}_{\tau}}\right)^{\theta}\right)}, \tag{A.6}$$

leading to Equation (14). For the labor demand in Equation (19), we have

$$\omega_{s,j\tau} = \frac{w_s \ell_{js\tau}}{c_{j\tau} y_{j\tau}} = \frac{w_s \ell_{js\tau}}{\sum_{s'} w_{s'} \ell_{js'\tau}} \omega_{\ell,j\tau},$$

$$= \frac{\left(\frac{Z_{s\tau}}{w_s}\right)^{\theta}}{\sum_{s'}\left(\frac{Z_{s'\tau}}{w_{s'}}\right)^{\theta}}\omega_{\ell,j\tau},$$

which leads to Equation (15). □

Proof for Lemma 2. We first show that the automation shares $\varphi_{j\tau}$ are also the solutions

$$C_j(\boldsymbol{w}, r, y_j) \equiv \min_{(\varphi_{j\tau}, y_{j\tau})_\tau} \sum_\tau \left(\widetilde{c}_\tau^{1-\lambda}\,\gamma\left(-\log\varphi_{j\tau}; 1+\tfrac{1-\lambda}{\theta}\right) + \left(\frac{r}{z_{k\tau}}\right)^{1-\lambda}\varphi_{j\tau}\right)^{\frac{1}{1-\lambda}} y_{j\tau} + \sum_\tau \frac{r}{z^f_{k\tau}}\varphi_{j\tau},$$

subject to the constraint imposed by Equation (1). Substituting for $c_{j\tau}$ from Equation (10), the first-order condition for $\varphi_{j\tau}$ is given by

$$0 = \tfrac{1}{1-\lambda}c_{j\tau}^{\lambda}\left(-\widetilde{c}_\tau^{1-\lambda}\,\gamma'\left(-\log\varphi_{j\tau}; 1+\tfrac{1-\lambda}{\theta}\right)\tfrac{1}{\varphi_{j\tau}} + \left(\frac{r}{z_{k\tau}}\right)^{1-\lambda}\right) y_{j\tau} + \frac{r}{z^f_{k\tau}}, \tag{A.7}$$

$$= -\widetilde{c}_\tau^{1-\lambda}\left(-\log\varphi_{j\tau}\right)^{\frac{1-\lambda}{\theta}} + \left(\frac{r}{z_{k\tau}}\right)^{1-\lambda} + (1-\lambda)\frac{c_{j\tau}^{-\lambda} r}{y_{j\tau} z^f_{k\tau}},$$

$$= -\left(-\log\varphi_{j\tau}\right)^{\frac{1-\lambda}{\theta}} + \left(\frac{r}{z_{k\tau}\widetilde{c}_\tau}\right)^{1-\lambda}\left[1 + (1-\lambda)\frac{z_{k\tau}/z^f_{k\tau}}{y_{j\tau}c_{j\tau}^{\lambda}\left(r/z_{k\tau}\right)^{-\lambda}}\right]. \tag{A.8}$$

The first-order condition for $y_{j\tau}$ is given by $c_{j\tau} = c_j\, y_j^{1/\lambda}\,\xi_{i\tau}^{1/\lambda} y_{j\tau}^{-1/\lambda}$,where c_j is the Lagrange multiplier corresponding to Equation (1) which, by envelope condition, satisfies $c_j = \partial C_j/\partial y_j$. Substituting the resulting expression for $y_{j\tau} = \xi_{i\tau} c_j^{\lambda} c_{j\tau}^{-\lambda}$ in Equation (A.8) gives us Equations (7) and (8). Applying Shephard's lemma, we can re-derive the expressions for skill-group and machine demand:

$$\ell_{s,j} = \frac{\partial C_j}{\partial w_s} = \sum_\tau c_{j\tau}^{\lambda}\widetilde{c}_\tau^{1-\lambda}\,\gamma_{j\tau}\frac{1}{\widetilde{c}_\tau}\frac{\partial\widetilde{c}_\tau}{\partial w_s}y_{j\tau} = \frac{c_j y_j}{w_s}\sum_\tau \frac{c_{j\tau}y_{j\tau}}{c_j y_j}\omega_{L,j\tau}\omega_{s,L\tau},$$

$$k_j = \frac{\partial C_j}{\partial r} = \sum_\tau c_{j\tau}^{\lambda}\left(\frac{r}{z_{k\tau}}\right)^{1-\lambda}\frac{1}{r}y_{j\tau} + \sum_\tau \frac{\varphi_{j\tau}}{z^f_{k\tau}} = \frac{c_j y_j}{r}\sum_\tau \frac{c_{j\tau}y_{j\tau}}{c_j y_j}\omega_{K,j\tau} + \sum_\tau \frac{\varphi_{j\tau}}{z^f_{k\tau}}.$$

The envelope condition also leads to $\partial C_j/\partial y_j = c_j$.

Next, we derive expressions for the partial derivatives of marginal costs c_j with respect

to factor prices and scale. For the price of machines, we have

$$\begin{aligned}\frac{\partial c_{j\tau}}{\partial r} &= c_{j\tau}^{\lambda}\left[\left(\frac{r}{z_{k\tau}}\right)^{1-\lambda}\frac{1}{r}+\frac{1}{1-\lambda}\left(-\widetilde{c}_{\tau}^{1-\lambda}\,\gamma'\left(-\log\varphi_{j\tau};1+\tfrac{1-\lambda}{\theta}\right)\frac{1}{\varphi_{j\tau}}+\left(\frac{r}{z_{k\tau}}\right)^{1-\lambda}\right)\frac{\partial\varphi_{j\tau}}{\partial r}\right],\\ &= \frac{c_{j\tau}}{r}\omega_{K,j}-\frac{r}{z_{k\tau}^{f}y_{j\tau}}\frac{\partial\varphi_{j\tau}}{\partial r},\end{aligned}$$

where in the second equality, we have used Equation (A.7). Similarly, for the skill-s wages, we have

$$\begin{aligned}\frac{\partial c_{j\tau}}{\partial w_s} &= c_{j\tau}^{\lambda}\left[\widetilde{c}_{\tau}^{1-\lambda}\,\gamma\,\tfrac{1}{\widetilde{c}_{\tau}}\frac{\partial\widetilde{c}_{\tau}}{\partial w_s}+\frac{1}{1-\lambda}\left(-\widetilde{c}_{\tau}^{1-\lambda}\,\gamma'\left(-\log\varphi_{j\tau};1+\tfrac{1-\lambda}{\theta}\right)\frac{1}{\varphi_{j\tau}}+\left(\frac{r}{z_{k\tau}}\right)^{1-\lambda}\right)\frac{\partial\varphi_{j\tau}}{\partial w_s}\right],\\ &= \frac{c_{j\tau}}{w_s}\omega_{L,j}\omega_{s,L\tau}-\frac{r}{z_{k\tau}^{f}y_{j\tau}}\frac{\partial\varphi_{j\tau}}{\partial w_s}.\end{aligned}$$

Finally, the scale elasticity of the marginal cost is given by

$$\begin{aligned}\frac{\partial c_{j\tau}}{\partial y_j} &= c_{j\tau}^{\lambda}\tfrac{1}{1-\lambda}\left(-\widetilde{c}_{\tau}^{1-\lambda}\,\gamma'\left(-\log\varphi_{j\tau};1+\tfrac{1-\lambda}{\theta}\right)\frac{1}{\varphi_{j\tau}}+\left(\frac{r}{z_{k\tau}}\right)^{1-\lambda}\right)\frac{\partial\varphi_{j\tau}}{\partial y_j},\\ &= -\frac{r}{z_{k\tau}^{f}y_{j\tau}}\frac{\partial\varphi_{j\tau}}{\partial y_j},\end{aligned}$$

where we have again have used Equation (A.7) in the last line. Combining these results, we can write the elasticities of marginal cost as

$$\frac{\partial\log c_{j\tau}}{\partial\log r} = \omega_{K,j\tau}-\frac{r/z_{k\tau}^{f}\varphi_{j\tau}}{c_{j\tau}y_{j\tau}}\frac{\partial\log\varphi_{j\tau}}{\partial\log r}, \tag{A.9}$$

$$\frac{\partial\log c_{j\tau}}{\partial\log w_s} = \omega_{L,j\tau}\omega_{s,L\tau}-\frac{r/z_{k\tau}^{f}\varphi_{j\tau}}{c_{j\tau}y_{j\tau}}\frac{\partial\log\varphi_{j\tau}}{\partial\log w_s}, \tag{A.10}$$

$$\frac{\partial\log c_{j\tau}}{\partial\log y_j} = -\frac{r/z_{k\tau}^{f}\varphi_{j\tau}}{c_{j\tau}y_{j\tau}}\frac{\partial\log\varphi_{j\tau}}{\partial\log y_j}. \tag{A.11}$$

We now need to compute the elasticities of the automation share $\varphi_{j\tau}$ with respect to factor prices and scale. Using Equation

$$\frac{\partial\log\varphi_{j\tau}}{\partial\log r} = -\theta\left(\frac{c_{j\tau}^{*}}{\widetilde{c}_{\tau}}\right)^{\theta}\frac{\partial\log c_{j\tau}^{*}}{\partial\log r} = \theta\,\log\varphi_{j\tau}\left(1+\tfrac{\lambda}{1-\lambda}\left(1-\frac{\partial\log c_j}{\partial\log r}\right)\zeta_{j\tau}\right), \tag{A.12}$$

$$\frac{\partial \log \varphi_{j\tau}}{\partial \log w_s} = -\theta \left(\frac{c^*_{j\tau}}{\widetilde{c}_\tau}\right)^\theta \left(\frac{\partial \log c^*_{j\tau}}{\partial \log w_s} - \omega_{s,L\tau}\right) = \theta \log \varphi_{j\tau} \left(-\tfrac{\lambda}{1-\lambda}\zeta_{j\tau}\frac{\partial \log c_j}{\partial \log w_s} - \omega_{s,L\tau}\right), \tag{A.13}$$

$$\frac{\partial \log \varphi_{j\tau}}{\partial \log y_j} = -\theta \left(\frac{c^*_{j\tau}}{\widetilde{c}_\tau}\right)^\theta \frac{\partial \log c^*_{j\tau}}{\partial \log y_j} = -\theta \log \varphi_{j\tau} \tfrac{1}{1-\lambda}\zeta_{j\tau} \left(1 + \lambda \frac{\partial \log c_j}{\partial \log y_j}\right),$$

where we have used the definition of $\zeta_{j\tau}$ in Equation (31). Substituting these expressions in Equations (A.9)-(A.11), we then find

$$\begin{aligned}
\frac{\partial \log c_j}{\partial \log r} &= \sum_\tau \psi_{\tau,j} \frac{\partial \log c_{j\tau}}{\partial \log r}, \\
&= \omega_{K,j} - \theta \sum_\tau \psi_{\tau,j}\omega_{f,j\tau} \log \varphi_{j\tau} \left(1 + \tfrac{\lambda}{1-\lambda}\left(1 - \frac{\partial \log c_j}{\partial \log r}\right)\zeta_{j\tau}\right), \\
&= \tfrac{1}{1-\lambda\frac{\theta}{1-\lambda}\left(\sum_\tau \psi_{\tau,j}\omega_{f,j\tau} \log \varphi_{j\tau}\zeta_{j\tau}\right)} \left(\omega_{K,j} - \theta\left(\sum_\tau \psi_{\tau,j}\,\omega_{f,j\tau} \log \varphi_{j\tau}\left(1 + \tfrac{\lambda}{1-\lambda}\zeta_{j\tau}\right)\right)\right),
\end{aligned} \tag{A.14}$$

$$\begin{aligned}
\frac{\partial \log c_j}{\partial \log w_s} &= \sum_\tau \psi_{\tau,j} \frac{\partial \log c_{j\tau}}{\partial \log w_s}, \\
&= \omega_{s,j} + \theta \sum_\tau \psi_{\tau,j}\omega_{f,j\tau} \log \varphi_{j\tau} \left(\tfrac{\lambda}{1-\lambda}\zeta_{j\tau}\frac{\partial \log c_j}{\partial \log w_s} + \omega_{s,L\tau}\right), \\
&= \tfrac{1}{1-\lambda\frac{\theta}{1-\lambda}\left(\sum_\tau \psi_{\tau,j}\omega_{f,j\tau} \log \varphi_{j\tau}\zeta_{j\tau}\right)} \left(\omega_{s,j} + \theta \sum_\tau \psi_{\tau,j}\,\omega_{f,j\tau} \log \varphi_{j\tau}\,\omega_{s,L\tau}\right),
\end{aligned} \tag{A.15}$$

$$\begin{aligned}
\frac{\partial \log c_j}{\partial \log y_j} &= \sum_\tau \psi_{\tau,j} \frac{\partial \log c_{j\tau}}{\partial \log y_j}, \\
&= \theta \sum_\tau \psi_{\tau,j}\omega_{f,j\tau} \tfrac{1}{1-\lambda}\zeta_{j\tau} \left(1 + \lambda \frac{\partial \log c_j}{\partial \log y_j}\right), \\
&= \frac{\frac{\theta}{1-\lambda}\left(\sum_\tau \psi_{\tau,j}\omega_{f,j\tau} \log \varphi_{j\tau}\zeta_{j\tau}\right)}{1-\lambda\frac{\theta}{1-\lambda}\left(\sum_\tau \psi_{\tau,j}\omega_{f,j\tau} \log \varphi_{j\tau}\zeta_{j\tau}\right)}.
\end{aligned}$$

The last expression leads to Equation (30), which can be rewritten as

$$\frac{1}{1 + \lambda\varepsilon^m_j} = 1 - \lambda\frac{\theta}{1-\lambda}\left(\sum_\tau \psi_{\tau,j}\omega_{f,j\tau} \log \varphi_{j\tau}\,\zeta_{j\tau}\right). \tag{A.16}$$

Finally, we again use Shephard's lemma to relate the factor price elasticities of marginal

costs to scale elasticities of factor demand:

$$\frac{\partial \log c_j}{\partial \log r} = \frac{r}{c_j}\frac{\partial c_j}{\partial r} = \frac{r}{c_j}\frac{\partial^2 C_j}{\partial r \partial y_j} = \frac{r}{c_j}\frac{\partial k_j}{\partial y_j} = \frac{r k_j}{c_j y_j}\frac{\partial \log k_j}{\partial \log y_j} = \left(\omega_{K,j} + \omega_{f,j}\right)\, \varepsilon_{K,j}, \tag{A.17}$$

$$\frac{\partial \log c_j}{\partial \log w_s} = \frac{w_s}{c_j}\frac{\partial c_j}{\partial w_s} = \frac{w_s}{c_j}\frac{\partial^2 C_j}{\partial w_s \partial y_j} = \frac{w_s}{c_j}\frac{\partial \ell_{s,j}}{\partial y_j} = \frac{w_s \ell_{s,j}}{c_j y_j}\frac{\partial \log \ell_{s,j}}{\partial \log y_j} = \omega_{s,j}\, \varepsilon_{s,j}, \tag{A.18}$$

which we can combine with Equations (A.14), (A.15), and (A.16) to find Equations (32) and (33). □

Proof for Lemma 3. First, note that the firm-and-task level labor demand is given by

$$\begin{aligned}
\ell_{s,j\tau} &= \frac{c_j y_j}{w_s}\left(\frac{c_{j\tau}}{c_j}\right)^{1-\lambda}\left(\frac{\widetilde{c}_\tau}{c_{j\tau}}\right)^{1-\lambda}\gamma\left(-\log\varphi_{j\tau}\right)\, \omega_{s,L\tau}, \\
&= \frac{y_j}{w_s}\, c_j^{\lambda}\widetilde{c}_\tau^{1-\lambda}\gamma\left(-\log\varphi_{j\tau}\right)\, \omega_{s,L\tau}.
\end{aligned}$$

We now compute the elasticity of the skill demand with respect to the price of machines:

$$\begin{aligned}
\frac{\partial \log \ell_{s,j\tau}}{\partial \log r} &= \lambda\frac{\partial \log c_j}{\partial \log r} - \left(\frac{\gamma'(x)}{\gamma(x)}\right)\Bigg|_{x=-\log\varphi_{j\tau}}\frac{\partial \log \varphi_{j\tau}}{\partial \log r}, \\
&= \lambda\,\frac{\partial \log c_j}{\partial \log r} - \varphi_{j\tau}\frac{\left(-\log\varphi_{j\tau}\right)^{\frac{1-\lambda}{\theta}}}{\gamma\left(-\log\varphi_{j\tau}\right)}\frac{\partial \log \varphi_{j\tau}}{\partial \log r}, \\
&= \lambda\,\frac{\partial \log c_j}{\partial \log r} - \varphi_{j\tau}\frac{\left(\frac{r}{z_{k\tau}\widetilde{c}_\tau}\right)^{1-\lambda}\left[1+(1-\lambda)\frac{z_{k\tau}/z_{k\tau}^f}{y_{j\tau}c_{j\tau}^{\lambda}(r/z_{k\tau})^{-\lambda}}\right]}{\omega_{L,j\tau}/\left(\widetilde{c}_\tau/c_{j\tau}\right)^{1-\lambda}}\theta\,\log\varphi_{j\tau}\,\left(1+\frac{\lambda}{1-\lambda}\left(1-\frac{\partial \log c_j}{\partial \log r}\right)\zeta_{j\tau}\right), \\
&= \lambda\,\frac{\partial \log c_j}{\partial \log r} - \theta\frac{\omega_{K,j\tau}}{\omega_{L,j\tau}}\frac{\log\varphi_{j\tau}}{1-\zeta_{j\tau}}\,\left(1+\frac{\lambda}{1-\lambda}\left(1-\frac{\partial \log c_j}{\partial \log r}\right)\zeta_{j\tau}\right), \\
&= \lambda\left(1+\frac{\theta}{1-\lambda}\frac{\zeta_{j\tau}}{1-\zeta_{j\tau}}\frac{\omega_{K,j\tau}}{\omega_{L,j\tau}}\log\varphi_{j\tau}\right)\,\frac{\partial \log c_j}{\partial \log r} - \theta\frac{\omega_{K,j\tau}}{\omega_{L,j\tau}}\frac{1+\frac{\lambda}{1-\lambda}\zeta_{j\tau}}{1-\zeta_{j\tau}}\,\log\varphi_{j\tau}, \\
&= \lambda\left(1+\frac{\theta}{1-\lambda}\frac{\zeta_{j\tau}}{1-\zeta_{j\tau}}\frac{\omega_{K,j\tau}}{\omega_{L,j\tau}}\log\varphi_{j\tau}\right)\,\left(\omega_{K,j}+\omega_{f,j}\right)\varepsilon_{K,j} - \theta\frac{\omega_{K,j\tau}}{\omega_{L,j\tau}}\frac{1+\frac{\lambda}{1-\lambda}\zeta_{j\tau}}{1-\zeta_{j\tau}}\,\log\varphi_{j\tau}, \\
&= \lambda\left(\omega_{K,j}+\omega_{f,j}\right)\varepsilon_{K,j} + \left(\frac{\lambda\theta}{1-\lambda}\frac{\zeta_{j\tau}}{1-\zeta_{j\tau}}\frac{\omega_{K,j\tau}}{\omega_{L,j\tau}}\log\varphi_{j\tau}\right)\left(\left(\omega_{K,j}+\omega_{f,j}\right)\varepsilon_{K,j}-1\right) \\
&\quad - \theta\frac{\omega_{K,j\tau}}{\omega_{L,j\tau}}\frac{\log\varphi_{j\tau}}{1-\zeta_{j\tau}},
\end{aligned} \tag{A.19}$$

$$= \lambda \left(\omega_{K,j} + \omega_{f,j}\right) \varepsilon_{K,j} + \lambda\theta \frac{\omega_{f,j\tau}}{\omega_{L,j\tau}} \log \varphi_{j\tau} \left(\left(\omega_{K,j} + \omega_{f,j}\right) \varepsilon_{K,j} - 1\right) - \theta \frac{\omega_{K,j\tau}}{\omega_{L,j\tau}} \frac{\log \varphi_{j\tau}}{1-\zeta_{j\tau}}, \tag{A.20}$$

where in the second equality, we have used the definition of the incomplete gamma function from Lemma 1, in the third equality, we have used Equations (7), (31), (8), and (14), in the fourth equality, we have used Equation (A.12), in the sixth equation, we have substituted from Equation (A.17), and in the last equality, we have used the following result:

$$\begin{aligned}
\frac{\zeta_{j\tau}}{1-\zeta_{j\tau}} \omega_{K,j\tau} &= \left(\frac{r}{z_{k\tau} c_{j\tau}}\right)^{1-\lambda} \varphi_{j\tau} \left(1-\lambda\right) \frac{z_{k\tau}/z^f_{k\tau}}{\xi_{i\tau} \left(r/z_{k\tau} c_j\right)^{-\lambda} y_j/a_j}, \\
&= \left(1-\lambda\right) \frac{r\varphi_{j\tau}/z^f_{k\tau}}{\xi_{i\tau} \left(c_{j\tau}/c_j\right)^{1-\lambda} c_j y_j/a_j}, \\
&= \left(1-\lambda\right) \omega_{f,j\tau}.
\end{aligned}$$

Next, we aggregate across tasks within the firm:

$$\begin{aligned}
\sigma^{s,K}_j = \frac{1}{\Omega_{K,j}} \frac{\partial \log \ell_{s,j}}{\partial \log r} &= \frac{1}{\Omega_{K,j}} \sum_\tau \frac{\ell_{s,j\tau}}{\ell_{s,j}} \frac{\partial \log \ell_{s,j\tau}}{\partial \log r}, \\
&= \frac{1}{\Omega_{K,j}} \sum_\tau \psi_{\tau,j} \frac{\omega_{s,j\tau}}{\omega_{s,j}} \frac{\partial \log \ell_{s,j\tau}}{\partial \log r}, \\
&= \lambda \frac{\left(\omega_{K,j} + \omega_{f,j}\right) \varepsilon_{K,j}}{\Omega_{K,j}} \\
&\quad + \lambda \frac{\left(\omega_{K,j} + \omega_{f,j}\right) \varepsilon_{K,j} - 1}{\Omega_{K,j}} \left(\theta \frac{\omega_{f,j\tau}}{\omega_{L,j}} \mathbb{E}^f_{\tau|j} \left[\frac{\omega_{s,L\tau}}{\omega_{s,Lj}} \log \varphi_{j\tau}\right]\right) \\
&\quad - \theta \frac{\omega_{K,j}}{\Omega_{K,j}} \mathbb{E}_{\tau|j} \left[\frac{\omega_{s,L\tau}}{\omega_{s,j}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \frac{\log \varphi_{j\tau}}{1-\zeta_{j\tau}}\right], \\
&= \lambda \frac{\varepsilon_{K,j}}{\varepsilon_j} + \lambda \left(\frac{\varepsilon_{K,j}}{\varepsilon_j} - \frac{1}{\Omega_{K,j}}\right) \left(\frac{\varepsilon_{s,j}}{1+\lambda\varepsilon^m_j} - 1\right) \\
&\quad - \theta \left(1 + \omega_{f,i}\right) \frac{\omega_{K,j}}{\omega_{K,j} + \omega_{f,j}} \mathbb{E}_{\tau|j} \left[\frac{\omega_{s,L\tau}}{\omega_{s,j}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \frac{\log \varphi_{j\tau}}{1-\zeta_{j\tau}}\right], \\
&= \lambda \frac{\varepsilon_{K,j}}{\varepsilon_j} + \lambda \left(\frac{\varepsilon_{K,j}}{\varepsilon_j} - \frac{1}{\Omega_{K,j}}\right) \left(\frac{\varepsilon_{s,j}}{1+\lambda\varepsilon^m_j} - 1\right) \\
&\quad - \frac{\theta}{\varepsilon_j} \frac{\omega_{K,j}/\omega_{L,j}}{\omega_{K,j} + \omega_{f,j}} \mathbb{E}_{\tau|j} \left[\frac{\omega_{s,L\tau}}{\omega_{s,Lj}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \frac{\log \varphi_{j\tau}}{1-\zeta_{j\tau}}\right],
\end{aligned}$$

where in the third equality, we have substituted for the task-level elasticity from Equation (A.20), and in the fourth equality, we have substituted from Equation (33) to find

$$1-\tfrac{\theta\lambda}{1-\lambda}\omega_{f,j}\,\mathbb{E}^{f}_{\tau|j}\left[\log\varphi_{j\tau}\,\zeta_{j\tau}\right]=1-\frac{\lambda\varepsilon^{m}_{j}}{1+\lambda\varepsilon^{m}_{j}}=\frac{1}{1+\lambda\varepsilon^{m}_{j}}.$$

Similarly, we can compute the elasticity of task-level skill-s demand with respect to the price of skill-s':

$$\begin{aligned}
\frac{\partial\log\ell_{s,j\tau}}{\partial\log w_{s'}}&=(1-\lambda)\,\omega_{s',L\tau}+\frac{\partial\log\omega_{s,L\tau}}{\partial\log w_{s'}}+\lambda\frac{\partial\log c_j}{\partial\log w_{s'}}-\left(\frac{\gamma'(x)}{\gamma(x)}\right)\Bigg|_{x=-\log\varphi_{j\tau}}\frac{\partial\log\varphi_{j\tau}}{\partial\log w_{s'}},\\
&=(1-\lambda)\,\omega_{s',L\tau}+\frac{\partial\log\omega_{s,L\tau}}{\partial\log w_{s'}}+\lambda\,\frac{\partial\log c_j}{\partial\log w_{s'}}-\varphi_{j\tau}\frac{\left(-\log\varphi_{j\tau}\right)^{\frac{1-\lambda}{\theta}}}{\gamma\left(-\log\varphi_{j\tau}\right)}\frac{\partial\log\varphi_{j\tau}}{\partial\log w_{s'}},\\
&=(1+\theta-\lambda)\,\omega_{s',L\tau}+\lambda\,\frac{\partial\log c_j}{\partial\log w_{s'}}\\
&\quad-\varphi_{j\tau}\frac{\left(\frac{r}{z_{k\tau}\widetilde{c}_\tau}\right)^{1-\lambda}\left[1+(1-\lambda)\,\frac{z_{k\tau}/z^{f}_{k\tau}}{y_{j\tau}c^{\lambda}_{j\tau}(r/z_{k\tau})^{-\lambda}}\right]}{\omega_{L,j\tau}/\left(\widetilde{c}_\tau/c_{j\tau}\right)^{1-\lambda}}\\
&\quad\times\theta\,\log\varphi_{j\tau}\left(-\tfrac{\lambda}{1-\lambda}\zeta_{j\tau}\frac{\partial\log c_j}{\partial\log w_{s'}}-\omega_{s',L\tau}\right),\\
&=(1+\theta-\lambda)\,\omega_{s',L\tau}+\lambda\,\frac{\partial\log c_j}{\partial\log w_{s'}}+\theta\frac{\omega_{K,j\tau}}{\omega_{L,j\tau}}\frac{\log\varphi_{j\tau}}{1-\zeta_{j\tau}}\left(\omega_{s',L\tau}+\tfrac{\lambda}{1-\lambda}\zeta_{j\tau}\frac{\partial\log c_j}{\partial\log w_{s'}}\right),\\
&=(1+\theta-\lambda)\,\omega_{s',L\tau}+\lambda\left(1+\tfrac{\theta}{1-\lambda}\frac{\zeta_{j\tau}}{1-\zeta_{j\tau}}\frac{\omega_{K,j\tau}}{\omega_{L,j\tau}}\log\varphi_{j\tau}\right)\frac{\partial\log c_j}{\partial\log w_{s'}}\\
&\quad+\theta\frac{\omega_{K,j\tau}}{\omega_{L,j\tau}}\frac{1}{1-\zeta_{j\tau}}\,\log\varphi_{j\tau}\,\omega_{s',L\tau},\\
&=\left(1-\lambda+\theta\left(1+\frac{\omega_{K,j\tau}}{\omega_{L,j\tau}}\frac{\log\varphi_{j\tau}}{1-\zeta_{j\tau}}\right)\right)\omega_{s',L\tau}\\
&\quad+\lambda\left(1+\tfrac{\theta}{1-\lambda}\frac{\zeta_{j\tau}}{1-\zeta_{j\tau}}\frac{\omega_{K,j\tau}}{\omega_{L,j\tau}}\log\varphi_{j\tau}\right)\omega_{s',j}\,\varepsilon_{s',j},
\end{aligned}\tag{A.21}$$

where, again, in the third equality, we have substituted from Equation (A.13), in last equality, we have substituted from Equation (A.18)

Aggregating across tasks, we can find the elasticity at the firm level:

$$\sigma^{s,s'}_{j}\equiv\frac{1}{\Omega_{s',j}}\frac{\partial\log\ell_{s,j}}{\partial\log w_{s'}},$$

$$
\begin{aligned}
&= \frac{1}{\Omega_{s',j}} \sum_{\tau} \frac{\ell_{j\tau}}{\ell_j} \frac{\partial \log \ell_{s,j\tau}}{\partial \log w_{s'}}, \\
&= \frac{1}{\Omega_{s',j}} \sum_{\tau} \psi_{\tau,j} \frac{\omega_{s,j\tau}}{\omega_{s,j}} \frac{\partial \log \ell_{s,j\tau}}{\partial \log w_{s'}}, \\
&= \frac{1}{\Omega_{s',j}} \sum_{\tau} \psi_{\tau,j} \frac{\omega_{s,j\tau}}{\omega_{s,j}} \left(1 - \lambda + \theta \left(1 + \frac{\omega_{K,j\tau}}{\omega_{L,j\tau}} \frac{\log \varphi_{j\tau}}{1-\zeta_{j\tau}}\right)\right) \omega_{s',L\tau} \\
&\quad + \frac{1}{\Omega_{s',j}} \lambda \left(1 + \tfrac{\theta}{1-\lambda} \sum_{\tau} \psi_{\tau,j} \frac{\omega_{s,j\tau}}{\omega_{s,j}} \frac{\zeta_{j\tau}}{1-\zeta_{j\tau}} \frac{\omega_{K,j\tau}}{\omega_{L,j\tau}} \log \varphi_{j\tau}\right) \omega_{s',j}\, \varepsilon_{s',j}, \\
&= \tfrac{1-\lambda+\theta}{1+\omega_{f,j}} \mathbb{E}_{\tau|j} \left[\omega_{L,j\tau} \frac{\omega_{s,L\tau}}{\Omega_{s,j}} \frac{\omega_{s',L\tau}}{\Omega_{s',j}}\right] + \tfrac{\theta}{1+\omega_{f,j}} \mathbb{E}_{\tau|j} \left[\omega_{K,j\tau} \frac{\log \varphi_{j\tau}}{1-\zeta_{j\tau}} \frac{\omega_{s,L\tau}}{\Omega_{s,j}} \frac{\omega_{s',L\tau}}{\Omega_{s',j}}\right] \\
&\quad + \frac{\lambda}{\varepsilon_j} \left(1 + \theta \tfrac{\omega_{f,j}}{\omega_{L,j}} \mathbb{E}^{f}_{\tau|j} \left[\log \varphi_{j\tau} \frac{\omega_{s,L\tau}}{\omega_{s,Lj}}\right]\right) \varepsilon_{s',j}, \\
&= \tfrac{1-\lambda+\theta}{\varepsilon_j} \mathbb{E}_{\tau|j} \left[\omega_{L,j\tau} \frac{\omega_{s,L\tau}}{\omega_{s,j}} \frac{\omega_{s',L\tau}}{\omega_{s',j}}\right] + \tfrac{\theta}{\varepsilon_{f,j}} \mathbb{E}_{\tau|j} \left[\omega_{K,j\tau} \frac{\log \varphi_{j\tau}}{1-\zeta_{j\tau}} \frac{\omega_{s,L\tau}}{\omega_{s,j}} \frac{\omega_{s',L\tau}}{\omega_{s',j}}\right] \\
&\quad + \tfrac{\lambda}{\varepsilon_j} \tfrac{\varepsilon_{s,j}\, \varepsilon_{s',j}}{1+\lambda\, \varepsilon^{m}_{j}},
\end{aligned}
$$

where in the last equation, we have used Equation (33).

In the case without fixed costs ($z^{f}_{k\tau} = \infty$, $\zeta_{j\tau} = \omega_{f,j\tau} = 0$), Equations (A.20) and (A.21) simplify to

$$
\sigma^{s,K}_{j\tau} \equiv \frac{1}{\omega_{K,j\tau}} \frac{\partial \log \ell_{s,j\tau}}{\partial \log r} = \lambda - \theta \frac{1}{\omega_{L,j\tau}} \log \varphi_{j\tau}, \tag{A.22}
$$

$$
\sigma^{s,s'}_{j\tau} \equiv \frac{1}{\omega_{s',j\tau}} \frac{\partial \log \ell_{s,j\tau}}{\partial \log w_{s'}} = \lambda \frac{\omega_{s',j}}{\omega_{s',j\tau}} + \left(1 - \lambda + \theta \left(1 + \frac{\omega_{K,j\tau}}{\omega_{L,j\tau}} \log \varphi_{j\tau}\right)\right) \frac{1}{\omega_{L,j\tau}}. \tag{A.23}
$$

to the price of machines is increasing in the automation share across tasks. □

Proof for Proposition 1. First, note that that the for all $s \neq o$, we have

$$
\begin{aligned}
0 = \frac{d \log \overline{L}_s}{d \log r} = \frac{\partial \log L_s}{\partial \log r} &= \frac{\partial \log L_s}{\partial \log r} + \frac{\partial \log L_s}{\partial \log w_s} \frac{d \log w_s}{d \log r} + \sum_{s' \neq s} \frac{\partial \log L_{s'}}{\partial \log w_{s'}} \frac{d \log w_{s'}}{d \log r} \\
&\quad + \frac{\partial \log L_s}{\partial \log Y} \frac{d \log Y}{d \log r}, \\
&= \Omega_K \sigma^{s,K} - \left(\Omega_K \sigma^{s,K} + \sum_{s' \neq s} \Omega_{s'} \sigma^{s,s'}\right) \frac{d \log w_s}{d \log r} \\
&\quad + \sum_{s' \neq o,s} \Omega_{s'} \sigma^{s,s'} \frac{d \log w_{s'}}{d \log r} + \bar{\varepsilon}_s \frac{d \log Y}{d \log r},
\end{aligned}
$$

where in the second equality, we have used the fact that $\frac{\partial \log L_s}{\partial \log w_s} = -\Omega_K \sigma^{s,K} - \sum_{s' \neq s} \Omega_{s'} \sigma^{s,s'}$ and $\frac{d \log w_o}{d \log r} = 0$. This leads to the following set of equations for all $s \neq o$,

$$\left(\Omega_K \sigma^{s,K} + \sum_{s' \neq s} \Omega_{s'} \sigma^{s,s'} \right) \frac{d \log w_s}{d \log r} - \sum_{s' \neq o,s} \Omega_{s'} \sigma^{s,s'} \frac{d \log w_{s'}}{d \log r} - \bar{\varepsilon}_s \frac{d \log Y}{d \log r} = \Omega_K \sigma^{s,K},$$

and

$$- \sum_{s' \neq o} \Omega_o \sigma^{o,s'} \frac{d \log w_{s'}}{d \log r} - \bar{\varepsilon}_o \frac{d \log Y}{d \log r} = \Omega_K \sigma^{o,K},$$

which we can write as

$$\boldsymbol{M} \begin{bmatrix} \frac{d \log \boldsymbol{w}_-}{d \log r} \\ \frac{d \log Y}{d \log r} \end{bmatrix} = \boldsymbol{\sigma}^K \Omega_K,$$

for $\boldsymbol{M}$ and $\boldsymbol{\sigma}^K$ defined as in Equation (29). Additionally, we have $\frac{d \log P}{d \log r} = \Omega_K + \sum_{s \neq o} \Omega_{s'} \frac{d \log w_{s'}}{d \log r} + \varepsilon^m \frac{d \log Y}{d \log r}$, which implies

$$\begin{aligned} \frac{d \log \boldsymbol{w}^R}{d \log r} &= \frac{d \log \boldsymbol{w}}{d \log r} - \frac{d \log P}{d \log r}, \\ &= \boldsymbol{I}_{-o} \boldsymbol{M}^{-1} \boldsymbol{\sigma}^K \Omega_K - \Omega_K - \boldsymbol{\Omega}'_m \boldsymbol{M}^{-1} \boldsymbol{\sigma}^K \Omega_K, \\ &= -\Omega_K \left(1 + \left(\mathbf{1} \boldsymbol{\Omega}'_m - \boldsymbol{I}_{-o} \boldsymbol{M}^{-1} \right) \boldsymbol{M}^{-1} \boldsymbol{\sigma}^K \right) \end{aligned}$$

leading to the desired result. □

Proof for $\varepsilon_{K,j} = -\varepsilon^m_j / \omega^f_{j,i}$ in the case without marginal costs. In this case from Equation (30) with $\zeta_{j\tau} \equiv 1$ we find

$$\lambda \, \varepsilon^m_j = \frac{\frac{\lambda}{1-\lambda} \theta \, \omega_{f,j} \, \mathbb{E}^f_{\tau|j} \left[\log \varphi_{j\tau} \right]}{1 - \frac{\lambda}{1-\lambda} \theta \, \omega_{f,j} \, \mathbb{E}^f_{\tau|j} \left[\log \varphi_{j\tau} \right]},$$

which, in combination with Equation (32) implies

$$\begin{aligned} \varepsilon_{K,j} &= -\lambda \frac{\varepsilon^m_j}{\omega_{f,j}} - \theta \left(1 + \lambda \varepsilon^m_j \right) \mathbb{E}^f_{\tau|j} \left[\log \varphi_{j\tau} \right], \\ &= -\frac{1}{\omega_{f,j}} \left(\lambda \, \varepsilon^m_j + \frac{1-\lambda}{\lambda} \left(1 + \lambda \varepsilon^m_j \right) \frac{\lambda}{1-\lambda} \theta \, \omega_{f,j} \, \mathbb{E}^f_{\tau|j} \left[\log \varphi_{j\tau} \right] \right), \\ &= -\frac{1}{\omega_{f,j}} \left(\lambda \, \varepsilon^m_j + \frac{1-\lambda}{\lambda} \left(1 + \lambda \varepsilon^m_j \right) \frac{\lambda}{1-\lambda} \theta \, \omega_{f,j} \, \mathbb{E}^f_{\tau|j} \left[\log \varphi_{j\tau} \right] \right), \\ &= -\frac{1}{\omega_{f,j}} \left(\lambda \, \varepsilon^m_j + \frac{1-\lambda}{\lambda} \lambda \varepsilon^m_j \right), \end{aligned}$$

$$= -\frac{\varepsilon_j^m}{\omega_{f,j}}.$$

□

Proof for Corollary 1. From Equation (A.20) in the case without fixed costs, we have

$$\begin{aligned}
\frac{\partial \log \ell_{j\tau}}{\partial \log r} &= \left(\lambda - \frac{\theta}{\omega_{K,j}} \frac{\omega_{K,j\tau}}{1-\omega_{K,j\tau}} \log \varphi_{j\tau}\right) \omega_{K,j}, \\
&= \left(\lambda + \frac{\theta}{\omega_{K,j}} \frac{\widehat{\varphi}_{j\tau}^{1+\frac{1-\lambda}{\theta}} e^{-\widehat{\varphi}_{j\tau}}}{\gamma\left(\widehat{\varphi}_{j\tau}; 1+\frac{1-\lambda}{\theta}\right)}\right) \omega_{K,j}, \\
&= \left(\lambda + \frac{\theta}{\omega_{K,j}} F\left(\widehat{\varphi}_{j\tau}\right)\right) \omega_{K,j}, \qquad \text{(A.24)}
\end{aligned}$$

where we have defined $\widehat{\varphi}_{j\tau} \equiv -\log \varphi_{j\tau}$ and where in the last equality, we have defined

$$F\left(\widehat{\varphi}_{j\tau}\right) = \frac{\widehat{\varphi}_{j\tau}^{1+\frac{1-\lambda}{\theta}} e^{-\widehat{\varphi}_{j\tau}}}{\gamma\left(\widehat{\varphi}_{j\tau}; s\right)} = \frac{\widehat{\varphi}_{j\tau} \gamma'\left(\widehat{\varphi}_{j\tau}; s\right)}{\gamma\left(\widehat{\varphi}_{j\tau}; s\right)}, \qquad s \equiv 1 + \frac{1-\lambda}{\theta}. \qquad \text{(A.25)}$$

Noting that in this case $\Omega_{K,j} = \omega_{K,j}$leads to the desired result.

Next, we consider the case without marginal costs. Since we have $\omega_{K,j\tau} = 0$ and $\omega_{L,j\tau} = 0$ without marginal costs, we haveFrom Equation (A.19), we now have

$$\begin{aligned}
\frac{\partial \log \ell_{j\tau}}{\partial \log r} &= \lambda \frac{\partial \log c_j}{\partial \log r} - \varphi_{j\tau} \frac{\left(-\log \varphi_{j\tau}\right)^{\frac{1-\lambda}{\theta}}}{\gamma\left(-\log \varphi_{j\tau}\right)} \frac{\partial \log \varphi_{j\tau}}{\partial \log r}, \\
&= \lambda \frac{\partial \log c_j}{\partial \log r} - \theta\, \varphi_{j\tau} \frac{\left(-\log \varphi_{j\tau}\right)^{\frac{1-\lambda}{\theta}}}{\gamma\left(-\log \varphi_{j\tau}\right)} \log \varphi_{j\tau} \left(1 + \tfrac{\lambda}{1-\lambda}\left(1 - \tfrac{\partial \log c_j}{\partial \log r}\right)\right), \\
&= \lambda\, \varepsilon_{K,j} \omega_{f,j} + \theta \frac{\widehat{\varphi}_{j\tau}^{1+\frac{1-\lambda}{\theta}} e^{-\widehat{\varphi}_{j\tau}}}{\gamma\left(\widehat{\varphi}_{j\tau}; 1+\frac{1-\lambda}{\theta}\right)} \left(1 + \tfrac{\lambda}{1-\lambda}\left(1 - \varepsilon_{K,j}\right)\right), \\
&= \lambda\, \varepsilon_{K,j} \omega_{f,j} + \theta\, F\left(\widehat{\varphi}_{j\tau}\right) \left(1 + \tfrac{\lambda}{1-\lambda}\left(1 - \varepsilon_{K,j} \omega_{f,j}\right)\right).
\end{aligned}$$

In this case, we have $\Omega_{K,j} = \frac{\omega_{f,j}}{1+\omega_{f,j}} = \varepsilon_j \omega_{f,j}$, we can write

$$\sigma_{j\tau}^{s,K} = \lambda \frac{\varepsilon_{K,j}}{\varepsilon_j} + \theta\, F\left(\widehat{\varphi}_{j\tau}\right) \left(\frac{1}{1-\lambda} \frac{1}{\varepsilon_j \omega_{f,j}} - \frac{\lambda}{1-\lambda} \frac{\varepsilon_{K,j}}{\varepsilon_j}\right),$$

leading to the desired result.

Finally, we show that $F(\cdot)$ is a positive valued and a monotonically decreasing function of $\widehat{\varphi}_{j\tau}$, that is $F'(\cdot) \leq 0$ everywhere. First, note that we have

$$\lim_{x\to 0^+} F(x) = \lim_{x\to 0^+} \frac{e^{-x}x^s + s\,e^{-x}x^{s-1}}{e^{-x}x^{s-1}} = s,$$
$$\lim_{x\to\infty} F(x) = 0.$$

Next, we find that

$$\begin{aligned} F'(x) &= \frac{-e^{-x}x^s + s\,e^{-x}x^{s-1} - e^{-x}x^s\gamma'(x;s)/\gamma(x;s)}{\gamma(x;s)}, \\ &= \frac{\gamma'(x;s)}{\gamma(x;s)}\left(-x + s - \frac{x^s e^{-x}}{\gamma(x;s)}\right). \end{aligned}$$

The sign of $F'(x)$ is the negative of the sign of $G(x) = x^s e^{-x} - (s-x)\,\gamma(x;s)$. Let

$$G'(x) = sx^{s-1}e^{-x} - x^s e^{-x} - s\,x^{s-1}e^{-x} + x^s e^{-x} + \gamma(x;s) = \gamma(x;s) \geq 0.$$

Therefore, we conclude that the derivative in Equation (A.24) is an increasing function of the automation share $\varphi_{j\tau}$. □

Proof for Proposition 4. To perform the aggregation, we need to compute the elasticity of firm scale y_j with respect to factor prices $(r, \boldsymbol{w})$ and the scale of aggregate output Y. First, we compute these elasticity for a firm j in industry i by fixing industry-level scale Y_i. To this end, we first define an industry-level measure of the scale elasticity of marginal cost $\bar{\varepsilon}_i^m$ following the definition (36). Then, applying Lemma 3 of Lashkari et al. (2024) (page 1807), we have the following expressions for this elasticity

$$\left.\frac{\partial \log y_j}{\partial \log r}\right|_{\boldsymbol{w},Y_i} = -\eta\,\frac{\omega_{K,j}^m - \overline{\omega}_{K,i}^m}{1+\eta\,\varepsilon_j^m}, \tag{A.26}$$
$$\left.\frac{\partial \log y_j}{\partial \log w_s}\right|_{r,\boldsymbol{w}_{-s},Y_i} = -\eta\,\frac{\omega_{s,j}^m - \overline{\omega}_{s,i}^m}{1+\eta\,\varepsilon_j^m},$$
$$\left.\frac{\partial \log y_j}{\partial \log Y_i}\right|_{r,\boldsymbol{w}} = \frac{1+\eta\,\bar{\varepsilon}_i^m}{1+\eta\,\varepsilon_j^m}, \tag{A.27}$$

where we have followed the notation of Lashkari et al. (2024) by defining firm-level mea-

sures of factor prices elasticities of marginal costs

$$\omega^m_{K,j} \equiv \frac{\partial \log c_j}{\partial \log r} = (\omega_{K,j} + \omega_{f,j})\ \varepsilon_{K,j}, \qquad \omega^m_{s,j} \equiv \frac{\partial \log c_j}{\partial \log w_s} = \omega_{s,j}\,\varepsilon_{s,j},$$

using also Equations (A.17) and (A.18), and where the corresponding industry-level elasticities are defined using

$$\overline{\omega}^m_{K,i} \equiv \mathbb{E}^p_{j|i}\left[\omega^m_{K,j}\right], \qquad \overline{\omega}^m_{s,i} \equiv \mathbb{E}^p_{j|i}\left[\omega^m_{s,j}\right],$$

where $\mathbb{E}^p_{j|i}[\cdot]$ denotes the expectation with respect to the distribution of pass-through across firms within industry, given by $\Psi^p_{j|i} \equiv \Psi_{j|i}\frac{1+\eta\,\overline{\varepsilon}^m_i}{1+\eta\,\varepsilon^m_j}$.

Applying the above result again at the industry, level, we now find

$$\begin{aligned}
\left.\frac{\partial \log Y_i}{\partial \log r}\right|_{\boldsymbol{w},Y} &= -\rho\,\frac{\overline{\omega}^m_{K,i} - \overline{\omega}^m_K}{1+\rho\,\overline{\varepsilon}^m_i}, \\
\left.\frac{\partial \log Y_i}{\partial \log w_s}\right|_{r,\boldsymbol{w}_{-s},Y} &= -\rho\,\frac{\overline{\omega}^m_{s,i} - \overline{\omega}^m_s}{1+\rho\,\overline{\varepsilon}^m_i}, \\
\left.\frac{\partial \log Y_i}{\partial \log Y}\right|_{r,\boldsymbol{w}} &= \frac{1+\rho\,\overline{\varepsilon}^m}{1+\rho\,\overline{\varepsilon}^m_i},
\end{aligned} \tag{A.28}$$

where we have used the definition in Equation (37) for $\overline{\varepsilon}^m$ and where

$$\overline{\omega}^m_K \equiv \mathbb{E}^p_i\left[\overline{\omega}^m_{K,i}\right], \qquad \overline{\omega}^m_s \equiv \mathbb{E}^p_{j|i}\left[\overline{\omega}^m_{s,i}\right],$$

where $\mathbb{E}^p_{j|i}[\cdot]$ denotes the expectation with respect to the distribution of pass-through across firms within industry, given by $\Psi^p_i \equiv \Psi_i\frac{1+\rho\,\overline{\varepsilon}^m_i}{1+\rho\,\overline{\varepsilon}^m}$. We then have

Now, we compute the aggregate elasticity of substitution between skill groups and machines:

$$\begin{aligned}
\sigma^{s,K} = \frac{1}{\Omega_K}\left.\frac{\partial \log \sum_j \ell_{s,j}}{\partial \log r}\right|_{\boldsymbol{w},Y} &= \frac{1}{\Omega_K}\sum_{i,j}\frac{\ell_{s,j}}{L_s}\left(\frac{\partial \log \ell_{s,j}}{\partial \log r} + \frac{\partial \log \ell_{s,j}}{\partial \log y_j}\frac{\partial \log y_j}{\partial \log r}\right), \\
&= \frac{1}{\Omega_K}\sum_{i,j}\Psi_i\Psi_{j,i}\frac{\omega_{s,j}}{\overline{\omega}_s}\left(\Omega_{K,j}\sigma^{s,K}_j + \varepsilon_{s,j}\frac{\partial \log y_j}{\partial \log r}\right), \\
&= \frac{1}{\Omega_K}\sum_{i,j}\Psi_i\Psi_{j,i}\frac{\omega_{s,j}}{\overline{\omega}_s}\left(\Omega_{K,j}\sigma^{s,K}_j + \varepsilon_{s,j}\left.\frac{\partial \log y_j}{\partial \log r}\right|_{\boldsymbol{w},Y_i}\right)
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{1}{\Omega_K} \sum_{i,j} \Psi_i \Psi_{j,i} \frac{\omega_{s,j}}{\overline{\omega}_s} \left(\Omega_{K,j} \sigma_j^{s,K} + \varepsilon_{s,j} \frac{\partial \log y_j}{\partial \log Y_i} \bigg|_{r,\boldsymbol{w}} \frac{\partial \log Y_i}{\partial \log r} \bigg|_{\boldsymbol{w},Y} \right) \\
&= \frac{1}{\Omega_K} \sum_{i,j} \Psi_i \Psi_{j,i} \frac{\omega_{s,j}}{\overline{\omega}_s} \Omega_{K,j} \sigma_j^{s,K} \\
&\quad - \eta \frac{1}{\Omega_K} \sum_i \Psi_i \sum_{j \in \mathcal{J}_i} \Psi_{j,i} \frac{\omega_{s,j}}{\overline{\omega}_s} \varepsilon_{s,j} \frac{\omega_{K,j}^m - \overline{\omega}_{K,i}^m}{1 + \eta\, \varepsilon_j^m} \\
&\quad - \rho \frac{1}{\Omega_K} \sum_i \Psi_i \sum_{j \in \mathcal{J}_i} \Psi_{j,i} \frac{\omega_{s,j}}{\overline{\omega}_s} \varepsilon_{s,j} \frac{1 + \eta\, \bar{\varepsilon}_i^m}{1 + \eta\, \varepsilon_j^m} \frac{\overline{\omega}_{K,i}^m - \overline{\omega}_K^m}{1 + \rho\, \bar{\varepsilon}_i^m}, \\
&= \mathbb{E}_j \left[\frac{\omega_{s,j}}{\overline{\omega}_s} \frac{\Omega_{K,j}}{\Omega_K} \sigma_j^{s,K} \right] \\
&- \frac{\eta}{\Omega_K \overline{\omega}_s} \mathbb{E}_i \left[\frac{1}{1 + \eta\, \bar{\varepsilon}_i^m} \mathbb{E}_{j|i} \left[\frac{1 + \eta\, \bar{\varepsilon}_i^m}{1 + \eta\, \varepsilon_j^m} \left(\omega_{s,j}^m \left(\omega_{K,j}^m - \mathbb{E}_{j|i}^p \left[\omega_{K,j}^m \right] \right) \right) \right] \right] \\
&- \frac{\rho}{\Omega_K \overline{\omega}_s} \frac{1}{1 + \rho\, \bar{\varepsilon}^m} \mathbb{E}_i \left[\frac{1 + \rho\, \bar{\varepsilon}^m}{1 + \rho\, \bar{\varepsilon}_i^m} \left(\overline{\omega}_{s,i}^m \left(\overline{\omega}_{K,i}^m - \mathbb{E}_i^p \left[\overline{\omega}_{K,i}^m \right] \right) \right) \right], \\
&= \mathbb{E}_j \left[\frac{\bar{\varepsilon}}{\varepsilon_j} \frac{\Omega_{s,j}}{\Omega_s} \frac{\Omega_{K,j}}{\Omega_K} \sigma_j^{s,K} \right] \\
&- \frac{\eta}{\Omega_K \overline{\omega}_s} \mathbb{E}_i \left[\frac{1}{1 + \eta\, \bar{\varepsilon}_i^m} \mathbb{E}_{j|i}^p \left[\omega_{s,j}^m \left(\omega_{K,j}^m - \mathbb{E}_{j|i}^p \left[\omega_{K,j}^m \right] \right) \right] \right] \\
&- \frac{\rho}{\Omega_K \overline{\omega}_s} \frac{1}{1 + \rho\, \bar{\varepsilon}^m} \mathbb{E}_i^p \left[\overline{\omega}_{s,i}^m \left(\overline{\omega}_{K,i}^m - \mathbb{E}_i^p \left[\overline{\omega}_{K,i}^m \right] \right) \right], \\
&= \mathbb{E}_j^c \left[\frac{\Omega_{s,j}}{\Omega_s} \frac{\Omega_{K,j}}{\Omega_K} \sigma_j^{s,K} \right] - \frac{\eta\, \bar{\varepsilon}}{1 + \lambda\, \bar{\varepsilon}^m} \mathbb{E}_i^p \left[\mathbb{C}_{j|i}^p \left(\frac{\omega_{s,j}^m}{\Omega_s}, \frac{\omega_{K,j}^m}{\Omega_K} \right) \right] \\
&\quad - \frac{\rho\, \bar{\varepsilon}}{1 + \lambda\, \bar{\varepsilon}^m} \mathbb{C}_i^p \left(\frac{\overline{\omega}_{s,i}^m}{\Omega_s}, \frac{\overline{\omega}_{K,i}^m}{\Omega_K} \right),
\end{aligned}
$$

$$
\begin{aligned}
\sigma^{s,K} = \frac{1}{\Omega_K} \frac{\partial \log \sum_j \ell_{s,j}}{\partial \log r} \bigg|_{\boldsymbol{w},Y} &= \frac{1}{\Omega_K} \sum_{i,j} \frac{\ell_{s,j}}{L_s} \left(\frac{\partial \log \ell_{s,j}}{\partial \log r} + \frac{\partial \log \ell_{s,j}}{\partial \log y_j} \frac{\partial \log y_j}{\partial \log r} \right), \\
&= \frac{1}{\Omega_K} \sum_{i,j} \Psi_i \Psi_{j,i} \frac{\omega_{s,j}}{\overline{\omega}_s} \left(\Omega_{K,j} \sigma_j^{s,K} + \varepsilon_{s,j} \frac{\partial \log y_j}{\partial \log r} \right), \\
&= \frac{1}{\Omega_K} \sum_{i,j} \Psi_i \Psi_{j,i} \frac{\omega_{s,j}}{\overline{\omega}_s} \left(\Omega_{K,j} \sigma_j^{s,K} + \varepsilon_{s,j} \frac{\partial \log y_j}{\partial \log r} \bigg|_{\boldsymbol{w},Y_i} \right) \\
&\quad + \frac{1}{\Omega_K} \sum_{i,j} \Psi_i \Psi_{j,i} \frac{\omega_{s,j}}{\overline{\omega}_s} \left(\Omega_{K,j} \sigma_j^{s,K} + \varepsilon_{s,j} \frac{\partial \log y_j}{\partial \log Y_i} \bigg|_{r,\boldsymbol{w}} \frac{\partial \log Y_i}{\partial \log r} \bigg|_{\boldsymbol{w},Y} \right) \\
&= \frac{1}{\Omega_K} \sum_{i,j} \Psi_i \Psi_{j,i} \frac{\omega_{s,j}}{\overline{\omega}_s} \Omega_{K,j} \sigma_j^{s,K}
\end{aligned}
$$

$$
\begin{aligned}
&- \eta \frac{1}{\Omega_K} \sum_i \Psi_i \sum_{j \in \mathcal{J}_i} \Psi_{j,i} \frac{\omega_{s,j}}{\overline{\omega}_s} \varepsilon_{s,j} \varepsilon_{K,j} \frac{\omega^m_{K,j} - \overline{\omega}^m_{K,i}}{1 + \eta\, \varepsilon^m_j} \\
&- \rho \frac{1}{\Omega_K} \sum_i \Psi_i \sum_{j \in \mathcal{J}_i} \Psi_{j,i} \frac{\omega_{s,j}}{\overline{\omega}_s} \varepsilon_{s,j} \frac{1 + \eta\, \bar{\varepsilon}^m_i}{1 + \eta\, \varepsilon^m_j} \bar{\varepsilon}_{K,i} \frac{\overline{\omega}^m_{K,i} - \overline{\omega}^m_K}{1 + \rho\, \bar{\varepsilon}^m_i}, \\
= \mathbb{E}_j &\left[\frac{\omega_{s,j}}{\overline{\omega}_s} \frac{\Omega_{K,j}}{\Omega_K} \sigma^{s,K}_j \right] \\
- \tfrac{\eta}{\Omega_K \overline{\omega}_s} \mathbb{E}_i &\left[\tfrac{1}{1+\lambda\, \bar{\varepsilon}^m_i} \mathbb{E}_{j|i} \left[\tfrac{1+\lambda\, \bar{\varepsilon}^m_i}{1+\lambda\, \bar{\varepsilon}^m_j} \left(\frac{\omega^m_{s,j}}{\overline{\omega}_s} \left(\omega^m_{K,j} - \mathbb{E}^p_{j|i} \left[\omega^m_{K,j} \right] \right) \right) \right] \right] \\
- \tfrac{\rho}{\Omega_K \overline{\omega}_s} \tfrac{1}{1+\lambda\, \bar{\varepsilon}^m} \mathbb{E}_i &\left[\tfrac{1+\lambda\, \bar{\varepsilon}^m}{1+\lambda\, \bar{\varepsilon}^m_i} \left(\overline{\omega}^m_{s,i} \left(\overline{\omega}^m_{K,i} - \mathbb{E}^p_i \left[\overline{\omega}^m_{K,i} \right] \right) \right) \right], \\
= \mathbb{E}_j &\left[\frac{\bar{\varepsilon}}{\varepsilon_j} \frac{\Omega_{s,j}}{\Omega_s} \frac{\Omega_{K,j}}{\Omega_K} \sigma^{s,K}_j \right] \\
- \tfrac{\eta}{\Omega_K \overline{\omega}_s} \mathbb{E}_i &\left[\tfrac{1}{1+\lambda\, \bar{\varepsilon}^m_i} \mathbb{E}^p_{j|i} \left[\left(\omega^m_{s,j} \left(\omega^m_{K,j} - \mathbb{E}^p_{j|i} \left[\omega^m_{K,j} \right] \right) \right) \right] \right] \\
- \tfrac{\rho}{\Omega_K \overline{\omega}_s} \tfrac{1}{1+\lambda\, \bar{\varepsilon}^m} \mathbb{E}^p_i &\left[\left(\overline{\omega}^m_{s,i} \left(\overline{\omega}^m_{K,i} - \mathbb{E}^p_i \left[\overline{\omega}^m_{K,i} \right] \right) \right) \right], \\
= \mathbb{E}^c_j &\left[\frac{\Omega_{s,j}}{\Omega_s} \frac{\Omega_{K,j}}{\Omega_K} \sigma^{s,K}_j \right] - \mathbb{E}_i \left[\tfrac{\eta \bar{\varepsilon}}{1+\lambda\, \bar{\varepsilon}^m_i} \mathbb{C}^p_{j|i} \left(\tfrac{\omega^m_{s,j}}{\Omega_s}, \tfrac{\omega^m_{K,j}}{\Omega_K} \right) \right] \\
&- \tfrac{\rho\, \bar{\varepsilon}}{1+\eta\, \bar{\varepsilon}^m} \mathbb{C}^p_i \left(\tfrac{\overline{\omega}^m_{s,i}}{\Omega_s}, \tfrac{\overline{\omega}^m_{K,i}}{\Omega_K} \right),
\end{aligned}
$$

where in the second equality, we have substituted for the scale elasticity of skill group s and its elasticity of substitution with machines, in the third equality, we have used the fact that $\omega_{s,j} \varepsilon_{s,j} = \omega^m_{s,j}$ and that Equations (A.26), (A.27), and (A.28) together imply

$$
\begin{aligned}
\left. \frac{\partial \log y_j}{\partial \log r} \right|_{\boldsymbol{w}, Y} &= \left. \frac{\partial \log y_j}{\partial \log r} \right|_{\boldsymbol{w}, Y_i} + \left. \frac{\partial \log y_j}{\partial \log Y_i} \right|_{r, \boldsymbol{w}} \left. \frac{\partial \log Y_i}{\partial \log r} \right|_{\boldsymbol{w}, Y}, \\
&= -\eta\, \frac{\omega^m_{K,j} - \overline{\omega}^m_{K,i}}{1 + \eta\, \varepsilon^m_j} - \rho \frac{1 + \lambda\, \bar{\varepsilon}^m_i}{1 + \eta\, \varepsilon^m_j} \frac{\overline{\omega}^m_{K,i} - \overline{\omega}^m_K}{1 + \rho\, \bar{\varepsilon}^m_i}.
\end{aligned}
$$

Following similar steps, we can compute the aggregate elasticity of substitution between skill groups:

$$
\begin{aligned}
\sigma^{s,s'} = \frac{1}{\Omega_{s'}} \frac{\partial \log \sum_j \ell_{s,j}}{\partial \log w_{s'}} &= \frac{1}{\Omega_{s'}} \sum_{i,j} \frac{\ell_{s,j}}{L_s} \left(\frac{\partial \log \ell_{s,j}}{\partial \log w_{s'}} + \frac{\partial \log \ell_{s,j}}{\partial \log y_j} \frac{\partial \log y_j}{\partial \log w_{s'}} \right), \\
&= \frac{1}{\Omega_{s'}} \sum_{i,j} \Psi_i \Psi_{j,i} \frac{\omega_{s,j}}{\Omega_s} \left(\Omega_{s',j} \sigma^{s,s'}_j + \varepsilon_{s,j} \frac{\partial \log y_j}{\partial \log w_{s'}} \right), \\
&= \frac{1}{\Omega_{s'}} \sum_{i,j} \Psi_i \Psi_{j,i} \frac{\omega_{s,j}}{\Omega_s} \Omega_{s',j} \sigma^{s,s'}_j
\end{aligned}
$$

$$
\begin{aligned}
&- \eta \frac{1}{\Omega_{s'}} \sum_i \Psi_i \sum_{j \in \mathcal{J}_i} \Psi_{j,i} \frac{\omega_{s,j}}{\Omega_s} \varepsilon_{s,j} \varepsilon_{K,j} \frac{\omega^m_{s',j} - \overline{\omega}^m_{s',i}}{1 + \eta\, \varepsilon^m_j} \\
&- \rho \frac{1}{\Omega_{s'}} \sum_i \Psi_i \sum_{j \in \mathcal{J}_i} \Psi_{j,i} \frac{\omega_{s,j}}{\Omega_s} \varepsilon_{s,j} \frac{1 + \lambda\, \bar{\varepsilon}^m_i}{1 + \eta\, \varepsilon^m_j} \frac{\overline{\omega}^m_{s',i} - \overline{\omega}^m_{s'}}{1 + \eta\, \bar{\varepsilon}^m_i}, \\
= \mathbb{E}^c_j \left[\frac{\Omega_{s,j}}{\Omega_s} \frac{\Omega_{s',j}}{\Omega_{s'}} \sigma^{s,s'}_j \right] &- \mathbb{E}_i \left[\frac{\eta\, \bar{\varepsilon}}{1 + \eta\, \bar{\varepsilon}^m} \mathbb{C}^p_{j|i} \left(\frac{\omega^m_{s,j}}{\Omega_s}, \frac{\omega^m_{s',j}}{\Omega_{s'}} \right) \right] \\
&- \frac{\rho\, \bar{\varepsilon}}{1 + \rho\, \bar{\varepsilon}^m} \mathbb{C}^p_i \left(\frac{\overline{\omega}^m_{s,i}}{\Omega_s}, \frac{\overline{\omega}^m_{s',i}}{\Omega_{s'}} \right).
\end{aligned}
$$

Finally, we can compute the scale elasticity of skill demand at the industry

$$
\begin{aligned}
\bar{\varepsilon}_{s,i} = \frac{\partial \log \sum_{j \in \mathcal{J}_i} \ell_{s,j}}{\partial \log Y_i} &= \sum_{j \in \mathcal{J}_i} \frac{\ell_{s,j}}{L_{s,i}} \left(\left. \frac{\partial \log \ell_{s,j}}{\partial \log y_j} \right|_{r,\boldsymbol{w}} \times \left. \frac{\partial \log y_j}{\partial \log Y_i} \right|_{r,\boldsymbol{w}} \right), \\
&= \sum_j \Psi_{j,i} \frac{\omega_{s,j}}{\overline{\omega}_{s,i}} \varepsilon_{s,j} \frac{1 + \eta\, \bar{\varepsilon}^m_i}{1 + \eta\, \varepsilon^m_j}, \\
&= \frac{1}{\overline{\omega}_{s,i}} \mathbb{E}^p_{j|i} \left[\overline{\omega}^m_s \right] = \frac{\overline{\omega}^m_{s,i}}{\overline{\omega}_{s,i}},
\end{aligned}
$$

and at the aggregate levels:

$$
\begin{aligned}
\bar{\varepsilon}_s = \frac{\partial \log \sum_j \ell_{s,j}}{\partial \log Y} &= \sum_{i,j} \frac{\ell_{s,j}}{L_s} \left(\left. \frac{\partial \log \ell_{s,j}}{\partial \log y_j} \right|_{r,\boldsymbol{w}} \times \left. \frac{\partial \log y_j}{\partial \log Y_i} \right|_{r,\boldsymbol{w}} \times \left. \frac{\partial \log Y_i}{\partial \log Y} \right|_{r,\boldsymbol{w}} \right), \\
&= \sum_{i,j} \Psi_i \Psi_{j,i} \frac{\omega_{s,j}}{\overline{\omega}_s} \varepsilon_{s,j} \frac{1 + \eta\, \bar{\varepsilon}^m_i}{1 + \eta\, \varepsilon^m_j} \frac{1 + \rho\, \bar{\varepsilon}^m}{1 + \rho\, \bar{\varepsilon}^m_i}, \\
&= \frac{1}{\overline{\omega}_s} \sum_i \Psi_i\, \mathbb{E}_{j|i} \left[\frac{1 + \eta\, \bar{\varepsilon}^m_i}{1 + \eta\, \varepsilon^m_j} \omega^m_{s,j} \right] \frac{1 + \rho\, \bar{\varepsilon}^m}{1 + \rho\, \bar{\varepsilon}^m_i}, \\
&= \frac{1}{\overline{\omega}_s} \mathbb{E}_i \left[\frac{1 + \rho\, \bar{\varepsilon}^m}{1 + \rho\, \bar{\varepsilon}^m_i} \overline{\omega}^m_s \right] = \frac{\overline{\omega}^m_s}{\overline{\omega}_s}.
\end{aligned}
$$

□

Proof. Noting that firm-level labor demand is given by $\ell_{s,j} = \sum_\tau \ell_{s,j\tau}$, and that firm-and-task-type level input demand is given by $y_{j\tau} = \left(c_{j\tau} / c_j \right)^{-\lambda} y_j$, we find

$$
\sigma^{s,K}_j = \frac{1}{\omega_{K,j}} \left. \frac{\partial \log \ell_{s,j}}{\partial \log r} \right|_{y_j},
$$

$$
\begin{aligned}
&= \frac{1}{\omega_{K,j}} \sum_{\tau} \frac{\ell_{s,j\tau}}{\ell_{s,j}} \left(\frac{\partial \log \ell_{s,j\tau}}{\partial \log r}\bigg|_{y_{j\tau}} + \frac{\partial \log y_{j\tau}}{\partial \log r} \right), \\
&= \frac{1}{\omega_{K,j}} \sum_{\tau} \frac{\ell_{s,j\tau}}{\ell_{s,j}} \left[\omega_{K,j\tau}\, \sigma_{j\tau}^{L,K} - \lambda \left(\frac{\partial \log c_{j\tau}}{\partial \log r} - \frac{\partial \log c_{j}}{\partial \log r} \right) \right], \\
&= \frac{1}{\omega_{K,j}} \sum_{\tau} \frac{\ell_{s,j\tau}}{\ell_{s,j}} \left[\omega_{K,j\tau}\, \sigma_{j\tau}^{L,K} - \lambda \left(\omega_{K,j\tau} - \omega_{K,j} \right) \right], \\
&= \sum_{\tau} \frac{c_{j\tau} y_{j\tau}}{c_j y_j} \frac{\omega_{s,j\tau}}{\omega_{s,j}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \sigma_{j\tau}^{L,K} - \lambda \sum_{\tau} \frac{c_{j\tau} y_{j\tau}}{c_j y_j} \frac{\omega_{s,j\tau}}{\omega_{s,j}} \left(\frac{\omega_{K,j\tau}}{\omega_{K,j}} - 1 \right), \\
&= \sum_{\tau} \xi_{i\tau} \left(\frac{c_{j\tau}}{c_j} \right)^{1-\lambda} \frac{\omega_{s,j\tau}}{\omega_{s,j}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \sigma_{j\tau}^{L,K} - \lambda \sum_{\tau} \xi_{i\tau} \left(\frac{c_{j\tau}}{c_j} \right)^{1-\lambda} \frac{\omega_{s,j\tau}}{\omega_{s,j}} \left(\frac{\omega_{K,j\tau}}{\omega_{K,j}} - 1 \right), \\
&= \mathbb{E}_{\tau|j} \left[\frac{\omega_{s,j\tau}}{\omega_{s,j}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \right] \lambda + \mathbb{E}_{\tau|j} \left[\frac{\omega_{s,j\tau}}{\omega_{s,j}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \widehat{\lambda}_{j\tau} \right] - \mathbb{C}_{\tau|j} \left(\frac{\omega_{s,j\tau}}{\omega_{s,j}}, \frac{\omega_{K,j\tau}}{\omega_{K,j}} \right) \lambda, \\
&= \lambda + \mathbb{E}_{\tau|j} \left[\frac{\omega_{s,j\tau}}{\omega_{s,j}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \widehat{\lambda}_{j\tau} \right].
\end{aligned}
$$

where $\mathbb{E}_{\tau|j}[\cdot]$ and $\mathbb{C}_{\tau|j}(\cdot,\cdot)$ are the expectation and covariance operators for the distributions of variable costs $\frac{C_{j\tau}^{v}}{C_{j}^{v}} = \frac{c_{j\tau} y_{j\tau}}{c_j y_j}$across tasks at the firm level. Applying similar steps, we can derive the elasticity of substitution between labor and machines

$$
\begin{aligned}
\sigma_{j}^{L,K} &= \frac{1}{\omega_{K,j}} \frac{\partial \log \ell_j}{\partial \log r}\bigg|_{y_j}, \\
&= \frac{1}{\omega_{K,j}} \sum_{\tau} \frac{\ell_{j\tau}}{\ell_j} \left(\frac{\partial \log \ell_{j\tau}}{\partial \log r}\bigg|_{y_{j\tau}} + \frac{\partial \log y_{j\tau}}{\partial \log r} \right), \\
&= \frac{1}{\omega_{K,j}} \sum_{\tau} \frac{\ell_{j\tau}}{\ell_j} \left[\omega_{K,j\tau}\, \sigma_{j\tau}^{L,K} - \lambda \left(\frac{\partial \log c_{j\tau}}{\partial \log r} - \frac{\partial \log c_{j}}{\partial \log r} \right) \right], \\
&= \frac{1}{\omega_{K,j}} \sum_{\tau} \frac{\ell_{j\tau}}{\ell_j} \left[\omega_{K,j\tau}\, \sigma_{j\tau}^{L,K} - \lambda \left(\omega_{K,j\tau} - \omega_{K,j} \right) \right], \\
&= \sum_{\tau} \frac{c_{j\tau} y_{j\tau}}{c_j y_j} \frac{\omega_{L,j\tau}}{\omega_{L,j}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \sigma_{j\tau}^{L,K} - \lambda \sum_{\tau} \frac{c_{j\tau} y_{j\tau}}{c_j y_j} \frac{\omega_{s,j\tau}}{\omega_{s,j}} \left(\frac{\omega_{K,j\tau}}{\omega_{K,j}} - 1 \right), \\
&= \mathbb{E}_{\tau|j} \left[\frac{\omega_{L,j\tau}}{\omega_{L,j}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \right] \lambda + \mathbb{E}_{\tau|j} \left[\frac{\omega_{L,j\tau}}{\omega_{L,j}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \widehat{\lambda}_{j\tau} \right] - \mathbb{C}_{\tau|j} \left(\frac{\omega_{L,j\tau}}{\omega_{L,j}}, \frac{\omega_{K,j\tau}}{\omega_{K,j}} \right) \lambda, \\
&= \lambda + \mathbb{E}_{\tau|j} \left[\frac{\omega_{L,j\tau}}{\omega_{L,j}} \frac{\omega_{K,j\tau}}{\omega_{K,j}} \widehat{\lambda}_{j\tau} \right].
\end{aligned}
$$

The elasticity of substitution between different skill groups follow similarly:

$$\begin{aligned}
\sigma_j^{s,s'} &= \frac{1}{\omega_{s',j}} \frac{\partial \log \ell_{s,j}}{\partial \log w_{s'}}\bigg|_{y_j}, \\
&= \frac{1}{\omega_{s',j}} \sum_\tau \frac{\ell_{s,j\tau}}{\ell_{s,j}} \left(\frac{\partial \log \ell_{s,j\tau}}{\partial \log w_{s'}}\bigg|_{y_{j\tau}} + \frac{\partial \log y_{j\tau}}{\partial \log w_{s'}} \right), \\
&= \frac{1}{\omega_{s',j}} \sum_\tau \frac{\ell_{s,j\tau}}{\ell_{s,j}} \left[\omega_{s',j\tau}\, \sigma_{j\tau}^{s,s'} - \lambda \left(\frac{\partial \log c_{j\tau}}{\partial \log w_{s'}} - \frac{\partial \log c_j}{\partial \log w_{s'}} \right) \right], \\
&= \sum_\tau \frac{\ell_{s,j\tau}}{\ell_{s,j}} \left[\frac{\omega_{s',j\tau}}{\omega_{s',j}} \frac{1}{\omega_{L,j\tau}} \left(1 + \theta - \omega_{K,j\tau} \left(\lambda + \widehat{\lambda}_{j\tau} \right) \right) - \lambda \left(\omega_{s',j\tau} - \omega_{s',j} \right) \right], \\
&= \mathbb{E}_{\tau|j} \left[\frac{\omega_{s,j\tau}}{\omega_{s,j}} \frac{\omega_{s',j\tau}}{\omega_{s',j}} \frac{\omega_{L,j}}{\omega_{L,j\tau}} \right] \frac{1+\theta}{\omega_{L,j}} - \mathbb{E}_{\tau|j} \left[\frac{\omega_{s,j\tau}}{\omega_{s,j}} \frac{\omega_{s',j\tau}}{\omega_{s',j}} \frac{\omega_{K,j}}{\omega_{L,j\tau}} \left(\lambda + \widehat{\lambda}_{j\tau} \right) \right] \\
&\quad - \mathbb{C}_{\tau|j} \left(\frac{\omega_{s,j\tau}}{\omega_{s,j}}, \frac{\omega_{s',j\tau}}{\omega_{s',j}} \right) \lambda.
\end{aligned}$$

□

Lemma 2. First, let $\widetilde{c}_\tau(\nu) \equiv \min_s c_{s\tau}(\nu)$ and note that if the following condition is satisfied

$$\widetilde{c}_\tau(\nu) \leq \frac{r}{z_{k\tau}},$$

there is no profit to be made by producing task-variety τ in any industry i.

Each firm j is distinguished by a productivity a_j, a cost $c_{j\tau} = c_{i\tau}(a_j)$ for an industry-task-type-specific and monotonically decreasing function of productivity $c_{i\tau}(\cdot)$, and a self-automation threshold $c_j^* = c_{i\tau}^*(a_j)$ for an industry-task-type-specific and monotonically decreasing function of productivity $c_{i\tau}^*(\cdot)$. Since outsourcing the product of task-type-and-variety (τ, ν) only involves paying the unit cost $q_{i\tau}(\nu)$, the firm treats the cost of outsourcing the task just like producing it with labor. Thus, the self-automation threshold of the firm $c_{i\tau}^*(a_j)$ is still given by

$$c_{i\tau}^*(a) \equiv \left(1 + (1-\lambda) \frac{z_{k\tau}/z_{k\tau}^f}{y_{j\tau}^k(a)} \right)^{\frac{1}{1-\lambda}} \frac{r}{z_{k\tau}}, \tag{A.29}$$

$$y_{i\tau}^k(a) \equiv \left(\frac{z_k}{r} \right)^\lambda \frac{\gamma_i}{J_i} \xi_{i\tau}\, \overline{c}_i^\eta\, c_i(a)^{-(\eta-\lambda)}\, a^{\eta-1}. \tag{A.30}$$

where $c_i(a)$ is the overall marginal cost of production of output for a firm with productivity a in industry i. If the platform price $q_{i\tau}(\nu)$ is below the lowest-cost labor $\widetilde{c}_{i\tau}(\nu)$ and then the demand by the downstream firm from the upstream TaS provider demand is nonzero and is given by

$$y^m_{j\tau}(\nu) = \xi_{i\tau}\, q_{i\tau}(\nu)^{-\lambda}\; c_j^{-(\eta-\lambda)}\; a_j^{\eta}\; \tfrac{Y_i}{J_i}\overline{c}_i^{\eta}, \qquad q_{i\tau}(\nu) \leq c_j^* = c^*_{i\tau}\left(a_j\right), \widetilde{c}_{i\tau}(\nu). \tag{A.31}$$

where $c_j = c_i\left(a_j\right)$.

Let $g_i(a)\,\Delta a$ denote the share of all J_i firms in industry i with productivity $a \pm \frac{1}{2}\Delta a$ and $a^*_{i\tau}(q)$ denote the highest level of productivity such that $q \leq c^*_{i\tau}(a)$. Then, aggregating across all firms, the total demand of all TaS providers in task variety (τ, ν) if the price is q from all downstream firms is given by

$$\begin{aligned} Y^m_{i\tau}(q) &= \xi_{i\tau}\, q^{-\lambda}\; \left(\tfrac{Y_i}{J_i}\overline{c}_i^{\eta}\right) \left(\sum_0^{a^*_{i\tau}(q)} c_{i\tau}(a)^{-(\eta-\lambda)}\; a^{\eta}\, g_i(a)\; \Delta a\right), \\ &\approx \xi_{i\tau}\, q^{-\lambda}\; \left(\tfrac{Y_i}{J_i}\overline{c}_i^{\eta}\right) \left(\int_0^{a^*_{i\tau}(q)} c_{i\tau}(a)^{-(\eta-\lambda)}\; a^{\eta}\, g_i(a)\; da\right), \end{aligned}$$

where $a^*_{i\tau}(q)$ is found by solving $q = c^*_{i\tau}(a)$, and where the second line provides an approximation given the fact that the number of firms in each industry J_i is very large. Since $a^*_{i\tau}(\cdot)$ is a monotonically decreasing function, this demand function is monotonic in q. Let the inverse demand function be denoted by $q^m_{i\tau}(Y)$.

Now let us consider the firm's problem. Each TaS provider is going to maximize

$$\Pi^m_{i\tau} \equiv \max_y \left[q^m_{i\tau}\left(y + (n_{i\tau} - 1)\,\overline{y}_{i\tau}\right) - \frac{r}{z_{k\tau}}\right] y - \frac{r}{z^f_{k\tau}} - f^m_e,$$

where we have already incorporated the assumption of a symmetric Nash equilibrium with $n_{i\tau}$ firms where the firm under question takes the choices $\overline{y}_{i\tau}$ of other firms as given. Then, we have

$$\begin{aligned} 0 &= q^m_{i\tau} + y\; \left(q^m_{i\tau}\right)' - \frac{r}{z_{k\tau}}, \\ &= q^m_{i\tau}\left(1 + \tfrac{1}{n_{i\tau}} \frac{Y^m_{i\tau}\left(q^m_{i\tau}\right)'}{q^m_{i\tau}}\right) - \frac{r}{z_{k\tau}}, \end{aligned}$$

where in the second equality we have used the fact that $y/Y^m_{i\tau} = n_{i\tau}$ in the symmetric

equilibrium. which leads to:

$$q_{i\tau}^{m} = \frac{\lambda_{i\tau}^{m}}{\lambda_{i\tau}^{m} - 1/n_{i\tau}} \frac{r}{z_{k\tau}},$$

where we have defined the industry-level demand elasticity

$$\lambda_{i\tau}^{m} = -\frac{q\left(Y_{i\tau}^{m}\right)'(q)}{Y_{i\tau}^{m}}.$$

□

Proof for Lemma 3. The Lagrangian for the compute minimization problem in Equation 51 is

$$\mathscr{L} = 6\, m\, d + \lambda \left[\psi_m\, m^{-\alpha_m} + \psi_d\, d^{-\alpha_d} + \underline{h} - h\right].$$

The first order conditions are:

$$\frac{\partial \mathscr{L}}{\partial m} = 6d - \alpha_m\, \lambda \psi_m\, m^{-\alpha_m} \frac{1}{m} = 0 \tag{A.32}$$

$$\frac{\partial \mathscr{L}}{\partial d} = 6m - \alpha_d\, \lambda \psi_d\, d^{-\alpha_d} \frac{1}{d} = 0 \tag{A.33}$$

Rearrange and take the ratio of Equations A.32 and A.33:

$$\alpha_m\, \lambda \psi_m\, m^{-\alpha_m} = \alpha_d\, \lambda \psi_d\, d^{-\alpha_d},$$

implying

$$\frac{6\, m\, d}{\lambda} = \alpha_d\, \lambda \psi_d\, d^{-\alpha_d} = \alpha_m\, \lambda \psi_m\, m^{-\alpha_m}.$$

Plugging this relationship into the scaling law and rearranging for m yields a power-law relationship between optimal model m, data size m, and loss h:

$$m = \left(\frac{1}{\psi_m} \frac{\alpha_d}{\alpha_d + \alpha_m}\right)^{-\frac{1}{\alpha_m}} (h - \underline{h})^{-\frac{1}{\alpha_m}},$$
$$d = \left(\frac{1}{\psi_d} \frac{\alpha_m}{\alpha_d + \alpha_m}\right)^{-\frac{1}{\alpha_d}} (h - \underline{h})^{-\frac{1}{\alpha_d}}.$$

Multiplying these together yields:

$$\frac{1}{z_{k\tau}^{f}} = 6\, m\, d = 6 \left(\frac{1}{\psi_m} \frac{\alpha_m}{\alpha_d + \alpha_m}\right)^{-\frac{1}{\alpha_d}} \left(\frac{1}{\psi_d} \frac{\alpha_d}{\alpha_d + \alpha_m}\right)^{-\frac{1}{\alpha_m}} (h - \underline{h})^{-\left(\frac{1}{\alpha_m} + \frac{1}{\alpha_d}\right)}.$$

□

B Appendix for Uncovering the Patterns of Productive Advantage

B.1 Task Survey

The survey utilized Prolific, a widely recognized online platform, to recruit participants, targeting U.S. adults aged 18 and older who had at least two years of work experience and familiarity with occupations involving computer vision tasks. Survey implementation was conducted through Qualtrics. The final sample exhibited significant demographic diversity, with respondents ranging in age from 18 to 85 years, and a median age of 35 years. Gender representation included 44.34 percent female, 41.21 percent male, and the remainder identifying as non-binary, other genders, or preferring not to disclose their gender. The sample also reflected variation in income levels, with an average annual household income of approximately $94,000, 8.44 percent of respondents reporting household incomes below $25,000, and 9.26 percent reporting incomes above $150,000.

The first stage of the survey required participants to identify their industry, occupation, and relevant computer vision tasks. If a participant indicated familiarity with a given task, they were directed to a series of detailed questions; otherwise, they were prompted to move to the next available task. To mitigate cognitive fatigue and response deterioration, participants had the option to discontinue participation at any point. Across the 263 occupations included in the survey, the distribution of tasks was uneven, with 153 occupations being associated with at least two computer vision tasks, yielding an average of 1.75 tasks per occupation.

To assess task accuracy requirements, the survey incorporated a series of questions on error tolerance. Respondents provided estimates of the typical error rate observed under normal working conditions for competent workers, the worst acceptable error rate beyond which a worker would no longer be considered qualified, and the expected error rate in a scenario where an individual was forced to make decisions at random. These measures provide a structured framework for evaluating the precision requirements of different tasks and the implications of automation for maintaining acceptable levels of performance.

To ensure the reliability and validity of the collected data, the survey incorporated several quality control mechanisms. Attention checks were embedded to detect inconsistent

or illogical responses. For example, responses were flagged for removal if a participant reported a typical worker error rate that was higher than the maximum acceptable error rate. Additionally, screening mechanisms were implemented to verify that respondents were sufficiently familiar with the occupation and tasks under evaluation. Before each task, participants were required to confirm their familiarity, and those who failed this screening were excluded from further analysis. Response time monitoring was also used to detect rushed or careless submissions, with responses falling below a predetermined duration threshold being removed.

The survey was conducted in two waves, following an iterative design process to improve data clarity and accuracy. An initial trial survey (Wave 1) was conducted in July 2023, collecting 451 valid responses. The findings from this preliminary survey informed revisions to the survey design, including rewording ambiguous questions, expanding response options to capture more nuanced perspectives, and refining instructional materials. The second wave (Wave 2), conducted in November 2023, yielded 3,327 responses and incorporated these refinements. To maintain data integrity, no respondents participated in both waves. A post-survey analysis confirmed that responses from the two waves could be seamlessly combined after appropriate adjustments. The survey handbook documents the specific modifications implemented between the trial and formal surveys.

B.2 Concordance of Occupations and Industries

The CPS permits respondents to specify employment industry at the four-digit NAICS code level. We use our estimates of labor supply and occupation-level hourly wages to calculate the wage-bill proportion of occupations in each industry. We aggregate task characteristics to the industry-level through

$$\gamma_{i\tau} = \sum_{o} \pi_{io}\gamma_{o\tau}, \text{ where } \pi_{io} = \frac{w_o \ell_{io}}{\sum_o w_o \ell_{io}},$$

where π_{io} corresponds to the proportion of industry i's wage-bill attributed to occupation o, w_o are occupation-level wage averages, and ℓ_{io} are labor hours of occupation o in industry i. Industry-task information include minimum task accuracy requirements and task complexity as well as task importance in the O*NET database.

B.3 AI Scaling Law Experiments

Parameter	$\beta^{\psi}_{m,0}$	$\beta^{\psi}_{m,1}$	$\beta^{\alpha}_{m,0}$	$\beta^{\alpha}_{m,1}$	$\beta^{\psi}_{d,0}$	$\beta^{\psi}_{d,1}$	$\beta^{\alpha}_{d,0}$	$\beta^{\alpha}_{d,1}$	$\underline{h}$	δ
Value	4.93	0.0049	−0.0001	0.0003	−1.11	0.73	0.0034	0.04	−139.08	339.15

Table B.1: Estimated Parameter Values for the AI Fine-Tuning Scaling Laws (CV Experiment)

Notes: Values rounded to four decimal places where appropriate. Parameters were estimated using [estimation method] with data from the AI scaling law experiment.

B.3.1 Effective Data Size

The specification we take to the data is modified relative to the one discussed in the main text in terms of the treatment of the data size. The value of the data size that we use in the specification of the scaling law in Equation (50) is specified in terms of *effective data per class*, which is a concept introduced by Muennighoff et al. (2023) to allow for the potential repeated data use during fine-tuning. The effective data in a given setting s of our experiment is defined as

$$\widehat{d^s} = \frac{d^s}{\chi^s}\left(1 + \delta\left(1 - e^{-(epoch^s - 1)/\delta}\right)\right), \tag{B.1}$$

where δ is a discount factor of each repeated use of the data, *epoch* represents the number of complete passes the same data is fed to the iterative fine-tuning process, and, as before, χ stands for task complexity. Intuitively, the n-th pass of using the same data in fine-tuning reduces its effectiveness by a factor $(1 - 1/\delta)^n$. We additionally estimate the value of the discount factor δ in our analysis.

To better understand the interpretation of each variable, consider a practical example where a computer vision model is implemented for automated defect detection in a manufacturing plant. The model categorizes components into five distinct classes ($\chi = 5$) based on different defect types. Suppose a dataset of 750 images is collected ($d = 750$), meaning that, on average, each defect/non-defect classification category contains 150 images ($\frac{d}{\chi} = 150$). Now, assuming the training process involves four epochs, the dataset is reused four times to iteratively update the model parameters. According to Equation B.1, the effective data per class, adjusted for repeated utilization during training, is computed as: $\widehat{d} = 598$.

B.4 Scaling Laws Estimation Results

Table B.1 presents the estimated coefficients of the scaling law using nonlinear least squares (NLS) based on our experiment. To better interpret the results, we computed the coeffi-

Number of Classes	ψ_m	ψ_d	α_m	α_d	$\underline{h}$
2	139.05	0.54	0.0001	0.03	-139.08
5	139.68	1.06	0.0004	0.07	-139.08
10	140.15	1.75	0.0006	0.09	-139.08
50	141.26	5.62	0.0010	0.16	-139.08
100	141.74	9.29	0.0012	0.19	-139.08
500	142.86	29.91	0.0017	0.25	-139.08
1000	143.35	49.47	0.0019	0.28	-139.08

Table B.2: Table of estimated scaling law parameters

cients ψ_m, ψ_d, α_m, and α_d under different task complexities χ according to Equation **??** and presented them in Table B.2, which offers empirical insights into the relationship between AI inputsâmodel size (m) and effective data per class ($\widehat{d}$)—and cross-entropy loss (h) in vision classification tasks of varying complexity. The parameters ψ_m, ψ_d, α_m, and α_d reflect the responsiveness of entropy loss reduction to changes in model size and dataset size.

Specifically, ψ_m and α_d jointly determine the rate at which cross-entropy loss declines as model size increases, whereas ψ_d and α_d dictate the slope of cross-entropy loss reduction with respect to data size expansion. The estimated values of α_m range from 1.08×10^{-4} for binary classification to 1.90×10^{-3} for a 1000-class problem. α_d varies from 3.07×10^{-2} for 2-class classification to 2.76×10^{-1} for 1000-class classification. While the value of ψ_m remains relatively stable around 140, exhibiting only a slow increase with the number of classes, ψ_d has grown significantlyâby almost 100 timesârising from 0.54 to 49.47. This increasing trend suggests that as task complexity grows, both model size and data size scaling become more influential in driving loss reductions.

Parameters α_m and α_d also govern the second-order effect of varying m and $\widehat{d}$, respectively — the larger the α_m and α_d, the greater the diminishing return to scale of the corresponding input factor. Holding the number of classes fixed, we observe that α_d is consistently an order of magnitude larger than α_m. This disparity implies that as data size increases, the marginal reduction in cross-entropy loss diminishes at a decreasing rate —reflected in a flattening slope. In contrast, the marginal loss reduction from increasing model size remains nearly constant across different initial settings, as evidenced by the persistently low values of α_m.

To better understand how different input factors would affect the cross entropy loss, Table B.3 presents an analysis of the effect of doubling the model size or data per class on the cross entropy loss under different fine-tuning configurations. Each row corresponds to an initial fine-tuning setting, characterized by a specific model size and data per class.

The cross-entropy loss (h) is reported for each setting (column 3). Notably, as we move from the first to the third row, the initial cross-entropy loss decreases, indicating improved model performance with larger model sizes and data sizes.

To assess the impact of scaling model size and data, two modifications are considered for each initial setting: (i) doubling the model size while keeping the data fixed and (ii) doubling the data while keeping the model size fixed. The results show that the effects of doubling the model size and doubling the data size are almost always comparable. However, doubling the model size has a more consistent effect — it consistently reduces cross-entropy loss by approximately 0.036 across all initial settings. In contrast, the effect of doubling the data per class is more pronounced when the initial data size is small (150 instances per class) but starts to diminish when the initial data size is large (3000 instances per class).

B.5 Mapping Error Rates to Cross-entropy Loss

Finally, Table B.4 reports the estimated parameters for the relationship between error rates and cross-entropy loss in Equation (56). The model explains over 97% of the variation in the cross-entropy values across experimental settings based on variations in error rates and problem complexity.

B.6 Inferring Task-Level Labor Productivities

B.6.1 Patterns of Skill Group Comparative Advantage and Industrial Employment

Assuming that there are no fixed costs, we find the following expression for the share of workers of skill group s in labor payments in industry i:

$$\omega_{s,Li} = \frac{w_s L_{si}}{\sum_{s'} w_{s'} L_{s'i}} = \sum_{\tau} \psi_{\tau,i} \frac{\left(Z_{s\tau}/w_s\right)^{\theta}}{\sum_{s'} \left(Z_{s'\tau}/w_{s'}\right)^{\theta}},$$

Initial Fine-Tuning Settings			**Cross-Entropy Loss (h)**		**Loss Reduction (Δh)**	
Model Size	*Data per Class*	h	*Model Size* $\times 2$	*Data* $\times 2$	*Model Size* $\times 2$	*Data* $\times 2$
2M	150	0.527	0.491	0.496	0.036	0.031
2M	3000	0.402	0.366	0.377	0.036	0.026
200M	150	0.289	0.253	0.258	0.036	0.031

Table B.3: Cross-entropy loss under different settings (problem complexity = 5 classes).

Parameter	β_1^h	β_2^h	β_3^h	β_0^h	β_4^h	β_5^h
Estimate	0.28	0.21	0.13	0.51	-1.75	0.36

Table B.4: Estimated Parameters of the Mapping from Error to Cross-Entropy Loss

Notes: Estimates of the parameter of the model mapping error rates and problem complexity to the measures of cross-entropy loss based on Equation (56). Values rounded to two decimal places where appropriate. Parameters were estimated using nonlinear least squares with data from the AI scaling law experiment.

$$= \sum_{\tau} \xi_{i\tau} \left(\frac{c_{i\tau}}{c_i} \right)^{1-\lambda} \frac{\left(Z_{s\tau} / w_s \right)^{\theta}}{\sum_{s'} \left(Z_{s'\tau} / w_{s'} \right)^{\theta}},$$
$$= \sum_{\tau} \xi_{i\tau} \left(\frac{c_{i\tau}}{c_i} \right)^{1-\lambda} \left(\frac{Z_{s\tau} \widetilde{c}_{\tau}}{w_s} \right)^{\theta},$$

where, since all firms j in the same industry have the same expression for the marginal costs we have let $c_{j\tau} \equiv c_{i\tau}$ and $c_j \equiv c_i$ for all firms j in industry i. In the case in which the automation costs are prohibitively high, such that $\varphi_{j\tau} \equiv \varphi_{i\tau} \approx 0$ for all firms in all tasks and industries, we have$c_{i\tau} = \widetilde{c}_{\tau}$ and therefore find

$$\frac{w_s L_{si}}{\sum_{s'} w_{s'} L_{s'i}} = \sum_{\tau} \xi_{i\tau} \left(\frac{\widetilde{c}_{\tau}}{c_i} \right)^{1-\lambda} \left(\frac{Z_{s\tau} \widetilde{c}_{\tau}}{w_s} \right)^{\theta} \propto w_s^{-\theta} \sum_{\tau} \xi_{i\tau} \widetilde{c}_{\tau}^{\theta+1-\lambda} Z_{s\tau}^{\theta}.$$

This also implies that the share of task τ in the time of skill s in industry i is given by

$$\psi_{\tau,is} \equiv \frac{\ell_{s,i\tau}}{\sum_{\tau'} \ell_{s,i\tau'}} = \frac{\psi_{\tau,i}\, \omega_{s,L\tau}}{\sum_{\tau'} \psi_{\tau',i}\, \omega_{s,L\tau'}} = \frac{\xi_{i\tau} \widetilde{c}_{\tau}^{1+\theta-\lambda}\, Z_{s\tau}^{\theta}}{\sum_{\tau} \xi_{i\tau} \widetilde{c}_{\tau}^{1+\theta-\lambda}\, Z_{s\tau}^{\theta}}.$$

Next, let us write the expression for the relative demand in industry i for two skill groups s and s' as

$$\frac{L_{is}}{L_{is'}} = \left(\frac{w_s}{w_{s'}} \right)^{-(1+\theta)} \frac{\sum_{\tau} \xi_{i\tau} \widetilde{c}_{\tau}^{1+\theta-\lambda}\, Z_{s\tau}^{\theta}}{\sum_{\tau'} \xi_{i\tau'} \widetilde{c}_{\tau'}^{1+\theta-\lambda}\, Z_{s'\tau}^{\theta}}, \tag{B.2}$$

and the relative demand of skill group s across two industries i and i' as

$$\frac{L_{is}}{L_{i's}} = \frac{\sum_{\tau} \xi_{i\tau} \widetilde{c}_{\tau}^{1+\theta-\lambda}\, Z_{s\tau}^{\theta}}{\sum_{\tau'} \xi_{i'\tau'} \widetilde{c}_{\tau'}^{1+\theta-\lambda}\, Z_{s\tau'}^{\theta}}.$$

B.6.2 Estimation of the Patterns of Comparative Advantage of Skill Groups

Let us next take some base skill group b and task type o as the unit productivity $Z_{bo} \equiv 1$ and state all other productivities relative to this one and let

$$\widetilde{Z}_{s\tau} \equiv \frac{Z_{s\tau}}{Z_{bo}}.$$

Assume a known value for the parameter θ. Our data allows us to measure relative employment of different skill groups within industry, relative wages, and the time (cost) share of each task for each industry-skill pair. We consider the following relationship between observed data $\widehat{\boldsymbol{d}}$ and the unobserved matrix of skill-task comparative advantage $\widetilde{\boldsymbol{Z}}$:

$$\widehat{\ell}_{is} \equiv \log\left(\frac{L_{is}}{L_{ib}}\right) + \widehat{(1+\theta)\log\left(\frac{w_s}{w_b}\right)} = \log\left(\sum_{\tau} \widetilde{\omega}_{i\tau}\left(\widetilde{\boldsymbol{Z}}\right)\, \widetilde{Z}_{s\tau}^{\theta}\right) + v_{is},$$

where $\widetilde{\omega}_{i\tau}$ is a matrix of intensities given by Equation (B.2) and where we assume v_{is} to be an *i.i.d* normally distributed measurement error. The above model leads to the following log likelihood function

$$\log \mathbb{P}\left(\widehat{\boldsymbol{\ell}} \,|\, \widetilde{\boldsymbol{z}}\right) = -\frac{1}{2}\sum_{i,s\neq b}\left[\frac{1}{\chi}\left(\widehat{\ell}_{is} - \log\left(\sum_{\tau}\widetilde{\omega}_{i\tau}\left(\widetilde{\boldsymbol{Z}}\right)\, e^{\theta\widetilde{z}_{s\tau}}\right)\right)^2 + \log\chi\right],$$

where we have set $\widetilde{z}_{s\tau} \equiv \log \widetilde{Z}_{s\tau}$ and where χ denotes the variance of measurement errors.

We take a Bayesian approach to the estimation of the matrix $\widetilde{\boldsymbol{z}} \equiv [\widetilde{z}_{s\tau}]$. We impose a Cauchy prior with parameter γ_z on the unobserved matrix of skill-task comparative advantage $\widetilde{\boldsymbol{z}}$ and an inverse-Gamma prior over the variance χ^2 following

$$\begin{aligned}
\mathbb{P}\left(\widetilde{z}_{s\tau}\right) &= \frac{1}{\pi\gamma_z\left(1 + \left(\widetilde{z}_{s\tau}/\gamma_z\right)^2\right)}, \\
\mathbb{P}\left(\chi\right) &= \frac{\left(\beta_\chi\right)^{\alpha_\chi}}{\Gamma\left(\alpha_\chi\right)}\chi^{\alpha_\chi - 1}\exp\left(-\frac{\beta_\chi}{\chi}\right),
\end{aligned}$$

where $(\alpha_\chi, \beta_\chi)$ are the parameters of the inverse-Gamma function. This allows us to write the log posterior function as

$$\log \mathbb{P}\left(\widetilde{\boldsymbol{z}}, \chi | \,\widehat{\boldsymbol{d}}\right) = -\frac{1}{2\chi}\sum_{i,s\neq b}\left[\left(\widehat{d}_{is} - \log\left(\sum_{\tau}\widetilde{\omega}_{i\tau}\left(\widetilde{\boldsymbol{z}}\right)\, e^{\theta\widetilde{z}_{s\tau}}\right)\right)^2 + \log\chi\right] - \sum_{\tau,s\neq b}\log\left(1 + \left(\frac{\widetilde{z}_{s\tau}}{\gamma_z}\right)^2\right)$$

$$-\frac{\beta_\chi}{\chi} - (\alpha_\chi + 1)\log\chi + \text{const.}$$

We use a Hamiltonian Monte Carlo algorithm for Bayesian inference to compute the posterior above over the space of comparative advantage.

C Additional Tables and Figures

Figure C.1: Distribution of AI Costs

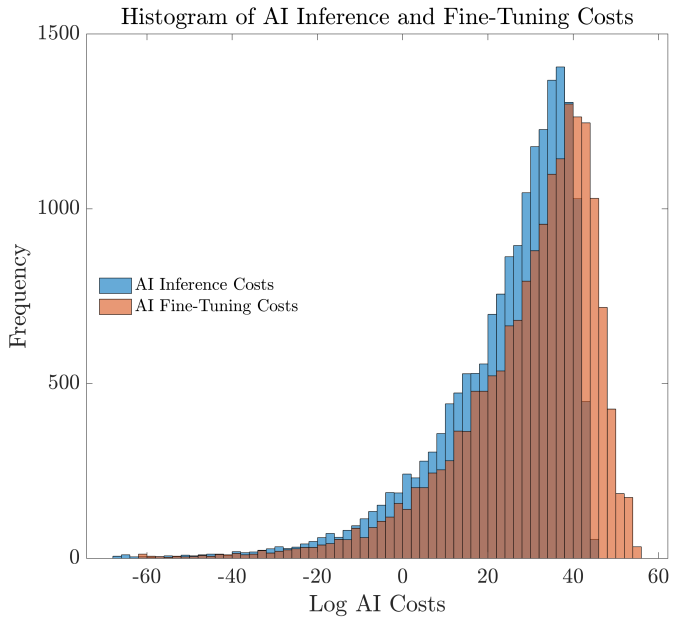
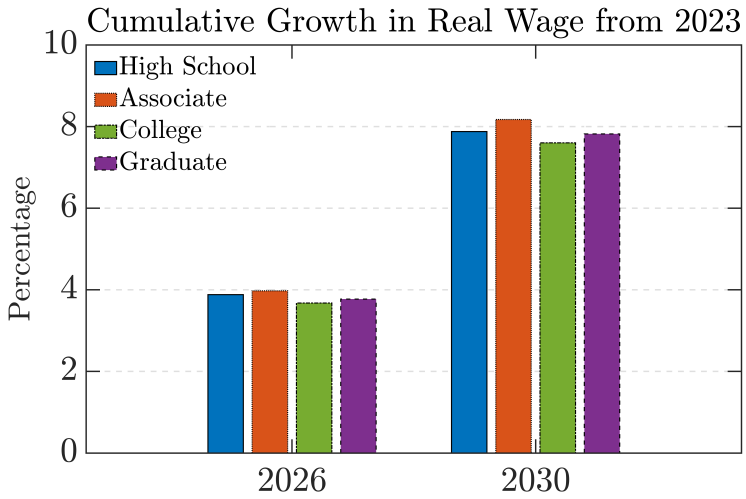


Figure C.2: Evolution of Real Wage on Automation Path



Note: The figure illustrates real wage growth relative to the baseline year of 2023. Wage growth is calculated at the projected price of compute in years 2026 and 2030.

References

Acemoglu, Daron, "The simple macroeconomics of AI," *Economic Policy*, 2025, *40* (121), 13–58. 6

_ **and David Autor**, "Skills, tasks and technologies: Implications for employment and earnings," in "Handbook of labor economics," Vol. 4, Elsevier, 2011, pp. 1043–1171. 6, 11

_ **and Pascual Restrepo**, "Artificial intelligence, automation, and work," in "The economics of artificial intelligence: An agenda," University of Chicago Press, 2018, pp. 197–236. 2

_ **and** _ , "The race between man and machine: Implications of technology for growth, factor shares, and employment," *American Economic Review*, 2018, *108* (6), 1488–1542. 6, 11

_ **and** _ , "Automation and new tasks: How technology displaces and reinstates labor," *Journal of Economic Perspectives*, 2019, *33* (2), 3–30. 6, 11

_ **and** _ , "The wrong kind of AI? Artificial intelligence and the future of labour demand," *Cambridge Journal of Regions, Economy and Society*, 2020, *13* (1), 25–35. 2

_ , **David Autor, Jonathon Hazell, and Pascual Restrepo**, "Artificial intelligence and jobs: Evidence from online vacancies," *Journal of Labor Economics*, 2022, *40* (S1), S293–S340. 7

Achiam, Josh, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat et al., "GPT-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023. 7

Altman, Sam, "Planning for AGI and beyond," 2023. 7

Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly journal of economics*, 2003, *118* (4), 1279–1333. 6

Babina, Tania, Anastassia Fedyk, Alex X He, and James Hodson, *Firm investments in artificial intelligence technologies and changes in workforce composition*, Vol. 31325, National Bureau of Economic Research, 2023. 7

Bahri, Yasaman, Ethan Dyer, Jared Kaplan, Jaehoon Lee, and Utkarsh Sharma, "Explaining neural scaling laws," *Proceedings of the National Academy of Sciences*, 2024, *121* (27), e2311878121. 3, 26

Berlingieri, Giuseppe, Filippo Boeri, Danial Lashkari, and Jonathan Vogel, "Capital-skill complementarity in firms and in the aggregate economy," Technical Report, National Bureau of Economic Research 2024. 21

Bonney, Kathryn, Cory Breaux, Cathy Buffington, Emin Dinlersoz, Lucia S Foster, Nathan Goldschlag, John C Haltiwanger, Zachary Kroff, and Keith Savage, "Tracking firm use of AI in real time: A snapshot from the Business Trends and Outlook Survey," Technical Report, National Bureau of Economic Research 2024. 5, 7, 35, 40

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell et al., "Language models are few-shot learners," *Advances in neural information processing systems*, 2020, *33*, 1877–1901. 2

Brynjolfsson, Erik and Tom Mitchell, "What can machine learning do? Workforce implications," *Science*, 2017, *358* (6370), 1530–1534. 6

_ , _ , **and Daniel Rock**, "What can machines learn and what does it mean for occupations and the economy?," in "AEA papers and proceedings," Vol. 108 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2018, pp. 43–47. 2, 3, 6

Burstein, Ariel, Eduardo Morales, and Jonathan Vogel, "Changes in Between-Group Inequality: Computers, Occupations, and International Trade," *American Economic Journal: Macroeconomics*, April 2019, *11* (2), 348–400. 32

Cazzaniga, Mauro, Ms Florence Jaumotte, Longji Li, Mr Giovanni Melina, Augustus J Panton, Carlo Pizzinelli, Emma J Rockall, and Ms Marina Mendes Tavares, *Gen-AI: Artificial intelligence and the future of work*, International Monetary Fund, 2024. 6

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs are GPTs: Labor market impact potential of LLMs," *Science*, 2024, *384* (6702), 1306–1308. 3, 6

Felten, Edward, Manav Raj, and Robert Seamans, "Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses," *Strategic Management Journal*, 2021, *42* (12), 2195–2217. 6

Felten, Edward W, Manav Raj, and Robert Seamans, "A method to link advances in artificial intelligence to occupational abilities," in "AEA Papers and proceedings," Vol. 108 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2018, pp. 54–57. 2, 3, 6

Fleming, Martin, Neil Thompson, and Wensu Li, "The Last Mile Problem in AI," Report, Brookings Institution, United States of America 2024. Retrieved from https://coilink.org/20.500.12592/vg1bd71 on 21 Mar 2025. COI: 20.500.12592/vg1bd71. 2

Frank, Morgan R, David Autor, James E Bessen, Erik Brynjolfsson, Manuel Cebrian, David J Deming, Maryann Feldman, Matthew Groh, José Lobo, Esteban Moro et al., "Toward understanding the impact of artificial intelligence on labor," *Proceedings of the National Academy of Sciences*, 2019, *116* (14), 6531–6539. 6

Frey, Carl Benedikt and Michael A Osborne, "The future of employment: How susceptible are jobs to computerisation?," *Technological forecasting and social change*, 2017, *114*, 254–280. 2, 3

Hampole, Menaka, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller, "Artificial intelligence and the labor market," Technical Report, National Bureau of Economic Research 2025. 7

Handa, Kunal, Alex Tamkin, Miles McCain, Saffron Huang, Esin Durmus, Sarah Heck, Jared Mueller, Jerry Hong, Stuart Ritchie, Tim Belonax et al., "Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations," *arXiv preprint arXiv:2503.04761*, 2025. 3, 6

Hernandez, Danny, Jared Kaplan, Tom Henighan, and Sam McCandlish, "Scaling laws for transfer," *arXiv preprint arXiv:2102.01293*, 2021. 31

Hestness, Joel, Sharan Narang, Newsha Ardalani, Gregory Diamos, Heewoo Jun, Hassan Kianinejad, Md Mostofa Ali Patwary, Yang Yang, and Yanqi Zhou, "Deep learning scaling is predictable, empirically," *arXiv preprint arXiv:1712.00409*, 2017. 3, 26

Hoffmann, Jordan, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark et al., "Training compute-optimal large language models," *arXiv preprint arXiv:2203.15556*, 2022. 5, 26, 27

Huang, Yueling, "The Labor Market Impact of Artificial Intelligence: Local vs. Aggregate Effect," 2023. Working Paper. 7

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei, "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020. 3, 26, 27

Korinek, Anton and Donghyun Suh, "Scenarios for the Transition to AGI," Technical Report, National Bureau of Economic Research 2024. 6

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, 2012, *25*. 2

Lane, Marguerita and Anne Saint-Martin, "The impact of Artificial Intelligence on the labour market: What do we know so far?," *OECD Social, Employment, and Migration Working Papers*, 2021, (256), 0_1–60. 6

Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard, "Information technology and returns to scale," *American Economic Review*, 2024, *114* (6), 1769–1815. 21, 34, A14

Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in "Proceedings of the IEEE/CVF international conference on computer vision" 2021, pp. 10012–10022. 31

McElheran, Kristina, J Frank Li, Erik Brynjolfsson, Zachary Kroff, Emin Dinlersoz, Lucia Foster, and Nikolas Zolas, "AI adoption in America: Who, what, and where," *Journal of Economics & Management Strategy*, 2024, *33* (2), 375–415. 7

Muennighoff, Niklas, Alexander Rush, Boaz Barak, Teven Le Scao, Nouamane Tazi, Aleksandra Piktus, Sampo Pyysalo, Thomas Wolf, and Colin A Raffel, "Scaling data-constrained language models," *Advances in Neural Information Processing Systems*, 2023, *36*, 50358–50376. A25

Rosenfeld, Jonathan S, "Scaling laws for deep learning," *arXiv preprint arXiv:2108.07686*, 2021. 3, 26

Rosenfeld, Jonathan S., Amir Rosenfeld, Yonatan Belinkov, and Nir Shavit, "A Constructive Prediction of the Generalization Error Across Scales," 2019. 5

Svanberg, Maja, Wensu Li, Martin Fleming, Brian Goehring, and Neil Thompson, "Beyond AI exposure: which tasks are cost-effective to automate with computer vision?," *Available at SSRN 4700751*, 2024. 6, 7

Thompson, Neil C, Kristjan Greenewald, Keeheon Lee, Gabriel F Manso et al., "The computational limits of deep learning," *arXiv preprint arXiv:2007.05558*, 2020, *10*. 2

Thompson, Neil, Martin Fleming, Benny J Tang, Anna M Pastwa, Nicholas Borge, Brian C Goehring, and Subhro Das, "A model for estimating the economic costs of computer vision systems that use deep learning," in "Proceedings of the AAAI Conference on Artificial Intelligence," Vol. 38 2024, pp. 23012–23018. 5

Webb, Michael, "The impact of artificial intelligence on the labor market," *Available at SSRN 3482150*, 2019. 3

Zeira, Joseph, "Workers, machines, and economic growth," *The Quarterly Journal of Economics*, 1998, *113* (4), 1091–1117. 6

Zhai, Xiaohua, Alexander Kolesnikov, Neil Houlsby, and Lucas Beyer, "Scaling vision transformers," in "Proceedings of the IEEE/CVF conference on computer vision and pattern recognition" 2022, pp. 12104–12113. 3, 26

Zhang, Biao, Zhongtao Liu, Colin Cherry, and Orhan Firat, "When scaling meets LLM fine-tuning: The effect of data, model and fine-tuning method," *arXiv preprint arXiv:2402.17193*, 2024. 31