AI, Scale, and Task-Based Theories of Automation[*]

Danial Lashkari
FRB NY & MIT FutureTech

Wensu Li
MIT FutureTech

Christina Qiu
University of Miami

Neil Thompson
MIT FutureTech

July 13, 2026

Please click here for the most recent draft.

Abstract

AI automation exemplifies technologies that entail task-level fixed costs, e.g., model training or fine-tuning. We integrate such fixed costs into a general theory of automation and show that AI's comparative advantage becomes *scale-dependent* at the task level. We derive the elasticities characterizing the resulting production function, including the endogenous labor–AI substitution elasticity. We apply a quantitative version of the model to computer vision automation, disciplining fixed (training) and marginal (inference) costs using AI scaling laws, estimated from a fine-tuning experiment. Calibrated to 2023 U.S. data and recent estimates of the decline in compute prices, the model projects rapid automation, reaching 23% of firms (60% of employment) by 2035. The labor share traces a U-shape as the substitution elasticity declines from above to below one. Scale accounts for the majority (∼63%) of cross-task variation in AI comparative advantage.

1 Introduction

The recent surge in the capabilities of Artificial Intelligence (AI) systems has turned the labor-market consequences of automation into an urgent question for economic research.

[*]We thank Chad Jones, Pascual Restrepo, Lawrence Schmidt, Chris Tonetti, and many seminar audiences. Won-Jae Chang, Kayla Krupa, and Canaan Cui provided outstanding research assistance. The views expressed do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System.

A growing body of empirical (e.g., Brynjolfsson and Mitchell, 2017; Felten et al., 2018; Frey and Osborne, 2017) and theoretical (e.g., Acemoglu and Restrepo, 2018a, 2020) work studies this question at the level of the tasks required to fulfill production. Through this lens, the relevant question is whether machines (AI) have a comparative advantage over labor in performing any given task, in the sense of offering lower production costs. In canonical task-based models, this comparative advantage is governed by relative marginal costs, which, in the case of AI, are primarily driven by the computational costs of running the model to perform the task (i.e., the costs of inference).

Yet this standard approach abstracts from the defining role of *scale*, both in driving progress in AI performance and in deploying AI in production. AI scaling laws establish a robust empirical relationship between AI performance and the computational inputs used to train models (Hoffmann et al., 2022; Kaplan et al., 2020).[1] Since the corresponding training costs are fixed with respect to the volume of inference, a larger scale of AI output justifies developing a more capable model. This logic also applies at the task level since models require fine-tuning to acquire capabilities specific to a given production task (Fleming et al., 2024). Consider two tasks that are both exposed to automation by AI: detecting tumors in radiological images, which has so far resisted automation, and monitoring a retailer's in-store cameras.[2] Automating either task requires its own model, fine-tuned on data collected and labeled from that setting. Whether such a task is worth automating then turns not only on whether machines are cheaper than labor at the margin, but also on whether the task is performed at enough scale to spread this fixed cost. Production scale may thus be a source of AI comparative advantage across tasks, alongside the marginal-cost comparison that standard theories emphasize.

Motivated by this observation, we provide a general theory of task-based automation, in which automating a task involves, in addition to variable (marginal) costs, a fixed cost of adopting machines. In this environment, automation is governed by *scale-dependent comparative advantage*, the product of two terms: the standard marginal advantage, i.e., the marginal productivity of machines relative to labor, and a novel *scale advantage*, an increasing function of the ratio of variable costs to fixed costs and, therefore, of the task scale. As task output grows or fixed costs vanish, scale advantage approaches unity and

[1] Over the past decade, the computational resources dedicated to training the most capable AI foundation models have grown dramatically (Thompson et al., 2020). AlexNet, an early success 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 decade later, ViT-22B, a large vision transformer, had computational needs exceeding those of AlexNet by roughly six orders of magnitude: it featured 22 billion parameters and was trained on a dataset of around four billion images (Dehghani et al., 2023).

[2] See CBS News (2019).

we recover the standard model (Autor et al., 2003; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018b, 2019). This mechanism is not specific to AI: any automation technology involving task-level fixed costs generates the same scale dependence, though AI offers a leading example given the central role of data and training in its cost structure.

We develop a representation for this task-based production function that allows for a dual characterization of the resulting cost and demand functions. Relying on this representation, we further provide a quantitative version of the model that directly maps to the production tasks in the Occupational Information Network (O*NET) classification. Embedding this model of automation in a general-equilibrium, multi-industry framework featuring firm heterogeneity, we derive the implications of declining compute prices for output, wages, and the labor share. We show that the elasticity of substitution between labor and machines evolves endogenously (and often non-monotonically) as cheaper compute shifts the frontier of automation across tasks.

A central challenge in quantifying any theory of AI automation is obtaining credible estimates of AI's comparative advantage across tasks, especially because early adoption patterns may not reflect AI's actual underlying potential. We use AI scaling laws to address this challenge: if we can estimate the level of AI performance needed to automate a given task, scaling laws predict the model size and training-data size required to achieve it. The implied computational inputs at the training stage also provide estimates of AI's marginal productivity, as inference costs scale with model size. Scaling laws, therefore, permit us to fully characterize both fixed and marginal components of AI productivity provided we find measures of required performance at the task level.

We focus on computer vision because it is a domain where the performance requirements for automation can be readily mapped to measurable task characteristics. Computer vision models perform visual classification and detection tasks, where the complexity of the problem (i.e., the number of potential outcomes) and its error tolerance (i.e., acceptable error rate) define the performance requirements that enter computer vision scaling laws (Zhai et al., 2022). Detecting tumors in radiology and monitoring retailers' in-store cameras offer two examples of such visual tasks. Following this logic, we construct LLM-generated measures of error tolerance and complexity for all vision-related O*NET tasks in the US economy.[3] Estimating computer vision scaling laws in a fine-tuning experiment then allows us to infer AI comparative advantage for these tasks and calibrate the quantitative general equilibrium model.

The calibrated model confirms our emphasis on scale: scale advantage accounts for

[3] The radiologist's task has a complexity of 2280 potential outcomes and an error rate tolerance of 0.4%. By contrast, the retailer's task has a complexity of 5 potential outcomes and an error rate tolerance of 20.6%.

roughly 63% of the cross-task variation in AI's comparative advantage in 2023. As compute prices fall, adoption follows an S-curve, rising from under 1% of firms in 2023 to 23% by 2035 (60% in employment-weighted terms as larger firms automate first). Aggregate output rises by about 7% by 2075, and real wages increase throughout. The labor share, however, traces a U-shape, driven by the endogenous evolution of the aggregate elasticity of substitution: labor and AI begin as gross substitutes, but as automation deepens, this elasticity falls below one, so that they become gross complements and the labor share recovers. Moreover, the sectoral pattern of automation re-orders substantially over time, confirming that today's adoption reveals little about where automation will ultimately concentrate.

The rest of the paper proceeds as follows. In Section 2, we extend a canonical task-based environment with a continuum of production tasks (Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018b, 2019) by allowing machines to carry task-level fixed costs. We characterize the automation decisions in terms of the scale-dependent comparative advantage of machines. Bridging the general model to its quantitative version, we introduce a task-type representation that groups tasks with common machine productivities and frequencies, while allowing labor productivity to vary within each type. We use this representation to derive analytical expressions for key production function elasticities. Crucially, we decompose the labor–AI elasticity of substitution into three components: an intensive margin of task substitution, holding the set of automated tasks fixed; and two extensive margins of new task automation, operating through scale and marginal advantage. Finally, embedding this firm-level production function in our macro model, we follow Lashkari et al. (2024) to characterize the aggregate elasticities and the general-equilibrium response to falling compute prices.

In Section 3, we estimate scaling laws and use measured task characteristics to infer fixed and marginal AI productivities at the task level. To do so, we first conduct a controlled fine-tuning experiment on computer vision foundation models that systematically varies model size, training-data size, and the complexity of image-classification tasks. This experiment identifies how the fixed training cost and marginal inference cost of AI change with model performance. It also provides the empirical mapping from task-level complexity and error tolerance into the performance level required to automate each task. Second, we use a frontier large language model (LLM) to identify the O*NET task items with substantial vision content and to construct the remaining task characteristics needed for quantification, including frequency and time use. We validate these LLM-generated measures against worker-survey responses and map the resulting 659 vision tasks to industries.

Section 4 uses these estimates of AI's cost structure to quantify the model. We calibrate the remaining parameters using standard values from the literature, together with data on firm-size distributions and on firms' adoption of computer vision AI in 2023 (Bonney et al., 2024). The calibrated model underlies the results described above, which we further illustrate through the automation path of a single firm, where adoption arrives in waves across groups of tasks with similar complexity and error tolerance. We also show similar results when the fixed cost of fine-tuning is incurred at the industry rather than the firm level, and under an alternative scenario featuring a faster compute-cost decline that incorporates algorithmic progress.

Prior Work Our paper contributes to task-based theories of automation (Zeira, 1998; Autor et al., 2003; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018b, 2019) in three ways.[4] First, we incorporate a central feature of AI automation: task-level fixed costs that arise when a general-purpose foundation model must be specialized to a particular production task, for instance, through fine-tuning. Second, for this general model, we characterize the dual cost function and the full set of production elasticities it implies, including the labor–AI elasticity of substitution and the output elasticities of factor demand. We show that these objects are scale-dependent in the presence of fixed costs and evolve endogenously as machine prices fall. Third, we specialize the general theory to a quantitative model that directly maps to task classifications such as O*NET. The mapping draws on tools commonly used in quantitative trade. Viewed through this lens, our framework introduces fixed costs and scale dependence into canonical models of offshoring (Grossman and Rossi-Hansberg, 2008) and comparative advantage (Dornbusch et al., 1977), with production tasks playing the role of traded commodities.

Although AI's potential for automation is widely recognized, direct evidence on its productivity effects remains largely confined to narrow experimental settings (e.g., Dell'Acqua et al., 2023; Peng et al., 2023; Cui et al., 2024; Brynjolfsson et al., 2025). A broader empirical literature instead measures task or occupational *exposure* to AI. Following early contributions such as Brynjolfsson and Mitchell (2017) and Felten et al. (2018), this literature constructs proxies for the extent to which AI can affect production across tasks and combines them with data sources such as O*NET to measure occupational exposure (see also Brynjolfsson et al., 2018; Frank et al., 2019; Felten et al., 2021; Lane and Saint-Martin, 2021;

[4] For other recent theoretical applications of this framework to the study of the macroeconomic and labor-market impacts of AI, see, for example, Korinek and Suh (2024) and Acemoglu (2025). Our quantitative approach is also closely related to Humlum (2019), who develops a quantitative task-based model of firm-level technology adoption.

Cazzaniga et al., 2024).[5] These measures capture only whether AI capabilities are relevant for a task, not whether AI can meet the task's performance requirement, the magnitude of task-level productivity gains, or the costs that govern adoption decisions. Our framework instead turns task characteristics into cost primitives and incorporates the fixed costs of adoption, tracing their implications for the allocation of production across firms and industries and, ultimately, for aggregate outcomes.

A central methodological ingredient in our quantitative exercise is the use of AI scaling laws to infer how training and inference costs vary across tasks. An alternative is to infer comparative advantage from observed adoption patterns as such data become available (e.g., McElheran et al., 2024; Bonney et al., 2024).[6] Lindenlaub et al. (2026), for example, use observed adoption to estimate AI's *marginal* advantage across occupations. However, because AI diffusion remains at an early stage, a snapshot of current adoption may not offer sufficient information about the future path of automation. By contrast, scaling-law-based approaches have demonstrated forecasting value for AI capabilities (Achiam et al., 2023). Svanberg et al. (2024) show that cost measures based on scaling laws better predict realized unemployment shifts than a wide range of alternative automation measures. We therefore rely on scaling laws both for their empirical performance and because they reflect a central organizing principle emphasized by state-of-the-art model builders.[7]

Finally, our paper relates to a recent quantitative literature that studies the labor-market consequences of AI at the task level. Focusing on the consequences for inequality between different skill groups, Freund and Mann (2026) and Althoff and Reichardt (2026) construct rich models of worker sorting across occupations and rely on LLM-generated measures of AI's impact in the tradition of Eloundou et al. (2024) to discipline their models; Hampole et al. (2025) instead construct task-level exposure measures based on online job postings and professional profiles. Our approach is complementary but deliberately more parsimonious: we keep the production side close to the canonical task-based automation framework, add task-level fixed costs of AI specialization, and discipline the resulting cost primitives with scaling laws. Our emphasis on fixed costs echoes Svanberg et al. (2024), who likewise stress that automation depends on the cost of building AI systems for a task and not only on its technical feasibility; we differ in embedding this

[5]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).

[6]These data have also been used to study regional labor-market impacts of AI (Huang, 2023). Another fruitful source of information comes from textual overlap between AI-related terminology and job vacancies or resumes (e.g., Acemoglu et al., 2022; Babina et al., 2023).

[7]In the words of founders of OpenAI: "when we first started OpenAI, we didn't expect scaling to be as important as it has turned out to be" (Altman, 2023).

insight in a structural, task-based general-equilibrium model that delivers the resulting elasticities and aggregate dynamics.

2 Theory

We begin in Section 2.1 with a general model of task-based production in which using machines requires task-level fixed costs, and we characterize the firm's automation decisions and cost structure. Section 2.2 introduces a task-type representation that preserves the generality of this model while allowing us to derive the key elasticities characterizing the production function and the factor demand system. Section 2.3 then specializes this representation for quantitative analysis, illustrates its implications in two limiting cases, and embeds firms in a general-equilibrium environment that maps directly to task-level data.

2.1 A General Model of Task-Based Production with Fixed Costs

To produce output y, the firm performs a unit continuum of tasks $\nu \in \Upsilon$ according to

$$y = a \left(\int_{\Upsilon} \xi\left(\nu\right)^{\frac{1}{\lambda}} y\left(\nu\right)^{\frac{\lambda-1}{\lambda}} d\nu \right)^{\frac{\lambda}{\lambda-1}}, \tag{1}$$

where $y\left(\nu\right)$ denotes task output, λ the elasticity of substitution across tasks, a a firm-specific Hicks-neutral productivity, and $\xi\left(\nu\right)$ an exogenous shifter for the intensity of the task in production, such that $\int_{\nu \in \Upsilon} \xi\left(\nu\right) d\nu = 1$. Importantly, we assume that tasks are gross complements in production, $\lambda < 1$, implying they constitute "weak links" (Jones, 2011; Jones and Tonetti, 2026). Task output can be performed either using labor $\ell(\nu)$ or machines $k(\nu)$ following

$$y(\nu) = z_L(\nu)\, \ell(\nu) + z_K(\nu)\, \left(k(\nu) - 1/z_K^f(\nu) \right).$$

Workers have marginal productivity $z_L(\nu)$. Machines have marginal productivity $z_K(\nu)$ and require a fixed deployment cost of $1/z_K^f(\nu)$ units of machines, where $z_K^f(\nu)$ denotes the machine's *fixed-cost productivity*, so that a higher $z_K^f(\nu)$ implies a lower fixed deployment cost.[8]

[8]Likely worker-level fixed costs of on-boarding and training scale with the number of workers and are therefore variable costs at the firm level.

Cost Minimization Problem Assume now that firms take as given the wage rate w per unit of labor and the compute price r per unit of machines. They choose whether to use machines or labor to produce output for each task ν. We write the resulting cost-minimization problem as the optimal collection of task output $\boldsymbol{y} \equiv [y(\nu)]_0^1$ and production inputs $\boldsymbol{x} \equiv [x(\nu)]_0^1$ that solve

$$C(w, r; y) = \min_{\boldsymbol{x}, \boldsymbol{y}} \int_{\nu \in \Upsilon} \left(c(\nu)\, y(\nu) + c^f(\nu) \right) d\nu, \qquad \text{subject to Equation (1)}, \tag{2}$$

such that marginal $c(\nu)$ and fixed $c^f(\nu)$ costs satisfy

$$c(\nu) = \begin{cases} \frac{w}{z_L(\nu)}, & \text{if } x(\nu) = \ell(\nu) \\ \frac{r}{z_K(\nu)}, & \text{if } x(\nu) = k(\nu) \end{cases} \quad \text{and} \quad c^f(\nu) = \begin{cases} 0, & \text{if } x(\nu) = \ell(\nu) \\ \frac{r}{z_K^f(\nu)}, & \text{if } x(\nu) = k(\nu). \end{cases} \tag{3}$$

Proposition 1 below characterizes the solution to the firm's cost minimization problem. Crucially, the firm automates a task only if machines hold *scale-dependent* comparative advantage relative to labor: standard comparative advantage in marginal productivities, multiplied by a scale-advantage factor that rises with the task's equilibrium output, as we discuss below.

Proposition 1. *Given output y, the solution to the cost-minimization problem of Equations (2) and (3) is characterized by the following conditions. The automated set Υ_K of tasks for which $x(\nu) = k(\nu)$ satisfies*

$$\Upsilon_K \equiv \left\{ \nu \in \Upsilon : CA(\nu) \geq \frac{r}{w} \right\}, \quad \text{where } CA(\nu) = MA(\nu) \times SA(\nu). \tag{4}$$

Here, the machine comparative advantage *of a task $CA(\nu)$ is the product of* marginal *and* scale advantage*, defined, respectively, as*

$$MA(\nu) \equiv \frac{z_K(\nu)}{z_L(\nu)}, \qquad \text{and} \qquad SA(\nu) \equiv \left(1 + (1-\lambda) \frac{z_K(\nu)}{z_K^f(\nu)} \times \frac{1}{y_K(\nu)} \right)^{-\frac{1}{1-\lambda}}, \tag{5}$$

where task scale $y_K(\nu) \equiv \xi(\nu) (\frac{r/z_K(\nu)}{c})^{-\lambda} \frac{y}{a}$ is a function of firm-level marginal costs c given by

$$c^{1-\lambda} = \int_{\Upsilon \setminus \Upsilon_K} \xi(\nu) \left(\frac{w}{z_L(\nu)} \right)^{1-\lambda} d\nu + \int_{\Upsilon_K} \xi(\nu) \left(\frac{r}{z_K(\nu)} \right)^{1-\lambda} d\nu. \tag{6}$$

Proof. See the proof in Appendix A.4.1 (page A5). □

The key endogenous variable determining the patterns of task automation is a scale-dependent comparative advantage variable $CA(\nu)$. As defined in Equation (5), this variable has two multiplicative components: a marginal advantage $MA(\nu)$, the ratio of machine to labor marginal productivity, and a scale advantage $SA(\nu)$, which depends on the ratio of machine marginal to fixed-cost productivity and increases monotonically with task scale $y_K(\nu)$. As task scale $y_K(\nu)$ approaches zero, both scale and comparative advantage vanish and the task cannot be automated. As $y_K(\nu)$ grows large, scale advantage approaches one, and comparative advantage reflects only marginal productivity differences across factors.

This latter limit, of course, reduces to the standard case considered in the vast literature on task-based automation (e.g., Acemoglu and Restrepo, 2018b, 2019, 2022; Jones and Tonetti, 2026). Our framework, therefore, nests this standard model as the limiting case where either fixed costs converge to zero or the task scale approaches infinity. Beyond these limiting cases, scale advantage and comparative advantage depend on task scale $y_K(\nu)$, which varies with the firm's marginal cost c. But marginal cost itself depends on the set of tasks the firm automates, as shown in Equation (6). The firm's cost structure and task-level comparative advantages are therefore jointly determined, with the solution generally characterized by a fixed point in c and comparative advantages $\boldsymbol{CA}$ across all tasks.

2.2 A Task-Type Representation for the General Model

In this section, we introduce a representation for the model of Section 2.1 that, while maintaining the model's generality, allows us to examine the properties of the resulting production function. More specifically, under this representation, we derive the key elasticities that characterize the production function and the resulting factor demand system, including the elasticity of substitution between labor and machines and the output elasticity of demand for machines.

Task Types Without loss of generality, we can partition the space of production tasks Υ into a set of different types $\Upsilon_{\boldsymbol{\tau}}$. Assume that each task type is characterized by a tuple $\boldsymbol{\tau} \equiv (z_K, z_K^f, \xi)$, which gives the values of marginal and fixed machine productivity and production intensities for all tasks $\nu \in \Upsilon_{\boldsymbol{\tau}}$. Moreover, assume the measure of the tasks within this set of task types is given by $\mu(\boldsymbol{\tau})$.

Accordingly, we can denote the conditional distribution of labor productivity for all tasks within the type $\boldsymbol{\tau}$ as $G(z|\boldsymbol{\tau})$. Define the average labor productivity in a task type $\boldsymbol{\tau}$

as

$$\overline{Z}_{L,\boldsymbol{\tau}}^{\lambda-1} \equiv \int_0^\infty z^{\lambda-1} dG(z|\boldsymbol{\tau}). \tag{7}$$

It will prove useful to define a *selection function* that captures the relative rise in average labor productivity after applying a lower cutoff z_L. This function, and its corresponding elasticity, are given by

$$G_{\lambda-1}(z_L|\boldsymbol{\tau}) \equiv \int_{z_L}^\infty \left(\frac{z}{\overline{Z}_{L,\boldsymbol{\tau}}}\right)^{\lambda-1} dG(z|\boldsymbol{\tau}) \quad \text{and} \quad \widetilde{G}_{\lambda-1}(z_L|\boldsymbol{\tau}) \equiv -\frac{z_L\, G'_{\lambda-1}(z_L|\boldsymbol{\tau})}{G_{\lambda-1}(z_L|\boldsymbol{\tau})}. \tag{8}$$

We can now rearrange Equation (4) to construct a type-specific labor productivity cutoff

$$z^*_{\boldsymbol{\tau}} \equiv \frac{z_{K,\boldsymbol{\tau}}}{r/w} \times SA_{\boldsymbol{\tau}} \tag{9}$$

where $SA_{\boldsymbol{\tau}}$ evaluates scale advantage from Equation (5) for tasks with the tuple $\boldsymbol{\tau}$. For a task ν with type $\boldsymbol{\tau}$, the firm uses labor only if productivity $z_L(\nu)$ exceeds the cutoff $z^*_{\boldsymbol{\tau}}$; otherwise, the firm automates the task. Accordingly, the share of automated varieties in a task type is $\varphi_{K,\boldsymbol{\tau}} \equiv G(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau})$ and the type-level marginal costs follow

$$c_{\boldsymbol{\tau}}^{1-\lambda} = \varphi_{K,\boldsymbol{\tau}} \left(\frac{r}{z_{K,\boldsymbol{\tau}}}\right)^{1-\lambda} + \varphi_{L,\boldsymbol{\tau}} \left(\frac{w}{\overline{Z}_{L,\boldsymbol{\tau}}}\right)^{1-\lambda}, \qquad \varphi_{L,\boldsymbol{\tau}} \equiv G_{\lambda-1}(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau}). \tag{10}$$

The corresponding type-level shares of machines and labor in variable costs are

$$\omega_{K,\boldsymbol{\tau}} \equiv \varphi_{K,\boldsymbol{\tau}} \left(\frac{r/z_{K,\boldsymbol{\tau}}}{c_{\boldsymbol{\tau}}}\right)^{1-\lambda}, \qquad \omega_{L,\boldsymbol{\tau}} \equiv \varphi_{L,\boldsymbol{\tau}} \left(\frac{w/\overline{Z}_{L,\boldsymbol{\tau}}}{c_{\boldsymbol{\tau}}}\right)^{1-\lambda}. \tag{11}$$

As Equation (10) shows, the type-level shifters of the intensity of machines $\varphi_{K,\boldsymbol{\tau}}$ and labor $\varphi_{L,\boldsymbol{\tau}}$ are determined by two different functions of the automation cutoff: the c.d.f. $G(\cdot|\boldsymbol{\tau})$ and the selection function $G_{\lambda-1}(\cdot|\boldsymbol{\tau})$, respectively. As the type-level automation cutoff $z^*_{\boldsymbol{\tau}}$ rises, the share of tasks performed by machines $\varphi_{K,\boldsymbol{\tau}}$ rises. While the share of tasks performed by labor $1-\varphi_{K,\boldsymbol{\tau}}$ correspondingly falls, the shifter $\varphi_{L,\boldsymbol{\tau}}$ falls even faster, since the automated tasks are those in which labor is least productive and hence, with $\lambda < 1$, most cost-weighted; the resulting rise in the average labor productivity of un-automated tasks (the selection margin) is reflected in the falling ratio $\varphi_{L,\boldsymbol{\tau}}/(1-\varphi_{K,\boldsymbol{\tau}})$.

Firm-Level Cost Function Next, we integrate the automation decisions across task types to the firm level based on the measure μ across types. The task-weighted share of automated tasks is

$$\overline{\varphi}_K = \int \xi_{\boldsymbol{\tau}}\, \varphi_{K,\boldsymbol{\tau}}\, d\mu(\boldsymbol{\tau}). \tag{12}$$

As shown in the next proposition, the productivity-weighted integrals of machine and labor intensities across task types,

$$\varphi_K \equiv \int \xi_{\boldsymbol{\tau}}\, z_{K,\boldsymbol{\tau}}^{\lambda-1}\, \varphi_{K,\boldsymbol{\tau}}\, d\mu(\boldsymbol{\tau}), \qquad \varphi_L \equiv \int \xi_{\boldsymbol{\tau}}\, \overline{Z}_{L,\boldsymbol{\tau}}^{\lambda-1}\, \varphi_{L,\boldsymbol{\tau}}\, d\mu(\boldsymbol{\tau}), \tag{13}$$

help characterize the firm-level cost and production functions.

Proposition 2. *The firm's total cost function takes the additively separable form* $C(w,r;y) = c\,(y/a) + c^f$*, with* c *and* c^f *given by*

$$c^{1-\lambda} = \int \xi_{\boldsymbol{\tau}}\, c_{\boldsymbol{\tau}}^{1-\lambda}\, d\mu(\boldsymbol{\tau}) = \varphi_K\, r^{1-\lambda} + \varphi_L\, w^{1-\lambda}, \quad c^f = r\, k^f = r \int \frac{\varphi_{K,\boldsymbol{\tau}}}{z_{K,\boldsymbol{\tau}}^f}\, d\mu(\boldsymbol{\tau}). \tag{14}$$

Moreover, the firm's output satisfies

$$y = a \left(\varphi_K^{\frac{1}{\lambda}} \,(k - k^f)^{\frac{\lambda-1}{\lambda}} + \varphi_L^{\frac{1}{\lambda}}\, \ell^{\frac{\lambda-1}{\lambda}} \right)^{\frac{\lambda}{\lambda-1}}, \tag{15}$$

where $k - k^f$ *is total variable machine demand and* ℓ *is total labor demand.*

Proof. The proof is found in Appendix A.4.2 (page A7). □

The expression in Equation (15) parallels prior derivations of an aggregate (translated) CES form from task-based foundations (Acemoglu and Restrepo, 2018b, 2022; Aghion et al., 2017). As emphasized in particular by Acemoglu and Restrepo (2022), this form is not a standard primal representation and *does not* imply that the firm's underlying production function is (translated) CES. This is because the intensities φ_K and φ_L and the fixed machine requirement k^f are themselves functions of factor prices and scale $(r, w; y)$ through the dependence of the automation cutoffs $z^*_{\boldsymbol{\tau}}$. In contrast, the cost function $C = c\,(y/a) + c^f$ characterized above provides a well-defined dual representation of the firm's underlying production technology.

Similarly, firm-level shares for machines and labor in variable costs are given by

$$\omega_K \equiv \frac{r\left(k-k^f\right)}{c\,(y/a)} = \mathbb{E}\left[\omega_{K,\boldsymbol{\tau}}\right] = \varphi_K \left(\frac{r}{c}\right)^{1-\lambda}, \qquad \omega_L \equiv \frac{w\,\ell}{c\,(y/a)} = \mathbb{E}\left[\omega_{L,\boldsymbol{\tau}}\right] = \varphi_L \left(\frac{w}{c}\right)^{1-\lambda}, \tag{16}$$

where k and ℓ denote the firm-level demand for machines and labor, respectively, and where the expectation $\mathbb{E}[\cdot]$ is taken over types $\boldsymbol{\tau}$ under the distribution of variable costs across task types π, defined as $d\pi(\boldsymbol{\tau}) \equiv \zeta_{\boldsymbol{\tau}}(\frac{c_{\boldsymbol{\tau}}}{c})^{1-\lambda} d\mu(\boldsymbol{\tau})$.

Elasticities of the Cost Function To better understand the properties of the firm's cost function, we examine key elasticities that characterize factor demand. First, the elasticities of total cost with respect to factor prices are

$$\Omega_K \equiv \frac{\partial \log C}{\partial \log r} = \frac{\omega_K + \omega^f}{1 + \omega^f} \geq \omega_K, \qquad \Omega_L \equiv \frac{\partial \log C}{\partial \log w} = \frac{\omega_L}{1 + \omega^f} \leq \omega_L, \tag{17}$$

which corresponds to a firm's share of machines and labor in total factor payments, and where $\omega^f \equiv \frac{c^f}{c(y/a)}$ correspond to the firm's fixed-to-variable cost ratio. With fixed costs, total cost is more elastic with respect to the machine price than in the corresponding model without fixed costs. The reason is that a reduction in the factor price of machines not only reduces marginal costs for already-automated tasks on the intensive margin, but also enables the automation of additional tasks on the extensive margin. The latter margin uses machines that require fixed costs to deploy, amplifying machine demand.

We additionally examine how firm-level costs vary with output, using the scale elasticity given by

$$\varepsilon \equiv \frac{\partial \log C}{\partial \log y} = \frac{c\,(y/a)}{C} = \frac{1}{1 + \omega^f} \leq 1, \tag{18}$$

which decreases in the firm's fixed-to-variable cost ratio ω^f. Production exhibits *increasing returns to scale* when automation involves fixed costs, and the magnitude of these returns grows when the proportion of fixed to variable costs is large. Without fixed costs, the model exhibits constant returns to scale.

Elasticities of Marginal Costs In the presence of scale-dependence, the elasticities of total costs studied above may differ from those of the marginal cost. The latter elasticities determine how changes in wages, machine prices, and output pass through to firm-level marginal costs, and ultimately, firm output prices. These prices, in turn, govern how production is reallocated across firms and industries. Proposition 3 below characterizes these elasticities and shows how they vary in the presence and absence of fixed costs.

Proposition 3. *Denote the fixed-cost-weighted-average operator as $\mathbb{E}_f[\cdot]$, under which type weights are given by $d\pi_f(\boldsymbol{\tau}) \equiv \frac{1}{\omega^f} \frac{r}{c(y/a)} \frac{\varphi_{K,\boldsymbol{\tau}}}{z^f_{K,\boldsymbol{\tau}}} d\mu(\boldsymbol{\tau})$. The elasticity of the firm's marginal cost c with respect to output y is given by*

$$\varepsilon^m \equiv \frac{\partial \log c}{\partial \log y} = -\tfrac{1}{\lambda}\Theta, \qquad \Theta \equiv \frac{\frac{\lambda}{1-\lambda}\omega^f\, \mathbb{E}_f\left[\zeta_{\boldsymbol{\tau}} \widetilde{G}\left(z^*_{\boldsymbol{\tau}} | \boldsymbol{\tau}\right)\right]}{1 + \frac{\lambda}{1-\lambda}\omega^f\, \mathbb{E}_f\left[\zeta_{\boldsymbol{\tau}} \widetilde{G}\left(z^*_{\boldsymbol{\tau}} | \boldsymbol{\tau}\right)\right]}, \tag{19}$$

where $\widetilde{G}(x|\boldsymbol{\tau}) \equiv \frac{x\,g(x|\boldsymbol{\tau})}{G(x|\boldsymbol{\tau})}$ *and* $\zeta_{\boldsymbol{\tau}} = 1 - SA_{\boldsymbol{\tau}}^{1-\lambda}$. *The elasticity of the firm's marginal cost with respect to factor prices satisfies*

$$\Omega_K^m \equiv \frac{\partial \log c}{\partial \log r} = \Theta + (1-\Theta)\left(\omega_K + \omega^f \mathbb{E}_f\left[\widetilde{G}\left(z_{\boldsymbol{\tau}}^*|\boldsymbol{\tau}\right)\right]\right), \qquad \Omega_L^m \equiv \frac{\partial \log c}{\partial \log w} = 1 - \Omega_K^m. \tag{20}$$

Proof. The proof is found in Appendix A.4.3 (page A7). □

First, the term $\zeta_{\boldsymbol{\tau}} \in [0,1]$ represents a measure of the *importance* of scale in driving firm automation decisions. At the limit when firm output y approaches infinity, or when $SA_{\boldsymbol{\tau}} \to 1$, the term $\zeta_{\boldsymbol{\tau}}$ approaches zero; at the limit when firm output y approaches zero, or $SA_{\boldsymbol{\tau}} \to 0$, this term approaches one. Second, the term $\widetilde{G}\left(z_{\boldsymbol{\tau}}^*|\boldsymbol{\tau}\right)$ represents the elasticity of the labor productivity distribution evaluated at the automation cutoff in Equation (9). Intuitively, the elasticity $\widetilde{G}\left(z_{\boldsymbol{\tau}}^*|\boldsymbol{\tau}\right)$ represents the relative marginal rise in the share of automated tasks as the comparative advantage of machines, embedded in $z_{\boldsymbol{\tau}}^*$, increases.

The scaling properties of marginal cost depend on the fixed-cost-weighted mean of the interaction of the two terms above, i.e., $\mathbb{E}_f\left[\zeta_{\boldsymbol{\tau}}\widetilde{G}\left(z_{\boldsymbol{\tau}}^*|\boldsymbol{\tau}\right)\right]$. As seen in Equation (19), the scale elasticity of marginal cost ε^m is negative and decreasing in this statistic, suggesting that scaling output lowers marginal costs. The expression for Θ shows that this form of scalability becomes stronger if scale-related considerations are binding for task types with a wide margin of tasks close to the automation cutoff. The proposition further shows that the elasticity of a firm's marginal costs with respect to the price of machines r is positive and increasing in this term.

Elasticities of Factor Demand Having characterized the cost function, we can derive factor demand using Shephard's lemma. Proposition 4 characterizes how factor demand varies with output and factor prices.

Proposition 4. *Firm-level elasticity of substitution between workers and machines is*

$$\begin{aligned}\sigma \equiv \frac{1}{\Omega_K}\frac{\partial \log \ell}{\partial \log r} = \underbrace{\lambda}_{\textit{Intensive Margin}} + \underbrace{\tfrac{1}{\Omega_K}\mathbb{E}_L\left[\widetilde{G}_{\lambda-1}\left(z_{\boldsymbol{\tau}}^*|\boldsymbol{\tau}\right)\right]}_{\textit{Extensive–Marginal Advantage Margin}} \\ + \underbrace{\tfrac{\Omega_K^m-\Omega_K}{\Omega_K}\lambda + \tfrac{1-\Omega_K^m}{\Omega_K}\,\tfrac{\lambda}{1-\lambda}\,\mathbb{E}_L\left[\zeta_{\boldsymbol{\tau}}\widetilde{G}_{\lambda-1}\left(z_{\boldsymbol{\tau}}^*|\boldsymbol{\tau}\right)\right]}_{\textit{Extensive–Scale Advantage Margin}},\end{aligned} \tag{21}$$

where the labor-cost-weighted average operator is $\mathbb{E}_L[\cdot]$, *under which type weights are given by* $d\pi_L(\boldsymbol{\tau}) \equiv \frac{1}{\omega_L}\xi_{\boldsymbol{\tau}}\varphi_{L,\boldsymbol{\tau}}\left(\frac{w/\overline{Z}_{L,\boldsymbol{\tau}}}{c}\right)^{1-\lambda}d\mu(\boldsymbol{\tau})$. *The elasticities of firm demand with respect to output* *are, respectively,*

$$\eta_K \equiv \frac{\partial \log k}{\partial \log y} = \varepsilon \frac{\Omega_K^m}{\Omega_K}, \qquad \eta_L \equiv \frac{\partial \log \ell}{\partial \log y} = \varepsilon \frac{\Omega_L^m}{\Omega_L}. \tag{22}$$

Proof. The proof is found in Appendix A.4.4 (page A10). □

Equation (21) is one of the key results of our paper. It shows that the firm's labor demand elasticity always exceeds the task elasticity of substitution λ, which accounts for the standard intensive margin of substitution in a CES production function *without* the extensive margin. Beyond this margin, a rise in the price of machines r further raises demand for labor since automation cutoffs for all task types fall due to the decline in machine comparative advantage along two dimensions.

First, a rise in machine price r brings about a decline in marginal advantage, which is present in the standard models of task-based automation. This term, captured by the second term on the right-hand side of Equation (21), is higher when workers spend more time on task types with a higher local elasticity of the selection function, as captured by the expression $\mathbb{E}_L\left[\widetilde{G}_{\lambda-1}\left(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau}\right)\right]$. This elasticity gives the percentage decline in the labor cost shifter $\varphi_{L,\boldsymbol{\tau}}$ per percentage rise in the automation cutoff $z^*_{\boldsymbol{\tau}}$, capturing how strongly the labor component of production responds to automation's extensive margin.

Second, in the presence of fixed costs, a rise in machine price r brings about an additional decline in the scale advantage of machines, since the task scale $y_K(\nu)$ also falls across all tasks. This term, on the second line of Equation (21), is higher when workers spend more time on task types with a higher importance of scale advantage *and* a higher sensitivity of the selection function to changes in automation cutoff, as captured by the expression $\mathbb{E}_L\left[\zeta_{\boldsymbol{\tau}}\widetilde{G}_{\lambda-1}\left(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau}\right)\right]$. The contribution of this term suggests that the elasticity of substitution is *higher* due to the presence of fixed costs.

The expressions in Equation (22) illustrate the tight connection between the elasticities of marginal cost and the output elasticities of factor demand. We can additionally show that $\eta_K \geq 1 \geq \eta_L$, where the equality is only satisfied in the case without fixed costs. As such, factor demand is generically scale-dependent (nonhomothetic).

2.3 The Quantitative Model

In the previous section, we studied the task-based production function in the context of the task-type representation and provided a second-order (dual) characterization of the corresponding cost function. In this section, we discuss how we can use this representation to transform the abstract model of Section 2.1 into a quantitative model that directly maps to task-level data. In particular, we associate each task-type with specific task classifications in the observed data.[9] We use this representation to parameterize heterogeneity in labor and machine comparative advantage, while preserving the logic of the baseline task-based framework. Therefore, the quantitative model bridges theory and measurement, translating the primitives that determine marginal and scale advantages into objects that can be disciplined at the task-type level.

2.3.1 Firm-Level Production

Let $\tau \in \mathcal{T}$ denote a task item in a task-level classification $\mathcal{T}$, such as the US O*NET dataset. We map this task item to a task type $\boldsymbol{\tau} \equiv (z_{K,\tau}, z^f_{K,\tau}, \xi_\tau)$ in our representation of Section 2.2, assuming the values of the corresponding tuple $\boldsymbol{\tau}$ to be constant for each task ν belonging to task item τ. With slight abuse of notation, we henceforth use both τ and $\boldsymbol{\tau}$ interchangeably. We choose a corresponding counting measure μ on the set of all task items within the task classification data $\mathcal{T}$, and further assume the following conditional distribution of labor productivity within a task item τ

$$G(z|\boldsymbol{\tau}) \equiv F_\theta\left(\tfrac{z}{Z_\tau}\right) = \exp\left(-\left(\tfrac{z}{Z_\tau}\right)^{-\theta}\right), \tag{23}$$

where $F_\theta(\cdot/Z_\tau)$ is the c.d.f. of a Fréchet distribution with a shape parameter θ and a scale parameter Z_τ.

Now, define a task-item-specific *relative* automation cutoff $\widehat{z}_\tau$ as

$$\widehat{z}_\tau \equiv \frac{z^*_{\boldsymbol{\tau}}}{Z_\tau} = \frac{w}{r}\frac{z_{K,\tau}}{Z_\tau}\, SA_\tau, \qquad\qquad SA_\tau \equiv \left(1 + (1-\lambda)\frac{z_{K,\tau}/z^f_{K,\tau}}{y_{K,\tau}}\right)^{-\frac{1}{1-\lambda}}, \tag{24}$$

where task scale is $y_{K,\tau} = \xi_\tau \left(\frac{r/z_{K,\tau}}{c}\right)^{-\lambda} \frac{y}{a}$. Using this definition, Equations (23) and (10) imply that the automation share and the labor shifter for each task item are

$$\varphi_{K,\tau} = F_\theta(\widehat{z}_\tau) = \exp\left(-\widehat{z}_\tau^{\,-\theta}\right), \qquad \varphi_{L,\tau} = \frac{1}{\Gamma\left(1+\frac{1-\lambda}{\theta}\right)}\gamma\left(1 + \tfrac{1-\lambda}{\theta}, \widehat{z}_\tau^{\,-\theta}\right), \tag{25}$$

[9] Task data with broad occupational coverage are also increasingly available outside the United States. Examples include Germany's BERUFENET, used by Dengler and Matthes (2018) to study substitution potentials of occupations, and Italy's INAPP-ISTAT occupational survey, used by Cirillo et al. (2021) to study digitalization, routineness, and employment. Harmonized international task surveys such as PIAAC and STEP provide another route for cross-country task measurement; see, for example, De la Rica et al. (2020) and Lewandowski et al. (2022).

where $\Gamma(\cdot)$ and $\gamma(s,x) \equiv \int_0^x t^{s-1}e^{-t}\,dt$ denote the gamma function and the lower incomplete gamma function, respectively. Moreover, the conditional average productivity of each task item satisfies $\overline{Z}_{L,\boldsymbol{\tau}}^{\lambda-1} = Z_\tau^{\lambda-1}\Gamma\left(1+\frac{1-\lambda}{\theta}\right)$. The elasticities of the c.d.f. and the selection functions characterizing the demand elasticities of Section 2.2 are given by[10]

$$\widetilde{G}(z^*_{\boldsymbol{\tau}}\,|\,\boldsymbol{\tau}) = \theta\,\widehat{z}_\tau^{-\theta} = -\theta\,\log\varphi_{K,\boldsymbol{\tau}}, \qquad \widetilde{G}_{\lambda-1}(z^*_{\boldsymbol{\tau}}\,|\,\boldsymbol{\tau}) = \theta\,\widetilde{\gamma}\left(1+\tfrac{1-\lambda}{\theta},\widehat{z}_\tau^{-\theta}\right), \tag{26}$$

where $\widetilde{\gamma}\,(s,x)$ denotes the elasticity of the lower incomplete gamma function with respect to its second argument given by $\widetilde{\gamma}\,(s,x) \equiv \frac{x^s\,e^{-x}}{\gamma(s,x)}$.

Given our assumption on $\mu(\cdot)$, the results on the firm-level cost function in Equations (12), (13), and (14) lead to the firm-level automated task share $\overline{\varphi}_K = \sum_{\tau\in\mathcal{T}}\xi_\tau\varphi_{K,\tau}$ and

$$\varphi_K = \sum_{\tau\in\mathcal{T}}\xi_\tau z_{K,\tau}^{\lambda-1}\varphi_{K,\tau}, \qquad \varphi_L = \sum_{\tau\in\mathcal{T}}\xi_\tau\overline{Z}_{L,\tau}^{\lambda-1}\varphi_{L,\tau}, \qquad c^f = r\sum_{\tau\in\mathcal{T}}\varphi_{K,\tau}/z_{K,\tau}^f. \tag{27}$$

All other results of Section 2.2 in characterizing the firm-level cost and demand functions follow in a similar fashion (Appendix A.1 presents all the expressions specialized to the case of the quantitative model).

2.3.2 Two Illustrative Examples

To build intuitions about the workings of this quantitative model, let us study it in two limiting cases: first, production without fixed costs, and second, production without marginal costs. These results illustrate important properties of the elasticities derived in Propositions 3 and 4 as the cost of machines declines.

Automation Without Fixed Costs First, consider the case in which automation involves no fixed costs, that is, $z_{K,\tau}^f \to \infty$ and $\zeta_\tau = 0$, which corresponds to the benchmark environments (e.g., Acemoglu and Restrepo, 2018a, 2020). Here, the task-item automation cutoff $\widehat{z}_\tau$ reflects only the standard marginal advantage of Equation (5).[11]

Without fixed costs, the elasticity of total cost with respect to output, or Equation (18) as $z_{K,\tau}^f \to \infty$, simplifies to $\varepsilon = 1$, implying constant returns to scale. The elasticities of total and marginal cost with respect to machine prices, respectively Equations (17) and (20) as $z_{K,\tau}^f \to \infty$, are equal, as $\Omega_K = \Omega_K^m = \omega_K$. The elasticities of machine and labor demand to output in Equation (22) are unitary, or $\eta_K = \eta_L = 1$. Furthermore, the firm-level elasticity

[10] We derive Equations (25) and (26) in Appendix A.1.

[11] All derivations are found in Appendix A.4.4 (page A10).

of substitution between labor and machines in Equation (21), using the closed form of the selection function elasticity in Equation (26), simplifies to

$$\sigma = \lambda + \frac{\theta}{\omega_K} \mathbb{E}_L \left[\widetilde{\gamma} \left(1 + \tfrac{1-\lambda}{\theta}, \widehat{z}_\tau^{-\theta} \right) \right], \tag{28}$$

since $\Omega_K = \Omega_K^m = \omega_K$ and the extensive–scale-advantage margin in Equation (21) vanishes when $\zeta_\tau = 0$. The surviving extensive–marginal term reflects only the marginal advantage and does not depend on scale.

Panel A of Figure 1 displays firm outcomes with declining machine prices r, tracing the firm's automation path. Due to the large differences in machine productivity across tasks, the share of automated tasks evolves in nearly discrete steps, increasing sharply then flattening out. This pattern reflects endogenous cycles of adjustment from partial to full automation of successive tasks. These cycles appear more clearly in the non-monotonic paths of the substitution elasticity and machine cost shares. Consequently, along the automation path, labor demand elasticities may generically be "lumpy."

Automation Without Marginal Costs Now, let us consider the opposite case of a firm without machine marginal cost, in the limit of $z_{K,\tau} \to \infty$ and $\zeta_\tau = 1$. Here, the relative automation cutoff $\widehat{z}_\tau$ reflects only the scale advantage defined in Equation (5).

Without marginal cost, the elasticity of total cost with respect to output, or Equation (18) as $z_{K,\boldsymbol{\tau}} \to \infty$, reflects increasing returns to scale, as $\varepsilon < 1$. The scale elasticity of marginal costs, given by Equation (19) as $z_{K,\boldsymbol{\tau}} \to \infty$, is negative, or $\varepsilon^m < 0$, suggesting that a firm's marginal costs are scale-dependent. Marginal cost is more elastic to machine prices than total costs, or $\Omega_K^m > \Omega_K > \omega_K$. The elasticities of factor demand to output in Equation (22) are no longer unitary, as $\eta_K > 1 > \eta_L$. Finally, the substitution elasticity is given by

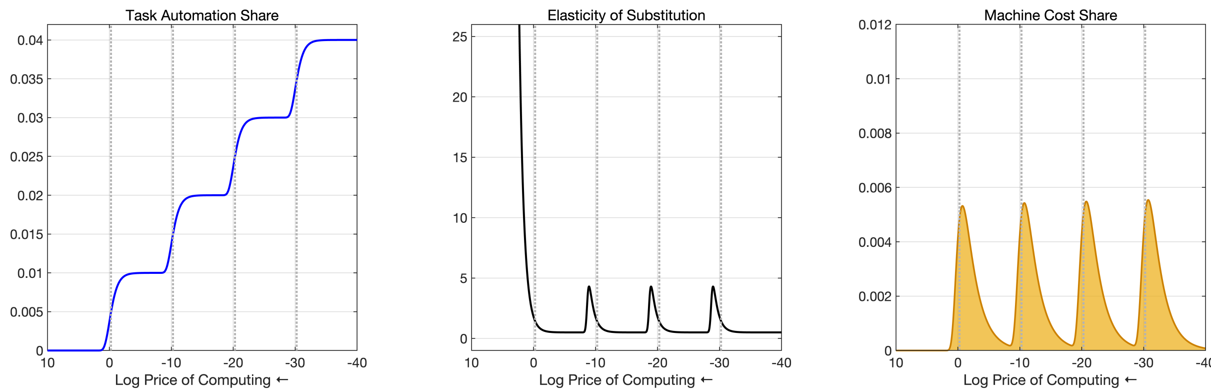
$$\sigma = \lambda \, \tfrac{\Omega_K^m}{\Omega_K} + \tfrac{1}{1-\lambda} \, \tfrac{1-\lambda\Omega_K^m}{\Omega_K} \, \mathbb{E}_L \left[\theta \, \widetilde{\gamma} \left(1 + \frac{1-\lambda}{\theta}, \widehat{z}_\tau^{-\theta} \right) \right]. \tag{29}$$

Although the intensive margin is inactive, scale advantage is highly responsive to machine prices, generating large labor-demand responses through the extensive margin of automation.

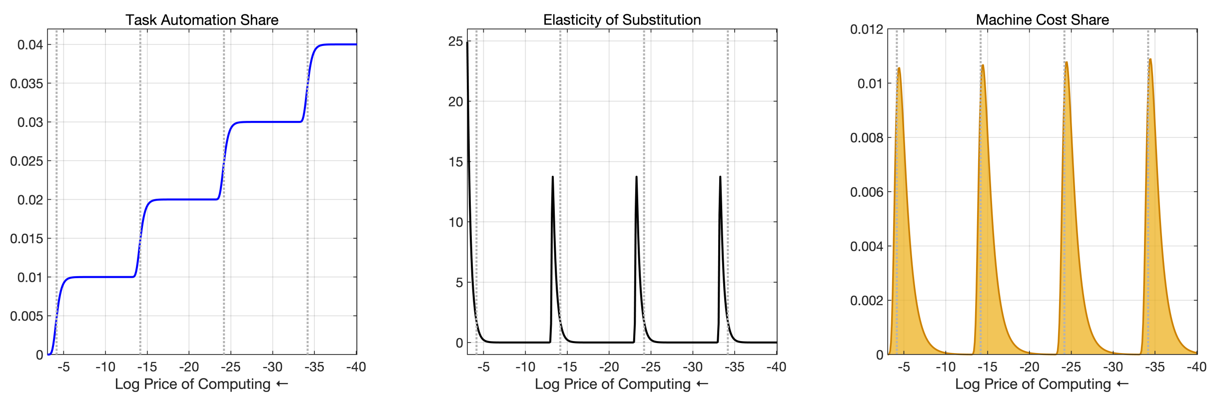
Panel B of Figure 1 shows the corresponding outcomes in the no-marginal-cost case. Many of the same qualitative patterns from the no-fixed-cost case in Panel A remain, including the step-like progression of the task automation share and the non-monotonicity of the firm's labor demand elasticities and machine cost shares along the adoption path. Here, since the intensive margin is shut down, the substitution elasticity falls to zero in

Figure 1: Outcomes of a Firm in Two Polar Cases as Prices of Machines Decline

Panel A. No Machine Fixed Costs



Panel B. No Machine Marginal Costs



Note: Both panels show toy examples for the parameters $\lambda = 0.5$, $\theta = 1.3$, and $|\mathcal{T}| = 5$, with $\xi_{\boldsymbol{\tau}} = 0.01$ for $\boldsymbol{\tau} \leq 4$ and $\overline{Z}_{L,\boldsymbol{\tau}} = 1$ for labor. Within each panel, the columns report, respectively, the share of automated tasks, the elasticity of substitution between labor and machines, and the machine share of costs, as the price of machines r declines. Panel A (no fixed costs) sets $z_K = (1, e^{-10}, e^{-20}, e^{-30}, 0)$ and $z_K^f = (\infty, \infty, \infty, \infty, 0)$, while Panel B (no marginal costs) sets $z_K = (\infty, \infty, \infty, \infty, 0)$ and $z_K^f = (1, e^{-10}, e^{-20}, e^{-30}, 0)$, for the variable and fixed productivities of machines, respectively.

the off-automation phases of the cycle.

In the no-marginal-cost case, three additional firm-level elasticities exhibit scale dependence along the automation path: the elasticity of marginal cost to machine prices Ω^m_K, the output elasticity of machine demand η_K, and the scale elasticity of marginal cost ε^m (Equations (20), (22), and (19)). These elasticities display non-monotonicities analogous to those discussed above: the elasticities of machine demand and prices spike when task items first get automated, then decrease as the labor productivity of the remaining tasks rises, while the scale elasticity of marginal cost is negative and drops sharply at the same points of discontinuity (see Appendix Figure A3).

2.3.3 General Equilibrium Environment

Finally, we embed the partial equilibrium framework of Section 2.3.1 into a general equilibrium environment and characterize the response of the resulting model to changes in machine prices.

We assume a set $\mathcal{J}_i$ of firms in industry i, each indexed by j, have Hicks-neutral productivity a_j distributed according to an industry-specific c.d.f. $H_i(a)$. Each firm operates with the production function described in Equation (1), where the vector of task-weights $\boldsymbol{\xi}_i$ is now industry-specific.

We assume an exogenous supply of labor $\overline{L}$ and machines $\overline{K}$. The supply of machines $\overline{K}$ corresponds to equilibrium price r and is fully owned by a unit mass of capitalist rentiers.

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

$$Y^{\frac{\rho-1}{\rho}} = \sum_{1\le i\le I} Y_i^{\frac{\rho-1}{\rho}}, \qquad Y_i^{\frac{\eta-1}{\eta}} = \sum_{j\in\mathcal{J}_i} y_j^{\frac{\eta-1}{\eta}}, \tag{30}$$

where industrial output Y_i is in turn a CES aggregator across the outputs of monopolistically competitive firms y_j within the industry. Here, η is the elasticity of substitution across firms and ρ is the elasticity of substitution across industries.

Given the simplicity of our environment, aside from the firm-level scale-dependence (nonhomotheticity), the characterization of the general equilibrium is otherwise standard and we relegate the derivations to Appendix A.2.

Elasticities of Aggregate Costs and Demand The market equilibrium product allocations are efficient in this model. As a result, we can define an aggregate production function and a corresponding cost function $\overline{C}(w,r;Y) = \bar{c}(w,r;Y)\,Y + \overline{C}^f(w,r;Y)$ where

Y denotes the aggregate consumption defined by Equation (30). Additionally, we define the aggregate counterparts to all the firm-level elasticities defined in Section 2.2. These include those of total costs with respect to output $\bar{\varepsilon} \equiv \frac{\partial \log \overline{C}}{\partial \log Y}$ and machine prices $\overline{\Omega}_K \equiv \frac{\partial \log \overline{C}}{\partial \log r}$, marginal costs with respect to output $\bar{\varepsilon}^m \equiv \frac{\partial \log \bar{c}}{\partial \log Y}$ and machine prices $\overline{\Omega}^m_K \equiv \frac{\partial \log \bar{c}}{\partial \log r}$, the scale elasticities of capital and labor demand $\bar{\eta}_K \equiv \frac{\partial \log K}{\partial \log Y}$ and $\bar{\eta}_L \equiv \frac{\partial \log L}{\partial \log Y}$, and the substitution elasticity $\bar{\sigma} \equiv \frac{1}{\overline{\Omega}_K} \frac{\partial \log L}{\partial \log r}$. Appendix Section A.3 extends the aggregation results of Lashkari et al. (2024) (derived in the case of a single-sector environment) to characterize all these aggregate elasticities to the multi-sector environment considered here.

General Equilibrium Response to Shock in Machine Prices We now consider the general equilibrium response of a small change $\Delta \log r$ in the price of machines on the aggregate output, wages, and factor shares. This shock captures the secular fall in the dollar cost of compute, due to hardware progress (Moore's-Law-type gains) and algorithmic improvement, which continuously expands the effective supply of machines. Formally, we model this as an exogenous shock $\Delta \log \overline{K}$ to the (effective) supply of machines, leading to the equilibrium change $\Delta \log r$ in the machine price.

Proposition 5. *Consider a decline in the price of machines, $\Delta \log r$. The general equilibrium responses of wage w, output Y, and the labor share $\overline{\Omega}_L$ follow*

$$\Delta \log w \approx -\frac{\bar{\eta}_L \overline{\Omega}^m_K - \bar{\varepsilon}^m \overline{\Omega}_K \bar{\sigma}}{\bar{\eta}_L \overline{\Omega}^m_L + \bar{\varepsilon}^m \overline{\Omega}_K \bar{\sigma}} \Delta \log r, \tag{31}$$

$$\Delta \log Y \approx -\frac{\left(\overline{\Omega}^m_K + \overline{\Omega}^m_L\right) \overline{\Omega}_K}{\bar{\eta}_L \overline{\Omega}^m_L + \bar{\varepsilon}^m \overline{\Omega}_K \bar{\sigma}} \bar{\sigma} \, \Delta \log r, \tag{32}$$

$$\Delta \log \overline{\Omega}_L \approx \frac{\left(\overline{\Omega}^m_K + \overline{\Omega}^m_L\right) \overline{\Omega}_K}{\bar{\eta}_L \overline{\Omega}^m_L + \bar{\varepsilon}^m \overline{\Omega}_K \bar{\sigma}} \left(\bar{\varepsilon} \bar{\sigma} - \bar{\eta}_L\right) \, \Delta \log r. \tag{33}$$

Proof. We can derive the response only based on two conditions: aggregate labor-market clearing $\overline{L} = L(w, r; Y)$ and the definition of the price index as the numeraire $\frac{\eta}{\eta - 1} \bar{c} = 1$. Noting that labor supply remains constant and taking first-order approximations of these conditions given a change $\Delta \log r$ in the price of machines yields

$$0 \approx \overline{\Omega}_K \bar{\sigma} \left(\Delta \log r - \Delta \log w\right) + \bar{\eta}_L \Delta \log Y,$$
$$0 \approx \overline{\Omega}^m_K \Delta \log r + \overline{\Omega}^m_L \Delta \log w + \bar{\varepsilon}^m \Delta \log Y.$$

Solving the two linear equations above in terms of the two unknowns $\Delta \log w$ and $\Delta \log Y$

leads to Equations (31) and (32). Equation (33) follows from $\overline{\Omega}_L = w\overline{L}/\overline{C}$ and the first-order approximation $\Delta \log \overline{C} = \overline{\Omega}_L \Delta \log w + \overline{\Omega}_K \Delta \log r + \bar{\varepsilon} \Delta \log Y$. □

In our empirical application, the aggregate share of automated tasks is small. As such, even if scale-dependence may be large at certain firms with higher degrees of adoption, the aggregate economy might remain nearly scale-invariant. In Corollary 1 below, we compare the responses derived in Proposition 5 to a standard scale-invariant approximation of the model, ignoring the fixed costs in the aggregate.

Corollary 1. *Suppose $z^f_{K,\tau} \to \infty$ for all tasks τ. Consider a decline in the price of machines, $\Delta \log r$. The general equilibrium responses of wage w, output Y, and the labor share $\overline{\Omega}_L$ follow*

$$\Delta \log w \approx -\frac{\overline{\Omega}_K}{\overline{\Omega}_L}\Delta \log r, \quad \Delta \log Y \approx -\overline{\sigma}\frac{\overline{\Omega}_K}{\overline{\Omega}_L}\Delta \log r, \quad \Delta \log \overline{\Omega}_L \approx (\overline{\sigma} - 1)\frac{\overline{\Omega}_K}{\overline{\Omega}_L}\Delta \log r. \quad (34)$$

3 Application: AI Scaling and Computer Vision

To showcase the potential of the quantitative model set up in Section 2, we apply it to study the economic impacts of a specific AI technology: computer vision. This application allows us to rely on AI scaling laws to uncover measures of AI marginal costs $1/z_{K,\tau}$ and fixed costs $1/z^f_{K,\tau}$ for all tasks subject to computer vision automation in the US economy. Using these measures, along with corresponding measures for labor productivity, our framework determines the patterns of marginal, scale, and comparative advantage for AI computer vision.

In Section 3.1, we discuss AI scaling laws, which are robust empirical relationships between model size, training data, and model performance. In Section 3.2, we estimate scaling laws for the fine-tuning of AI computer vision systems. Finally, in Section 3.3, we use LLMs to generate characteristics of production tasks involving substantial vision content to infer the level of performance required to automate each task. Combining these task-level performance requirements with the estimated scaling law allows us to infer the computation costs of automating each vision-related task with AI.

3.1 AI Scaling Laws and Task-Level Automation Costs

Along with the rapid progress of AI, an empirical literature has emerged that studies the scaling behavior of AI systems, demonstrating robust and predictable relationships between their scale and performance (Hestness et al., 2017; Rosenfeld, 2021; Bahri et al., 2024). These relationships hold across domains as varied as LLMs (Kaplan et al., 2020) and computer vision (Zhai et al., 2022). Here, we discuss how we can use the scaling laws to characterize the productivity of AI across different vision tasks.

Scaling Laws In principle, AI deployment entails both fixed and marginal costs. Fixed costs are associated with *training* a model of size m on a dataset of size d. At the training stage, computation costs grow with the product of model size and data size, or $m \times d$. Marginal costs are associated with making *inference* using the trained model: each instance of inference for a given task requires computation that scales with model size m.

Let h be the prediction loss of the AI system on the test data—e.g., the cross-entropy loss in out-of-sample classification. This loss is an inverse measure of model performance, where lower loss indicates that the model predicts more accurately, and is thus an empirical counterpart to the AI performance requirements introduced in Section 1. Among others, Hoffmann et al. (2022) have shown that AI performance is related to training inputs based on the following scaling law:

$$h = \mathcal{H}\left(m, d; \boldsymbol{\alpha}, \boldsymbol{\psi}, \underline{h}\right) \equiv \left(\psi_m m\right)^{-\alpha_m} + \left(\psi_d d\right)^{-\alpha_d} + \underline{h}, \tag{35}$$

where, as before, m and d denote model size and the size of training data, respectively. The parameters α_m and α_d govern the returns to increasing training inputs, while ψ_m and ψ_d capture model- and data-augmenting efficiency. These efficiency shifters may vary with task complexity or model architecture. The term $\underline{h}$ is irreducible loss, which denotes the lower bound on achievable prediction loss in the test data.

Typically, training costs constitute the majority of AI development and deployment costs. As such, the state-of-the-art approach to designing AI systems selects the combination of model and data sizes that minimize computation costs of the training stage, measured commonly in units of "compute," that is, the number of floating point operations (FLOPs), for achieving a given level of prediction loss h (Hoffmann et al., 2022):[12]

$$k^f = \min_{m,d} 6\, m \times d \qquad \text{subject to } \; h = \mathcal{H}\left(m, d; \boldsymbol{\alpha}, \boldsymbol{\psi}, \underline{h}\right), \tag{36}$$

where k^f denotes the fixed training units of compute, the scaling function $\mathcal{H}\left(m, d; \boldsymbol{\alpha}, \boldsymbol{\psi}, \underline{h}\right)$ is given by Equation (35), and the multiplicative factor 6 accounts for the average number of FLOPs for each training step and token of data (Kaplan et al., 2020).

[12]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.

Task-Level AI Scaling We use this framework to recover the comparative and scale advantages of AI across tasks. Following Section 2.3, let τ denote a specific task item in the O*NET classification. For readability, henceforth, we refer to these task items simply as tasks when no confusion arises. For each task τ, we assign a required prediction loss h_τ, scaling parameters $\boldsymbol{\alpha}_\tau$, efficiency shifters $\boldsymbol{\psi}_\tau$, and irreducible loss $\underline{h}_\tau$. The parameter h_τ can be interpreted as a *task-specific tolerance* for prediction loss: an AI system can only automate task τ if it achieves a loss lower than h_τ. Sections 3.2 and 3.3 below describe our construction of the task-specific inputs $(h_\tau, \boldsymbol{\psi}_\tau, \boldsymbol{\alpha}_\tau, \underline{h}_\tau)$, which characterize the AI scaling problem.

Applying Equation (36) at the task-specific requirement $h = h_\tau$ gives the fixed cost of training a model that achieves the performance required to automate task τ in terms of "compute," which we henceforth associate with our units of machines. This equation also yields the marginal cost of automation, which is proportional to the training-cost-minimizing model size m_τ. Given these task-level characteristics, Lemma 1 characterizes the fixed and marginal components of AI productivity in task τ.

Lemma 1. *Given tuple $(h_\tau, \boldsymbol{\psi}_\tau, \boldsymbol{\alpha}_\tau, \underline{h}_\tau)$, consider the training compute-minimization problem in Equation (36) for task τ. Then, the marginal $z_{K,\tau}$ and fixed cost $z^f_{K,\tau}$ productivity of AI for the task are given by*

$$z^f_{K,\tau} = \tfrac{1}{6}\, \psi_{d,\tau}\psi_{m,\tau} \left(\frac{\alpha_{m,\tau}}{\alpha_{d,\tau} + \alpha_{m,\tau}}\right)^{\frac{1}{\alpha_{d,\tau}}} \left(\frac{\alpha_{d,\tau}}{\alpha_{d,\tau} + \alpha_{m,\tau}}\right)^{\frac{1}{\alpha_{m,\tau}}} (h_\tau - \underline{h}_\tau)^{\left(\frac{1}{\alpha_{m,\tau}} + \frac{1}{\alpha_{d,\tau}}\right)}, \tag{37}$$

$$z_{K,\tau} = \psi_{m,\tau} \left(\frac{\alpha_{d,\tau}}{\alpha_{d,\tau} + \alpha_{m,\tau}}\right)^{\frac{1}{\alpha_{m,\tau}}} (h_\tau - \underline{h}_\tau)^{\frac{1}{\alpha_{m,\tau}}}. \tag{38}$$

Proof. The proof is found in Appendix A.4.5 (page A12). □

Equations (37) and (38) provide formulas for calculating AI fixed (i.e., training) and marginal (i.e., inference) productivities in terms of the scaling parameters. All else equal, tasks with higher tolerance for prediction loss h_τ require less compute to automate, lowering both fixed training and marginal inference costs. The effect is stronger for training than for inference productivity, and the gap between the two elasticities amounts to $1/\alpha_{d,\tau}$. This gap determines how quickly fixed costs fall relative to marginal costs as tolerance for prediction loss increases.

Both components of AI productivity also depend on the efficiency shifters $\boldsymbol{\psi}_\tau$. For training, productivity depends on both model- and data-augmenting efficiency, through the product of the two shifters. For inference, only model-augmenting efficiency matters. As a result, the data-augmenting efficiency term, $\psi_{d,\tau}$, creates an additional wedge between fixed training and marginal inference costs.

If we know labor productivity Z_τ across tasks, Lemma 1 fully characterizes the comparative and scale advantages of AI across tasks based on information on the AI scaling tuple $(h_\tau, \boldsymbol{\psi}_\tau, \boldsymbol{\alpha}_\tau, \underline{h}_\tau)$. In Section 3.2, we discuss how we estimate scaling law parameters from computer vision fine-tuning experiments.

3.2 Estimating Scaling Laws for Computer Vision Fine-Tuning

We next estimate the scaling law parameters that allow us, when combined with task information, to infer our measures of marginal, scale, and comparative advantage using Lemma 1. We use data from a computer vision experiment focused on *fine-tuning*, rather than the initial training of foundation models (Hernandez et al., 2021; Zhang et al., 2024).[13] Fine-tuning captures the "last mile" of AI deployment, in which a general-purpose model is adapted to a specific task, thus bridging the gap between generic models and the economic needs of a firm.

In our experiment, we fine-tune a foundation computer vision model to perform tasks at varying degrees of accuracy.[14] As described in Section 3.1, the model's performance is evaluated as the out-of-sample prediction loss when performing an object classification task, where higher values of loss indicate lower accuracy. We vary three dimensions of the fine-tuning process: (i) the *task complexity*, measured by the number of distinct classes or outcomes the model must distinguish in the underlying classification task (2, 10, 100, or 500 classes), (ii) *size of the training data*, measured by the number of sample images used in training (13, 65, 130, 650, and 1300 samples per class), and (iii) the *number of model parameters*, specified to be 7.3 thousand, 0.4 million, 28.3 million, and 87.8 million parameters. We repeat each combination of fine-tuning specifications 50 times, resulting in 4000 total observations.

We use outcomes from the experiment to estimate the scaling laws of Equation (35). To do so, we allow the power-law parameters $\boldsymbol{\alpha}$ and efficiency-shifters $\boldsymbol{\psi}$ of the scaling law to vary with task complexity. For a task τ with complexity χ_τ, the parameters follow

[13] Although our experiment focuses on fine-tuning compute, task-specific upfront costs also include collecting and labeling the training data itself, as well as other forms of last-mile adaptation such as prompt engineering. These costs are conceptually equivalent to the fixed costs in our theory, and the data-collection component likely co-moves with our compute-based measure, since the same compute-minimization problem determines the required data size d_τ.

[14] The foundation model is a Swin Transformer pre-trained on 500 randomly selected classes from the ImageNet dataset (Liu et al., 2021).

the log-log functional form of

$$\log \alpha_{f,\tau} = \beta^{\alpha}_{f,0} + \beta^{\alpha}_{f,1} \log \chi_{\tau}, \quad \log \psi_{f,\tau} = \beta^{\psi}_{f,0} + \beta^{\psi}_{f,1} \log \chi_{\tau}, \quad f \in \{m, d\}. \tag{39}$$

The specification above reflects the possibility that more complex tasks may be more difficult for the foundation model to "learn" to sufficient accuracy, which results in higher computational training costs. We estimate Equations (35) and (39) by nonlinear least squares; Appendix Table A2 reports the estimated coefficients.

A final consideration concerns the economic content of a model's task-level performance. The scaling law and compute-minimization problem in Equation (36) are expressed in terms of prediction loss h_{τ}. However, workers and LLMs cannot accurately assess the cross-entropy or prediction loss threshold a production task typically requires. Instead, in Section 3.3, we elicit performance requirements of real production tasks in terms of *error tolerance*, or the maximum error rate a worker can incur while still performing the task satisfactorily. For this reason, we also record each fine-tuned model's prediction accuracy in our experiment and estimate an empirical mapping between error rates and prediction loss h_{τ}, conditional on task complexity (Appendix B.1). This mapping translates the task-level error tolerance measures we construct next into the required loss levels h_{τ}.

3.3 Measures of Task Characteristics

We use a frontier large language model to construct task-level measures for 18,939 task items across 924 occupations in O*NET.[15] For those classified as vision-related, we elicit the characteristics needed to compute AI productivity from the scaling law, including error tolerance and task complexity. We also elicit the characteristics needed to compute labor productivity, including task frequency and time use.

The LLM-generated measures extend and complement survey-based measures of task characteristics, which are available for a smaller set of tasks and occupations (Svanberg et al., 2024). To validate our approach, we compare the LLM-generated responses to the average worker responses in that survey, using its questions adapted as LLM prompts, and find a strong positive correlation across task complexity, error tolerance, and frequency (Appendix Figure A1). This section describes the elicitation procedure and summarizes the resulting task characteristics. Appendix B provides additional details, including the prompt text and survey design.

[15] All model-generated measures in this section are produced with GPT-5; the complete set of prompts is provided in Appendix B.3.

Categorization of Task Items First, we use the LLM to determine whether task items in the O*NET data contain enough visual content to be considered "automatable" by computer vision, in which case their automation costs follow the scaling law estimated in Section 3.2. To do so, we ask the LLM to consider five automation types and to assign a decimal weight to the proportion of the task the technology can automate, where weights must sum to one. Automation types are (i) robotics and physical automation, (ii) analytical/predictive automation, (iii) computer vision, (iv) LLM or natural language processing automation, and (v) other or miscellaneous. We identify a task as vision-related if answers to all of ten repeated prompts assign at least half the weight to computer vision (Appendix Section B.3); this yields 659 vision tasks, constituting 3.5% of total tasks in the economy.

We map the tasks to 4-digit NAICS industries using O*NET measures of task frequency and different industries' occupational composition.[16] Across 247 industries, an average of 9.3% and a median of 8.2% of frequency-weighted tasks can be automated by computer vision.[17]

Task Error Tolerance and Complexity Among tasks that can be automated with computer vision, we use the LLM-generated data to construct two measures: (i) task error tolerance, and (ii) task complexity, which acts as a shifter on scaling law coefficients.[18] Previous estimates of these task-level characteristics have been elicited from workers in a survey completed by Svanberg et al. (2024). The average of workers' responses to "What's the worst error rate that a worker doing this task could have and still be considered qualified to do it?" constitutes a measure of task error tolerance. The average of workers' responses to "How many possible choices/categories are in the outcome?" reflects a measure of task complexity. We use these questions as prompts for the LLM, which responds considering only the "visual" component of tasks deemed automatable via computer vision. We elicit ten answers from the LLM for each task and take the average response.

To better understand the interpretation of each variable, let us consider practical examples where computer vision could automate the relevant visual component of a task.

[16] We map tasks from occupation to industries using the occupation's wage-bill share and weigh tasks in each occupation by their frequency, as elicited by the LLM (Section 3.3). The details of this concordance are in Appendix B.4.

[17] The most vision-intense industry is "Child Care Services" (NAICS 6244), where 36.9% of tasks by frequency have a vision component. The least vision-intense industry is Legal Services (NAICS 5411), where only 0.6% of tasks by frequency have a vision component.

[18] The functional form of the mapping of task error tolerance to prediction loss and the functional form of the scaling law coefficients with respect to task complexity are given respectively by Equations (A29) and (39).

Talent directors must "attend or view productions to maintain knowledge of available actors," an extremely complex task with an estimated 1520 potential outcomes. Air traffic controllers complete a less complex task: "monitoring aircraft within a specific airspace, using radar, computer equipment, or visual references" is associated with only 9 potential outcomes. The latter, however, can be similarly costly to automate, as the talent director's error tolerance is large at 20.5% while the air traffic controller's is extremely small at 0.1%.

The correspondence between the survey and LLM estimates is shown in Figure A1 (Section B.5 of the Appendix). Relative to survey estimates alone, the LLM procedure allows us, first, to reduce dispersion in our estimates of task characteristics, and second, to estimate task characteristics across all 659 vision tasks identified in O*NET, instead of only the 445 tasks for which valid responses were recorded in the survey.

For each task, we calculate the prediction loss associated with task characteristics from the LLM data in Equation (A29). We then use Equations (37), (38), and (39) to estimate the costs of fine-tuning a foundation model to the appropriate specifications.

Labor Productivity in Tasks Finally, we also use the LLM to construct measures of average task-level labor productivity as the ratio of the frequency of fulfilling the task (output) to the share of worker time spent on the task (time input). Since each task item belongs to a single occupation, we write $o = o(\tau)$ for the occupation of task τ. Denote the share of time spent on a task τ as a share of total working hours in occupation o as $TimeUse_{o\tau}$ and the task's monthly frequency as $Freq_{o\tau}$. We ask the LLM to provide estimates for both characteristics across all tasks, where prompts are shown in Section B.3 of the Appendix. We define average labor productivity of a task τ as $Z_\tau = Freq_{o\tau} / \left(TimeUse_{o\tau} \times w_o\right)$ where w_o are wages obtained from the 2023 Occupational Employment and Wage Statistics (OEWS) for an occupation o.

4 Quantitative Exercise

In this section, we use our estimates of the comparative advantage of AI in Section 3 to quantify the theory developed in Section 2 and study the aggregate consequences of computer vision AI automation across different firms and industries. We present our approach to the calibration in Section 4.1 and the results in Section 4.4.

4.1 Model Calibration

In addition to the primitives that characterize the patterns of comparative advantage of AI, 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-variety dispersion in task-level labor productivities. We calibrate these parameters to match values common in the prior literature.[19] Table 1 shows these calibrated parameters.

Firm Productivity Distribution Since scale is an important driver of automation, we carefully calibrate the distribution of firm size at the industry level. We use the 2022 Business Dynamics Statistics (BDS) to infer the within-industry firm productivity distribution using employment shares across firm-size bins. We assume that firm productivities in an industry are distributed Pareto, with an industry-specific scale parameter $\underline{a_i}$ and a shape parameter ϑ. For each 4-digit industry i, BDS reports total employment within each firm size bin. Each bin is assigned a representative employment threshold ℓ_j, which corresponds to the bin's lower bound. We construct total employment L_i as the sum of all employment across all bins within an industry.

We use the following regression to estimate parameters of the distribution

$$\log \tfrac{L_i(\ell_j \geq \ell)}{L_i} = \left(1 - \tfrac{\vartheta}{\eta - 1}\right) \log \ell + \alpha_i + \epsilon_i, \tag{40}$$

where $\frac{L_i(\ell_j \geq \ell)}{L_i}$ represents the share of labor in industry i with employment above the cutoff ℓ and α_i are industry-level fixed effects. This approach yields the value $\vartheta = 3.7$.

We obtain the industry scale parameters $\underline{a_i}$ by inverting the empirical industry-level employment shares $\widehat{\Psi}_i$ in the 2023 OEWS, such that $\underline{a_i} = \widehat{\Psi}_i^{\frac{1}{\rho-1}} J_i^{-\frac{1}{\eta-1}} \frac{w}{\overline{Z}_{L,i}}$ and $\overline{Z}_{L,i}^{\lambda-1} = \sum_\tau \xi_{i\tau} \overline{Z}_{L,\tau}^{\lambda-1}$ denotes average labor productivity in an industry. Industry-task weights $\boldsymbol{\xi}$ are computed in Appendix Section B.4.[20]

Inferring the Computing Costs Taking the task-level AI and labor productivities constructed in Section 3 as given, two degrees of freedom remain: the baseline price of com-

[19] 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 $\min_i \bar{\varepsilon}_i^m > -1/\rho$.

[20] Derivations are provided in Appendix A.4.6.

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	Broda and Weinstein (2006)			
Industry EoS	ρ	0.8	Berlingieri et al. (2024)			
Labor productivity	θ	1.3	Burstein et al. (2019)			
Capital price change	$\Delta \log r$	-0.31	Del Sozzo et al. (2025)			
		-1.35	Ho et al. (2024)			
Calibration						
Productivity Pareto Tail	ϑ	3.7	BDS			
Capital price	$\log r$	-14.9	Bonney et al. (2024)	Fall 2023 Computer Vision adoption rate (unweighted)	0.5%	0.6%
Fixed cost shifter	$\log B$	-17.0	Bonney et al. (2024)	Fall 2023 Computer Vision adoption rate (weighted)	2.4%	2.4%

Note: Calibrated model parameters. The top panel reports parameters set to values from the literature; the bottom panel reports calibrated parameters, where the column "Moment" describes the targeted data moment and the columns "Literature" and "Model" report its empirical value and model counterpart, respectively. The elasticities of substitution across firms (η) and industries (ρ) are set to conventional values from the literature. The fixed cost shifter B scales the training productivities $z^f_{K,\tau}$ of Equation (37), capturing how frequently firms incur the fixed cost of training AI models in our static model (Section 4.1).

pute r and the relative size of fixed to variable costs B. Because the model is static, mapping task-level fixed automation costs to annual data requires an assumption about how often firms incur the fixed cost of training an AI model. We capture this assumption with a constant shifter B applied to all training productivities $z^f_{K,\tau}$ in Equation (37).

We calibrate these two parameters in the baseline economy to match AI adoption rates from the September 2023 Business Trends and Outlook Survey (BTOS), as reported by Bonney et al. (2024). At that time, 0.5% of firms reported using computer vision tools, while the employment-weighted adoption rate was 2.4%, suggesting that larger firms were more likely to adopt. Intuitively, the first moment pins down the overall level of automation costs, while the second pins down the role of scale in adoption. The calibrated values are reported in Table 1. The baseline model generates an unweighted adoption rate of 0.6% and an employment-weighted adoption rate of 2.4%.

In the baseline quantification, we assume that each firm incurs the fixed training cost for every task it automates. As an alternative, we also consider the case in which fine-tuning costs are incurred once per task at the industry level, with each industry composed of a representative firm. In this case, productivities $\underline{a_i}$ are recalibrated to match the observed labor shares across industries in the 2023 OEWS. We present the results for this alternative case in Section 4.5.

We project the trends in AI automation forward in time to examine workers' economic outcomes as the cost of compute r declines. There is some uncertainty regarding the rate of decline in compute costs, although most estimates suggest a rapid decrease. As our baseline, we use recent estimates from Del Sozzo et al. (2025), which imply an annual decline of 26.4% in the dollar cost per FLOP. We also consider an alternative decline in the price of machines corresponding to 74.0%, which combines the baseline estimate with a rate of decline that accounts for algorithmic progress (Ho et al., 2024), as reported in Table 1 and Appendix Figure A10.

4.2 Calibrated Patterns of Comparative Advantage

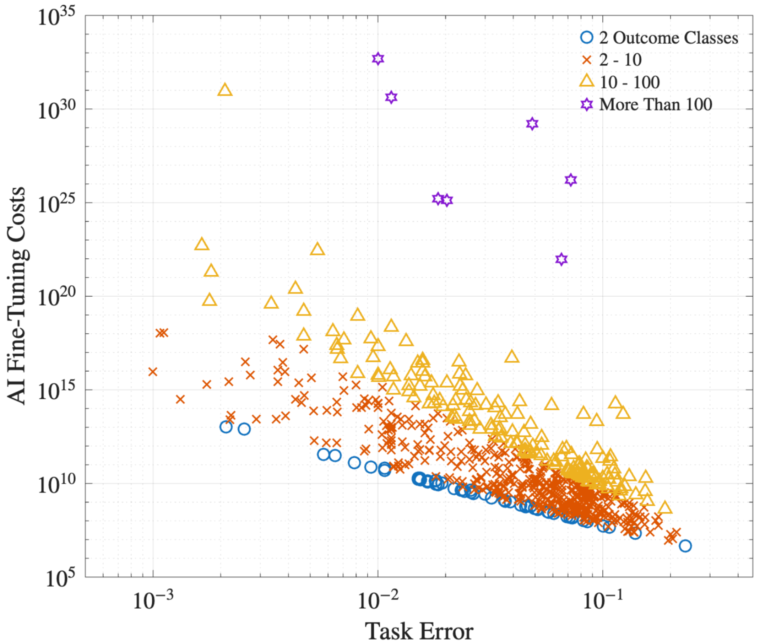
We begin by exploring the patterns of comparative advantage under our baseline calibration in 2023. Figure 2 visualizes the scaling law of Equation (35), calculating the relationship between task characteristics and fine-tuning costs for all vision tasks in the economy. More complex tasks are costlier to automate, and within each complexity group, tasks with lower error tolerance have higher automation costs. The slope of the relationship between error tolerance and fine-tuning costs becomes steeper for more complex tasks. Appendix Figure A4 illustrates the relationship between task error tolerance and estimated AI costs using a number of specific task examples to build intuition.

The Importance of Scale Advantage Next, we assess the quantitative importance of scale advantage in shaping comparative advantage across tasks. From Equation (5), for a task item τ, comparative advantage follows $\log CA_\tau = \log MA_\tau + \log SA_\tau$. This log-additive structure allows us to decompose the variance of comparative advantage into scale- and marginal-advantage components by regressing each component on log comparative advantage.[21] Figure 3 displays the results of this exercise for the calibrated baseline: in 2023, the scale component accounts for 63% of the cross-task variation in comparative advantage. This finding is robust to performing the decomposition using averages across broader task categories, as shown in Appendix Figure A5. Thus, most variation in comparative advantage arises from scale rather than marginal cost differences, indicating that scale considerations are a crucial determinant of task-level automation patterns.

Proposition 1 implies that comparative advantage eventually converges to marginal advantage as aggregate output grows. Quantitatively, however, this asymptotic result has

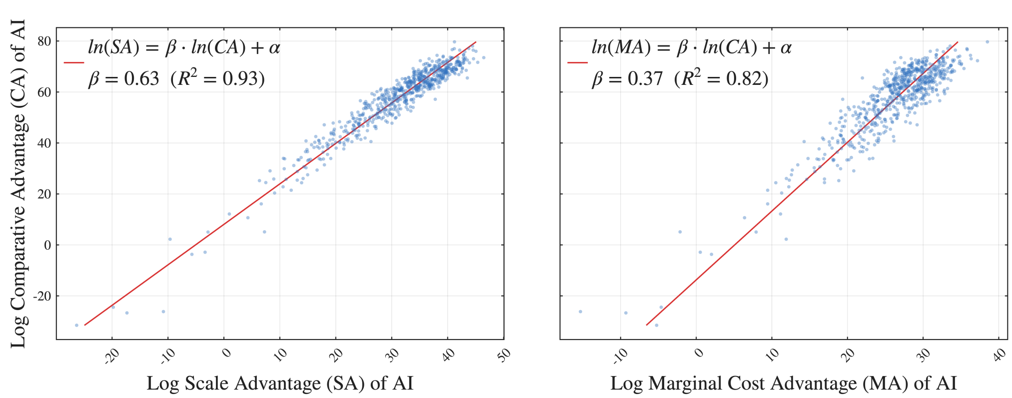
[21]Given the linear additivity, the following identity holds: $\beta_s + \beta_m \equiv \frac{\text{Cov}(\log SA_\tau, \log CA_\tau)}{\text{Var}(\log CA_\tau)} + \frac{\text{Cov}(\log MA_\tau, \log CA_\tau)}{\text{Var}(\log CA_\tau)} = 1$. Through OLS, we can uncover β using the regressions $\log SA_\tau = \alpha + \beta_s \log CA_\tau + \epsilon^s_\tau$ and $\log MA_\tau = \alpha + \beta_m \log CA_\tau + \epsilon^m_\tau$, the results of which are shown in Figure 3. The regressions use one observation per task, with $\log CA_\tau$, $\log MA_\tau$, and $\log SA_\tau$ constructed as unweighted means of their firm–industry-level log values; averaging preserves the log-additive identity, so $\beta_s + \beta_m = 1$ exactly.

Figure 2: Task Characteristics and Fine-Tuning Costs



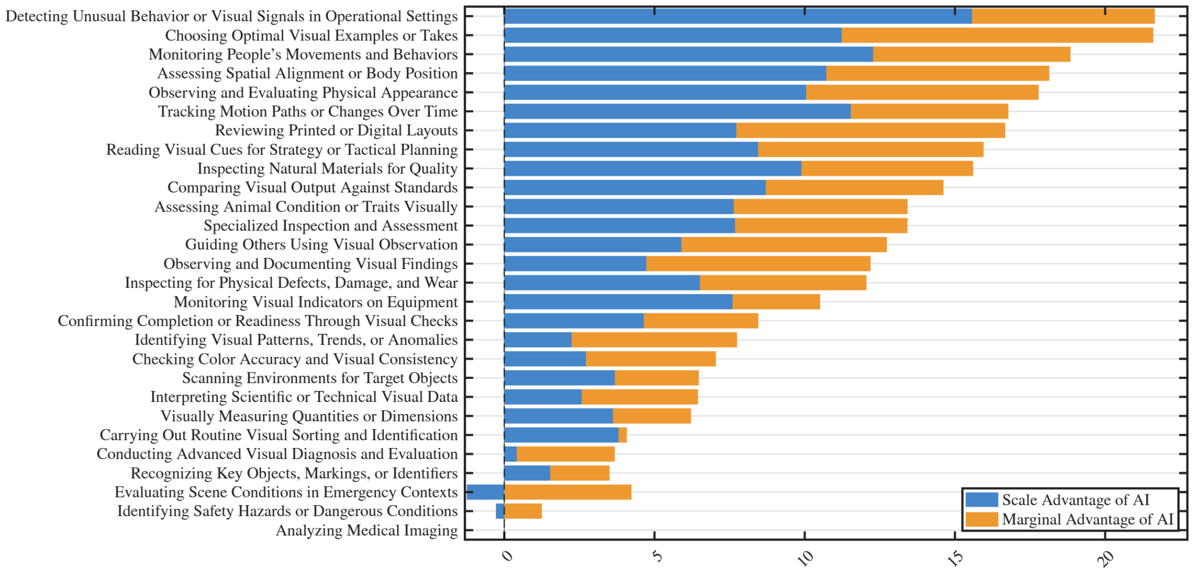
Note: This figure maps AI fine-tuning costs in terms of units of compute as a function of task error and complexity, representing how "scaling laws"—which assign high costs to models achieving lower prediction loss—look across a matrix of task characteristics. We split task complexity into four types: (i) *low*, corresponding to 2 outcome classes, (ii) *moderate*, more than 2 and less than 10 classes, (iii) *high*, more than 10 and less than 100 classes, and (iv) *very high*, more than 100 classes.

Figure 3: Variance Decomposition of AI Comparative Advantage Across Tasks in 2023



Note: This figure considers the economy at the calibrated 2023 baseline price of compute. It decomposes across-task variation in comparative advantage CA into that attributable to scale SA and marginal MA advantage. Scale accounts for 63% of the variation.

Figure 4: Average AI Comparative and Scale Advantage Across Task Categories



Note: The figure shows the unweighted average of AI's task-level marginal advantage and scale advantage in each category of tasks relative to "analyzing medical imaging", which has the smallest comparative advantage. We index the values associated with the "analyzing medical imaging" group as 0. The blue segment shows $\log SA_\tau - \log SA_0$ and the orange segment shows $\log MA_\tau - \log MA_0$. Task-level comparative advantage is computed as the average of firm-task-level comparative advantage estimates across firms. All 659 vision tasks are mapped into vision task categories generated by GPT-5.

limited relevance over economically meaningful horizons. When the model is projected forward to 2050, scale advantage still accounts for roughly 60% of the cross-task variation in comparative advantage. Appendix Figure A6 reports this variance decomposition.

Figure 4 shows the average of AI comparative advantage and its components within selected bundles of task groups at our baseline calibration.[22] Here, AI exhibits the greatest comparative advantage in "detecting unusual behavior or visual signals in operational settings" and the least comparative advantage in "analyzing medical imaging".

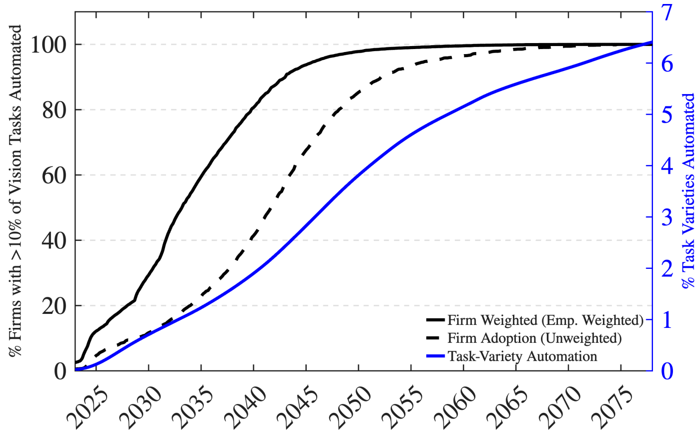
4.3 The Automation Path with Declining Compute Prices

We use the calibrated model to project the path of computer vision AI automation from 2023 to 2075 as computing costs decline. Figure 5 shows rapid growth in adoption over the first decade of the transition, especially when firms are weighted by employment. By 2025, the unweighted adoption rate reaches 4.7%, while the employment-weighted adoption rate reaches 12.2%. By 2035, these rates rise to 22.8% and 59.5%, respectively.

These adoption rates capture the extensive margin of firm-level automation: the share

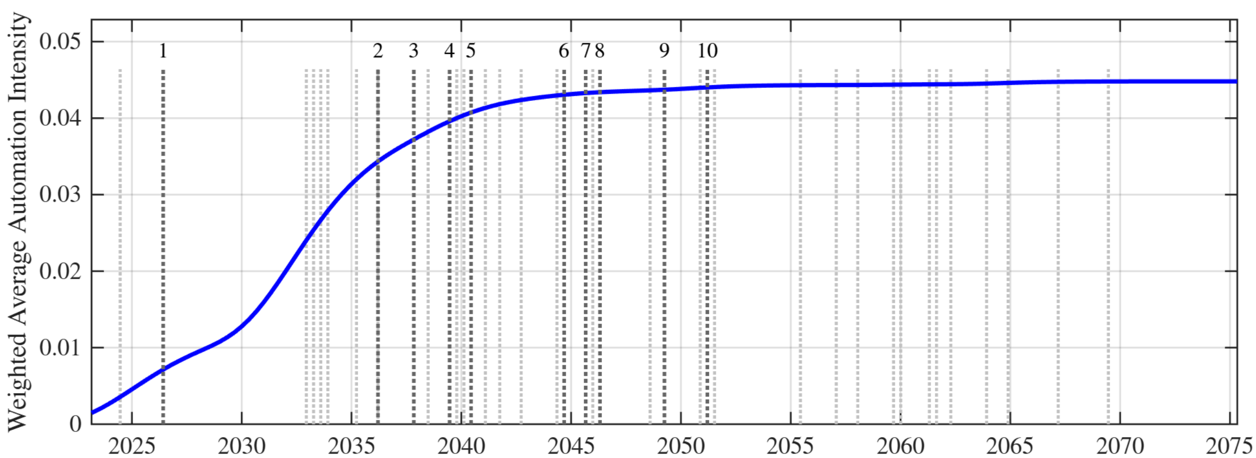
[22] Vision tasks were classified into vision task categories (i.e., "Visual Generalized Work Activities (GWAs)") generated by GPT-5.

Figure 5: The Automation Path of AI



Note: The figure shows the evolution of aggregate adoption as the cost of compute declines. The solid and dashed black lines (left axis) show the share of firms for which more than 10% of vision-task varieties are automated, employment-weighted and unweighted, respectively. The blue line (right axis) shows the cost-weighted average share of automated task varieties across *all* tasks and firms in the economy; because only the 659 vision tasks (3.5% of task items) are subject to computer-vision automation, this share remains far below 100%. We calibrate the cost of compute to match 2023 firm AI adoption in the Business Trends and Outlook Survey as the baseline.

Figure 6: Automation Path for a Gas Station (NAICS 4571)



Automated Tasks:

1. Watch for and recognize security risks and thefts and know how to prevent or handle these situations.
2. Observe precautions required to protect hotel and guest property and report damage, theft, and found articles to supervisors.
3. Observe color of products being baked, and adjust oven temperatures, humidity, or conveyor speeds accordingly.
4. Inspect vehicles for damage and record findings so that necessary repairs can be made.
5. Inspect and adjust rental items to meet needs of customer.
6. Monitor and authorize entrance and departure of employees, visitors, and other persons to guard against theft and maintain security of premises.
7. Inspect loads to ensure that cargo is secure.
8. Check the quality of raw materials to ensure that standards and specifications are met.
9. Examine shipment contents and compare with records, such as manifests, invoices, or orders, to verify accuracy.
10. Inspect and clean food preparation areas, such as equipment, work surfaces, and serving areas, to ensure safe and sanitary food-handling practices.

Note: This graph looks at a gas station's weighted variety automation rate along the adoption path. Gray, vertical lines denote when a task has achieved more than 50% base-weighted variety automation. The task list contains 10 tasks (each from one visual task category as defined in Figure 4), automated as computing costs decline and signified by the darker, numbered vertical lines.

of firms for which more than 10% of vision-task varieties are automated. To measure the intensive margin, Figure 5 also reports the cost-weighted average share of automated task varieties across all tasks and firms in the economy, $\mathbb{E}_j[\mathbb{E}_{\tau|j}[\varphi_{K,ij\tau}]]$. Because the 659 vision tasks account for only a small share of all production tasks, this task-level adoption measure only reaches 6.4% over the projected automation path. Throughout the transition, employment-weighted adoption remains above unweighted adoption, indicating that larger, more productive firms adopt computer vision AI earlier and more extensively.[23,24]

Our quantitative framework allows us to project the outcomes of specific firms along the automation path, including their AI expenditure and firm-level labor demand elasticities of substitution. To illustrate the rich predictions of our quantitative framework, we next focus on the outcomes of a firm in the industry "Gasoline Stations" (NAICS 4571).

Figure 6 shows the model-implied automation path for a representative firm in this industry.[25] The path mirrors the aggregate pattern in Figure 5, with automation expanding as compute costs decline. Of the 44 tasks in this industry that can potentially be automated using computer vision, the first to reach 50% automation does so in 2025: "examin[ing] products purchased for resale or received for storage to assess the condition of each product or item," which has an error tolerance of 9.6% and a complexity of 5 classification outcomes. The last task automated in the projection reaches this threshold in 2069: "date-stamp[ing], sort[ing], and rack[ing] incoming mail and messages," with an error tolerance of 5.3% and a complexity of 11.6 outcomes.

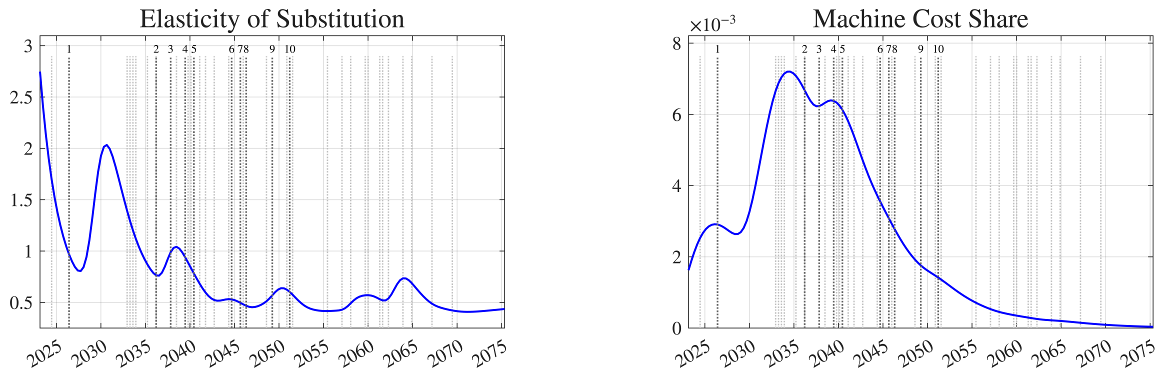
The left panel of Figure 7 displays the evolution of the firm-level elasticity of substitution, computed based on Equation (21), along the automation path. Consistent with our theoretical results in Proposition 4, this elasticity is non-monotonic, reflecting episodes of heightened capital–labor substitutability when new task varieties become automatable as compute prices fall. The right panel shows the machine share of costs for the firm in this industry. While also showing non-monotonicities, this share rises sharply when the initial tasks get automated, then gradually declines along the automation path.

[23] Automation results with alternative decline of 74.0% (hardware + algorithmic) can be found in Figure A10.

[24] Figure A9 reports similar results for the share of firms for which more than 1% of vision tasks are automated.

[25] 18.2% of labor in this industry is employed by firms smaller than the one illustrated.

Figure 7: Elasticity of Substitution and Machine Cost Shares for a Gas Station (NAICS 4571)



Note: The left panel of the figure examines the substitutability of labor to AI as the price of compute declines, where substitutability is defined as the firm-level elasticity of substitution between capital and labor. Higher values indicate greater substitutability. The right panel calculates the AI share of firm costs. Elasticities are calculated at the projected price of compute between years 2023 and 2075.

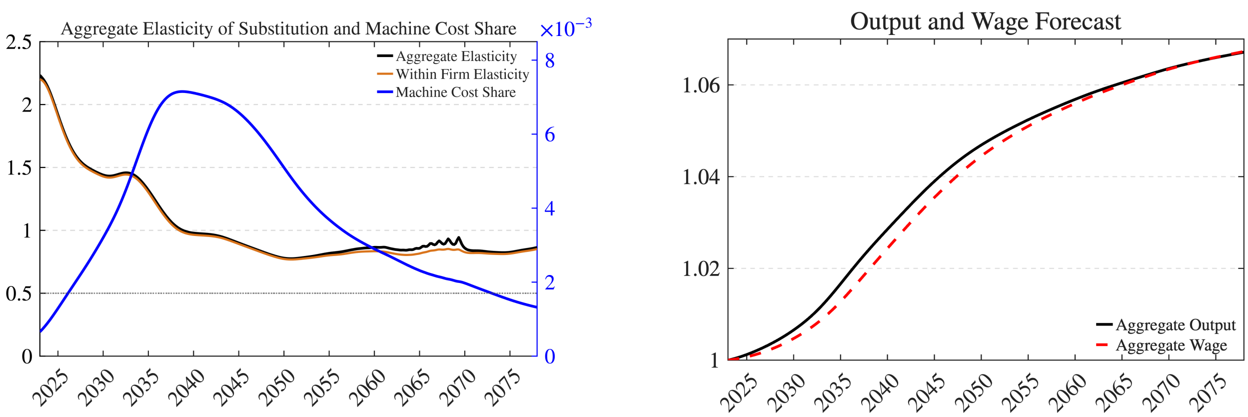
4.4 Aggregate Outcomes with Declining Compute Prices

Figure 8 shows how aggregate outcomes evolve along the automation path. The left panel shows the aggregate elasticity of substitution between AI and labor, defined in Equation (A15), and the aggregate AI cost share, defined in Equation (A11), as compute prices decline. The elasticity provides a short-run measure of workers' exposure to automation. It consistently exceeds the parameter λ, which accounts for the intensive margin of substitutability, reflecting the extensive margins of automation present in the aggregate economy. The figure also decomposes the aggregate elasticity into its components. Since there is little variation across industries in vision-task shares, the within-firm component (the first term in Equation (A15)) accounts for almost all of the aggregate elasticity.

As compute prices fall, substitutability between labor and computer vision AI declines, though non-monotonically. This decline reflects selection along the automation path: as AI becomes cheaper, the remaining non-automated workers are increasingly difficult to replace because they are relatively productive. More importantly, since the remaining tasks cannot be automated using vision, the scope for further extensive-margin automation narrows. Reflecting the same pattern, the aggregate AI cost share follows an inverse U-shape, increasing as AI and labor are gross substitutes, then declining as the substitution elasticity falls below unity and the two factors become gross complements. Correspondingly, this pattern implies a *U-shaped evolution for the labor share*.

The right panel of Figure 8 considers the evolution of aggregate real output over the adoption path relative to the 2023 baseline. Gains in real output initially rise around 2030 but then slow down as computer vision AI automation matures around 2040. In the

Figure 8: Elasticity of Substitution, Aggregate Output, and Wages on the Automation Path



Note: The left panel of the figure examines the substitutability of labor to AI as the price of compute declines, where substitutability is defined as the aggregate elasticity of substitution between capital and labor, and the aggregate AI cost share. Higher values indicate greater substitutability. The right panel shows aggregate output and the wage in proportion to their 2023 levels. Elasticities are calculated at the projected price of compute between years 2023 and 2075.

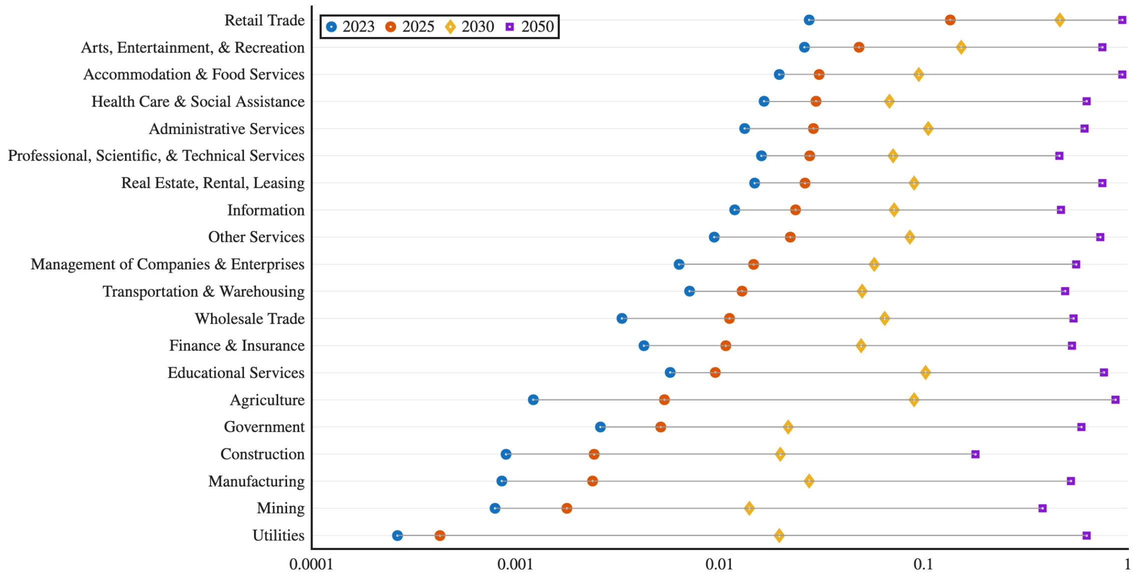
ten-year period between 2025 and 2035, real output increases from 0.1% to 1.7% beyond the 2023 baseline level, corresponding to a 1.6 percentage point rise (or 16 basis points contribution to aggregate growth annually). However, in the ten-year period between 2065 and 2075, real output increases only from 6.1% to 6.6% beyond the 2023 baseline level, contributing only 0.5 percentage points to growth.

Finally, we consider heterogeneity in the share of automated task varieties by industry in Figure 9. The top three sectors with the highest automation rates in 2023 are "Retail Trade" (NAICS 44–45), "Arts, Entertainment, and Recreation" (NAICS 71), and "Accommodation and Food Services" (NAICS 72) with cost-weighted firm-level variety automation rates of 13.5%, 4.8%, and 3.1%. We note that the rankings of most-automated sectors by computer vision change substantially over the automation path. For example, by 2050, the sector "Agriculture" (NAICS 11) is the third-most automated at 86.8%, despite being one of the least automated sectors in 2023.

4.5 Automation with Industry-Level Fixed Costs

We now consider the alternative case, introduced in Section 4.1, in which fine-tuning costs are incurred once per task at the industry level by a representative firm, rather than separately by each firm. Figure A8 in the Appendix reports the resulting outcome paths. Automation proceeds faster along the path than in the baseline, since the scale relevant for covering fixed costs is now industry rather than firm output. The aggregate elasticity of substitution between labor and AI again transitions from gross substitutes to gross

Figure 9: Sector-Level AI Automation Rates



Note: The figure displays average cost-weighted automated variety shares in 4-digit NAICS industries by 2-digit NAICS sectors. The automated variety shares are calculated at the projected price of compute in years 2023, 2025, 2030, and 2050.

complements as automation deepens, although its non-monotonic movements are more pronounced than in the baseline. The paths of aggregate output and wages remain very similar to the baseline. Overall, our aggregate results are robust to the level at which fixed costs are borne, while the pace of automation depends on the scale over which training costs are shared.

5 Conclusion

This paper develops a task-based theory of automation in which adopting machines entails task-level fixed costs. Fixed costs make machine comparative advantage scale-dependent: even when machines are cheap at the margin, automation depends on whether task scale is large enough to amortize the fixed cost. The canonical model appears as the limiting case in which fixed costs vanish or task scale becomes large.

The paper also proposes a task-type representation that bridges canonical theories of automation to their quantitative implementations. Using this representation, we can derive the endogenous elasticities of substitution between factors and output elasticities of factor demand. We use the representation to study computer vision automation, relying on AI scaling laws to discipline AI's marginal (inference) and fixed (training) compute costs across tasks. In our application, scale explains the majority of cross-task variation in comparative advantage, and its importance remains stable when projecting the economy forward along a path of declining compute prices. Along this path, the aggregate labor–AI elasticity of substitution declines from above to below one, producing a U-shaped labor share even as output and wages rise. We also demonstrate that scale remains important when assuming that fixed costs are sunk at the level of the industry, instead of the firm. Future work can examine how intermediate cases—where, for instance, firms outsource the fine-tuning process to upstream providers—may lead to different patterns of automation.

We emphasize that the theoretical results of the paper generalize to a variety of contexts beyond computer vision AI where fixed costs may shape comparative advantage, including other automation domains and applications in outsourcing, offshoring, and commodity trade. Our task-type representation can likewise quantify task-based models of automation for other technologies, such as LLMs or robotics—even ones where fixed costs play a minor role—provided credible task-level measures of their cost structure are available.

References

Acemoglu, Daron, "The simple macroeconomics of AI," *Economic Policy*, 2025, *40* (121), 13–58. 5

_ **and David Autor**, "Skills, tasks and technologies: Implications for employment and earnings," in "Handbook of labor economics," Vol. 4, Elsevier, 2011, pp. 1043–1171. 3, 4, 5

_ **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, 16

_ **and** _ , "The race between man and machine: Implications of technology for growth, factor shares, and employment," *American Economic Review*, 2018, *108* (6), 1488–1542. 3, 4, 5, 9, 11

_ **and** _ , "Automation and new tasks: How technology displaces and reinstates labor," *Journal of Economic Perspectives*, 2019, *33* (2), 3–30. 3, 4, 5, 9

— **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, 16

— **and** — , "Tasks, automation, and the rise in US wage inequality," *Econometrica*, 2022, *90* (5), 1973–2016. 9, 11

— **, David Autor, Jonathon Hazell, and Pascual Restrepo**, "Artificial intelligence and jobs: Evidence from online vacancies," *Journal of Labor Economics*, 2022, *40* (S1), S293–S340. 6

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. 6

Aghion, Philippe, Benjamin F. Jones, and Charles I. Jones, "Artificial Intelligence and Economic Growth," Working Paper 23928, National Bureau of Economic Research 2017. 11

Althoff, Lukas and Hugo Reichardt, "Task-Specific Technical Change and Comparative Advantage," 2026. CESifo Working Paper No. 12403, January 13, 2026. 6

Altman, Sam, "Planning for AGI and beyond," OpenAI 2023. 6

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. 3, 5

Babina, Tania, Anastassia Fedyk, Alex X He, and James Hodson, "Firm Investments in Artificial Intelligence Technologies and Changes in Workforce Composition," Working Paper 31325, National Bureau of Economic Research 2023. 6

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. 21

Berlingieri, Giuseppe, Filippo Boeri, Danial Lashkari, and Jonathan Vogel, "Capital-Skill Complementarity in Firms and in the Aggregate Economy," Working Paper 33000, National Bureau of Economic Research 2024. 29

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, 6, 29

Broda, Christian and David E Weinstein, "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, 2006, *121* (2), 541–585. 29

Brynjolfsson, Erik and Tom Mitchell, "What can machine learning do? Workforce implications," *Science*, 2017, *358* (6370), 1530–1534. 2, 5

_ , Danielle Li, and Lindsey R. Raymond, "Generative AI at Work," *The Quarterly Journal of Economics*, 2025, *140* (2), 889–942. 5

_ , Tom Mitchell, and Daniel Rock, "What can machines learn and what does it mean for occupations and the economy?," in "AEA Papers and Proceedings," Vol. 108 2018, pp. 43–47. 5

Burstein, Ariel, Eduardo Morales, and Jonathan Vogel, "Changes in Between-Group Inequality: Computers, Occupations, and International Trade," *American Economic Journal: Macroeconomics*, 2019, *11* (2), 348–400. 29

Cazzaniga, Mauro, Florence Jaumotte, Longji Li, Giovanni Melina, Augustus J Panton, Carlo Pizzinelli, Emma J Rockall, and Marina Mendes Tavares, "Gen-AI: Artificial Intelligence and the Future of Work," Staff Discussion Notes 2024/001, International Monetary Fund 2024. 6

CBS News, "Walmart's technology drive: AI and cameras can now mind the store," CBS News 2019. 2

Cirillo, Valeria, Rinaldo Evangelista, Dario Guarascio, and Matteo Sostero, "Digitalization, Routineness and Employment: An Exploration on Italian Task-Based Data," *Research Policy*, 2021, *50* (7), 104079. 15

Cui, Zheyuan Kevin, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz, "The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers," 2024. Available at SSRN 4945566. 5

De la Rica, Sara, Lucas Gortazar, and Piotr Lewandowski, "Job Tasks and Wages in Developed Countries: Evidence from PIAAC," *Labour Economics*, 2020, *65*, 101845. 15

Dehghani, Mostafa, Josip Djolonga, Basil Mustafa, Piotr Padlewski, Jonathan Heek et al., "Scaling Vision Transformers to 22 Billion Parameters," in "Proceedings of the

40th International Conference on Machine Learning," Vol. 202 of *Proceedings of Machine Learning Research* PMLR 2023, pp. 7480–7512. 2

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," Working Paper 24-013, Harvard Business School September 2023. 5

Dengler, Katharina and Britta Matthes, "The Impacts of Digital Transformation on the Labour Market: Substitution Potentials of Occupations in Germany," *Technological Forecasting and Social Change*, 2018, *137*, 304–316. 15

Dornbusch, Rudiger, Stanley Fischer, and Paul Anthony Samuelson, "Comparative advantage, trade, and payments in a Ricardian model with a continuum of goods," *The American Economic Review*, 1977, *67* (5), 823–839. 5

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs are GPTs: Labor market impact potential of LLMs," *Science*, 2024, *384* (6702), 1306–1308. 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. 5

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 2018, pp. 54–57. 2, 5

Fleming, Martin, Wensu Li, and Neil C. Thompson, "The Last Mile Problem in AI," Brookings 2024. 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. 5

Freund, Lukas B. and Lukas F. Mann, "Job Transformation, Specialization, and the Labor Market Effects of AI," 2026. Working paper. This version: March 18, 2026 (first version: August 16, 2025). 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

Grossman, Gene M and Esteban Rossi-Hansberg, "Trading tasks: A simple theory of offshoring," *American Economic Review*, 2008, *98* (5), 1978–1997. 5

Hampole, Menaka, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller, "Artificial intelligence and the labor market," Technical Report, National Bureau of Economic Research 2025. 6

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. 6

Hernandez, Danny, Jared Kaplan, Tom Henighan, and Sam McCandlish, "Scaling laws for transfer," *arXiv preprint arXiv:2102.01293*, 2021. 24

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. 21

Ho, Anson, Tamay Besiroglu, Ege Erdil, David Owen, Robi Rahman, Zifan Carl Guo, David Atkinson, Neil Thompson, and Jaime Sevilla, "Algorithmic Progress in Language Models," *arXiv preprint arXiv:2403.05812*, 2024. 29, 30

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. 2, 22

Huang, Yueling, "The Labor Market Impact of Artificial Intelligence: Local vs. Aggregate Effect," 2023. Working Paper. 6

Humlum, Anders, "Robot Adoption and Labor Market Dynamics," 2019. Princeton University working paper. 5, 29

Jones, Charles I., "Intermediate Goods and Weak Links in the Theory of Economic Development," *American Economic Journal: Macroeconomics*, 2011, *3* (2), 1–28. 7

— **and Christopher Tonetti**, "Past Automation and Future A.I.: Weak Links Tame the Growth Expansion," 2026. Working paper, May 2026. 7, 9

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. 2, 21, 22

Korinek, Anton and Donghyun Suh, "Scenarios for the Transition to AGI," Technical Report, National Bureau of Economic Research 2024. 5

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 256, OECD Publishing, Paris 2021. 5

Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard, "Information technology and returns to scale," *American Economic Review*, 2024, *114* (6), 1769–1815. 4, 20, 28, A4

Lewandowski, Piotr, Albert Park, Wojciech Hardy, Yang Du, and Saier Wu, "Technology, Skills, and Globalization: Explaining International Differences in Routine and Nonroutine Work Using Survey Data," *World Bank Economic Review*, 2022, *36* (3), 687–708. 15

Lindenlaub, Ilse, Ryungha Oh, Maria Alejandra Rodriguez, and Laura Veldkamp, "Beyond Exposure: Predicting AI Adoption Based on Comparative Advantage," Working Paper 35271, National Bureau of Economic Research 2026. 6

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. 24

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. 6

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer, "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot," *arXiv preprint arXiv:2302.06590*, 2023. 5

Rosenfeld, Jonathan S, "Scaling laws for deep learning," *arXiv preprint arXiv:2108.07686*, 2021. 21

Sozzo, Emanuele Del, Martin Fleming, Kenneth Flamm, and Neil Thompson, "How much progress has there been for NVIDIA GPUs?," 2025. Working Paper. 29, 30

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, 25, 26, A18

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

Zeira, Joseph, "Workers, machines, and economic growth," *The Quarterly Journal of Economics*, 1998, *113* (4), 1091–1117. 5

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, 22

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. 24

Appendix

A Theoretical Appendix

A.1 Specializing the Section 2.2 Results to the Quantitative Model

This subsection records the few objects from Section 2.2 whose notation changes when the task-type representation is specialized to the quantitative model of Section 2.3. The quantitative model replaces the type measure $d\mu(\boldsymbol{\tau})$ by the counting measure over task items $\tau \in \mathcal{T}$, so integrals over task types become sums over $\mathcal{T}$.

The firm-level marginal cost expression in Equation (14) becomes $c^{1-\lambda} = \sum_{\tau \in \mathcal{T}} \xi_\tau c_\tau^{1-\lambda}$. After substituting the task item marginal costs, this is $c^{1-\lambda} = \sum_{\tau \in \mathcal{T}} \xi_\tau [\varphi_{K,\tau}(r/z_{K,\tau})^{1-\lambda} + \varphi_{L,\tau}(w/\overline{Z}_{L,\tau})^{1-\lambda}]$. The remaining equality in Equation (14), $c^{1-\lambda} = \varphi_K r^{1-\lambda} + \varphi_L w^{1-\lambda}$, follows from the definitions of φ_K and φ_L in Section 2.3. Likewise, Equation (16) holds with discrete variable-cost weights $\pi_\tau \equiv \xi_\tau (c_\tau/c)^{1-\lambda}$, so that $\mathbb{E}_\pi[x_\tau] = \sum_{\tau \in \mathcal{T}} \pi_\tau x_\tau$. In particular, $\omega_K = \sum_{\tau \in \mathcal{T}} \pi_\tau \omega_{K,\tau} = \varphi_K (r/c)^{1-\lambda}$ and $\omega_L = \sum_{\tau \in \mathcal{T}} \pi_\tau \omega_{L,\tau} = \varphi_L (w/c)^{1-\lambda}$.

Define the fixed-cost-weighted task distribution by $\pi_{\tau,f} \equiv (1/\omega^f)[r/(c(y/a))]\varphi_{K,\tau}/z^f_{K,\tau}$, where $\omega^f \equiv c^f/[c(y/a)]$. Using Equation (26) to evaluate $\widetilde{G}(z^*_\tau|\tau)$, the definition of Θ in Equation (19) becomes

$$\Theta = \frac{\frac{\lambda}{1-\lambda}\omega^f \sum_{\tau \in \mathcal{T}} \pi_{\tau,f} \zeta_\tau \theta \widehat{z}_\tau^{-\theta}}{1 + \frac{\lambda}{1-\lambda}\omega^f \sum_{\tau \in \mathcal{T}} \pi_{\tau,f} \zeta_\tau \theta \widehat{z}_\tau^{-\theta}}, \qquad \zeta_\tau \equiv 1 - SA_\tau^{1-\lambda}. \tag{A1}$$

Equation (20) specializes to

$$\Omega_K^m = \Theta + (1-\Theta)\left(\omega_K + \omega^f \sum_{\tau \in \mathcal{T}} \pi_{\tau,f} \theta \widehat{z}_\tau^{-\theta}\right), \qquad \Omega_L^m = 1 - \Omega_K^m. \tag{A2}$$

Finally, with labor-cost weights $\pi_{L,\tau} \equiv \omega_L^{-1} \xi_\tau \varphi_{L,\tau}[(w/\overline{Z}_{L,\tau})/c]^{1-\lambda}$, Equation (21) becomes

$$\begin{aligned} \sigma = \lambda &+ \frac{1}{\Omega_K} \sum_{\tau \in \mathcal{T}} \pi_{L,\tau} \widetilde{G}_{\lambda-1}(z^*_\tau|\tau) \\ &+ \frac{\Omega_K^m - \Omega_K}{\Omega_K}\lambda + \frac{1-\Omega_K^m}{\Omega_K}\frac{\lambda}{1-\lambda} \sum_{\tau \in \mathcal{T}} \pi_{L,\tau} \zeta_\tau \widetilde{G}_{\lambda-1}(z^*_\tau|\tau), \end{aligned} \tag{A3}$$

where $\widetilde{G}_{\lambda-1}(z^*_\tau|\tau)$ is evaluated using Equation (26). All other equations from Section 2.2 carry over unchanged, apart from replacing integrals over $\boldsymbol{\tau}$ by sums over $\tau \in \mathcal{T}$ and using the Section 2.3 definitions of the task item objects.

Proof of Equations (25) and (26). Let $\widehat{z}_\tau = z^*_\tau / Z_\tau$. The automation share in Equation (25) follows directly from evaluating the Fréchet c.d.f. at z^*_τ. For the labor shifter, start from the definition of the selection function,

$$\varphi_{L,\tau} = \int_{z^*_\tau}^{\infty} \left(\frac{z}{\overline{Z}_{L,\tau}} \right)^{\lambda-1} dG(z|\tau).$$

Use the change of variables $t = (z/Z_\tau)^{-\theta}$. Then $z \in [z^*_\tau, \infty)$ corresponds to $t \in [\widehat{z}_\tau^{-\theta}, 0]$ and $dG = -e^{-t}dt$, so

$$\int_{z^*_\tau}^{\infty} z^{\lambda-1} dG(z|\tau) = Z_\tau^{\lambda-1} \int_0^{\widehat{z}_\tau^{-\theta}} t^{(1-\lambda)/\theta} e^{-t} dt = Z_\tau^{\lambda-1} \gamma \left(1 + \frac{1-\lambda}{\theta}, \widehat{z}_\tau^{-\theta} \right).$$

Since $\overline{Z}_{L,\tau}^{\lambda-1} = Z_\tau^{\lambda-1} \Gamma(1 + (1-\lambda)/\theta)$, division by $\overline{Z}_{L,\tau}^{\lambda-1}$ gives the expression for $\varphi_{L,\tau}$ in Equation (25).

For Equation (26), the first elasticity follows by differentiating $\log G(z^*_\tau|\tau) = -\widehat{z}_\tau^{-\theta}$ with respect to $\log z^*_\tau$, which also gives the equivalent expression in terms of $\log \varphi_{K,\tau}$. For the selection elasticity, Equation (25) implies that $G_{\lambda-1}(z^*_\tau|\tau)$ is proportional to $\gamma(1 + (1-\lambda)/\theta, \widehat{z}_\tau^{-\theta})$. Therefore, using the definition of $\widetilde{\gamma}$ in Equation (26),

$$\widetilde{G}_{\lambda-1}(z^*_\tau|\tau) \equiv -\frac{\partial \log G_{\lambda-1}(z^*_\tau|\tau)}{\partial \log z^*_\tau} = \theta\, \widetilde{\gamma} \left(1 + \frac{1-\lambda}{\theta}, \widehat{z}_\tau^{-\theta} \right).$$

Expanding the definition of $\widetilde{\gamma}$ gives the closed-form selection-elasticity expression in Equation (26).

A.2 General Equilibrium

The demands for firm-level and industry-level products are given by $y_j = p_j^{-\eta} P_i^{\eta} Y_i$ and $Y_i = P_i^{-\rho} P^{\rho} Y$ where p_j, P_i, and P are the firm-, industry-, and aggregate-level CES price indices, respectively. We assume that firms ignore their impact on the industry-level price index so that the product market is characterized by monopolistic competition.

Monopolistic competition implies that firms set prices at a constant markup over marginal costs $p_j = \frac{\eta}{\eta-1} c_j / a_j$, where the productivity-adjusted marginal cost c_j of the firm j is given by Equation (14). Aggregating across firms within industry i, the price index can be written as $P_i = \frac{\eta}{\eta-1} J_i^{\frac{1}{1-\eta}} \overline{c}_i$ where the average industry-level marginal cost $\overline{c}_i$

is given by

$$\bar{c}_i^{1-\eta} = \frac{1}{J_i} \sum_{j \in \mathcal{J}_i} \left(\frac{c_j}{a_j} \right)^{1-\eta}. \tag{A4}$$

From the expression for firm-level demand, we write the scale of firm j's output and its share in the industry-level variable costs as

$$y_j = J_i^{\frac{\eta}{1-\eta}} \left(\frac{c_j/a_j}{\bar{c}_i} \right)^{-\eta} Y_i, \qquad \Psi_{j,i} \equiv \frac{c_j y_j / a_j}{J_i^{\frac{1}{1-\eta}} \bar{c}_i Y_i} = \frac{1}{J_i} \left(\frac{c_j/a_j}{\bar{c}_i} \right)^{1-\eta}, \tag{A5}$$

where $J_i \equiv |\mathcal{J}_i|$ and $I \equiv |\mathcal{I}|$ denote 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

$$\bar{c}^{1-\rho} = \sum_i J_i^{\frac{1-\rho}{1-\eta}} \bar{c}_i^{1-\rho}. \tag{A6}$$

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

$$Y_i = J_i^{\frac{\rho}{\eta-1}} \left(\frac{\bar{c}_i}{\bar{c}} \right)^{-\rho} Y, \qquad \Psi_i \equiv \frac{J_i^{\frac{1}{1-\eta}} \bar{c}_i Y_i}{\bar{c}\, Y} = J_i^{\frac{1-\rho}{1-\eta}} \left(\frac{\bar{c}_i}{\bar{c}} \right)^{1-\rho}. \tag{A7}$$

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 (14). Aggregating across firms in each industry gives the following expression for industrial profits

$$\overline{\Pi}_i = \frac{1}{\eta - 1} \Psi_i\, \bar{c}\, Y - \overline{C}_i^f, \tag{A8}$$

where industry i's expenditure on fixed automation costs is $\overline{C}_i^f \equiv \sum_{j \in \mathcal{J}_i} c_j^f$.

We can express the share of labor and machines in aggregate variable costs as

$$\bar{\omega}_L \equiv \frac{w\overline{L}}{\bar{c}\, Y} = \sum_{i,j} \Psi_i \Psi_{j,i}\, \omega_{L,j} \quad \text{and} \quad \bar{\omega}_K \equiv \frac{r\overline{K} - \overline{C}^f}{\bar{c}\, Y} = \sum_{i,j} \Psi_i \Psi_{j,i}\, \omega_{K,j}, \tag{A9}$$

where we have defined the aggregate machine demand for fixed automation costs as $\overline{C}^f \equiv \sum_i \overline{C}_i^f$. Equation (A9) clears the factor markets and determines the equilibrium wage w, the rental price of machines r, and the scale of output Y. Under monopolistic competition, the market equilibrium is efficient.

A.3 Aggregate Elasticities

Aggregate Elasticities of Total Cost Let $\overline{C}(w,r;Y) = \bar{c}Y + \overline{C}^f$ denote total costs in the aggregate economy. Aggregate total cost elasticities maintain the same form as the micro-elasticity but use aggregate instead of firm-level factor shares. The aggregate scale elasticity of total costs is given by

$$\bar{\varepsilon} \equiv \frac{\partial \log \overline{C}}{\partial \log Y} = \frac{1}{1+\overline{\omega}^f}, \tag{A10}$$

where $\overline{\omega}^f = \sum_{i,j} \Psi_i \Psi_{j,i} \omega_j^f$ denotes the aggregate fixed to variable cost ratio. This ratio also impacts the aggregate elasticity of total costs with respect to machine prices following

$$\overline{\Omega}_K \equiv \frac{\partial \log \overline{C}}{\partial \log r} = \frac{\overline{\omega}_K + \overline{\omega}^f}{1+\overline{\omega}^f}. \tag{A11}$$

Aggregate Scale Elasticities of Marginal Cost and Factor Demand We define $\bar{\varepsilon}_i^m$ and $\bar{\varepsilon}^m$ as the scale elasticities of marginal costs for industry i and for the aggregate economy. The scale elasticity of marginal costs for industry i is

$$\bar{\varepsilon}_i^m \equiv \frac{\partial \log \bar{c}_i}{\partial \log Y_i} = \frac{1}{\eta}\left(\mathbb{E}_{j|i}\left[\left(1+\eta\, \varepsilon_j^m\right)^{-1}\right]^{-1} - 1\right), \tag{A12}$$

where the operator $\mathbb{E}_{j|i}[\cdot]$ denotes the cost-weighted average within industries under distribution $\Psi_{j,i}$ in Equation (A5). Likewise, the same elasticity at the aggregate level is

$$\bar{\varepsilon}^m \equiv \frac{\partial \log \bar{c}}{\partial \log Y} = \frac{1}{\rho}\left(\mathbb{E}_i\left[(1+\rho\, \bar{\varepsilon}_i^m)^{-1}\right]^{-1} - 1\right), \tag{A13}$$

where the operator $\mathbb{E}_i[\cdot]$ denotes the cost-weighted average across industries under distribution Ψ_i in Equation (A7).

Following Lashkari et al. (2024), we define the pass-through distribution of industry-to-firm-level output $\Psi_{j,i}^p$ and aggregate-to-industry-level output Ψ_i^p, as $\Psi_{j,i}^p \equiv \Psi_{j,i}\frac{1+\eta\bar{\varepsilon}_i^m}{1+\eta\varepsilon_j^m}$ and $\Psi_i^p \equiv \Psi_i \frac{1+\rho\bar{\varepsilon}^m}{1+\rho\bar{\varepsilon}_i^m}$. The two distributions aggregate the elasticities of firm-level marginal costs with respect to factor prices, which by Equation (22) satisfy $\Omega_{K,j}^m = \varepsilon_j^{-1}\Omega_{K,j}\eta_{K,j} = (\omega_{K,j} + \omega_j^f)\,\eta_{K,j}$ and $\Omega_{L,j}^m = \omega_{L,j}\eta_{L,j}$; note that fixed machine requirements enter the machine-side object. We define $\overline{\Omega}_{f,i}^m \equiv \mathbb{E}_{j|i}^p[\Omega_{f,j}^m]$ and $\overline{\Omega}_f^m \equiv \mathbb{E}_i^p[\overline{\Omega}_{f,i}^m]$ for $f \in \{K, L\}$. The scale elasticities of aggregate factor demand then follow from the aggregate counterpart of Equation (22):

$$\overline{\eta}_K \equiv \frac{\partial \log K}{\partial \log Y} = \overline{\varepsilon}\,\frac{\overline{\Omega}_K^m}{\overline{\Omega}_K}, \quad \overline{\eta}_L \equiv \frac{\partial \log L}{\partial \log Y} = \frac{1}{1-\overline{\Omega}_K}\left(\overline{\varepsilon} - \overline{\Omega}_K \overline{\eta}_K\right). \tag{A14}$$

Aggregate Substitution Elasticities of Demand Next, we characterize the aggregate elasticity of labor and machines as

$$\overline{\sigma} \equiv \frac{1}{\overline{\Omega}_K}\frac{\partial \log L}{\partial \log r} = \mathbb{E}_j^c\left[\frac{\Omega_{L,j}}{\overline{\Omega}_L}\frac{\Omega_{K,j}}{\overline{\Omega}_K}\sigma_j\right] - \mathbb{E}_i\left[\frac{\eta\overline{\varepsilon}}{1+\eta\overline{\varepsilon}_i^m}\mathbb{C}_{j|i}^p\left(\frac{\Omega_{L,j}^m}{\overline{\Omega}_L}, \frac{\Omega_{K,j}^m}{\overline{\Omega}_K}\right)\right] - \frac{\rho\,\overline{\varepsilon}}{1+\rho\,\overline{\varepsilon}^m}\mathbb{C}_i^p\left(\frac{\overline{\Omega}_{L,i}^m}{\overline{\Omega}_L}, \frac{\overline{\Omega}_{K,i}^m}{\overline{\Omega}_K}\right), \tag{A15}$$

where $\mathbb{E}_j^c[\cdot]$ denotes the total-cost-weighted average across all firms under the distribution $\frac{\overline{\varepsilon}}{\varepsilon_j}\Psi_i\Psi_{j,i}$, and $\mathbb{C}_{j|i}^p(\cdot,\cdot)$ and $\mathbb{C}_i^p(\cdot,\cdot)$ denote the within-industry and across-industry covariances under the pass-through distributions $\Psi_{j,i}^p$ and Ψ_i^p, respectively. Machines *automate* labor if this elasticity is positive, indicating that a fall in the price of machines lowers the demand for labor.

Three additive terms of Equation (A15) can decompose the elasticity. The first term averages firm-level substitution elasticities of labor demand σ_j. The second reflects between-firm, within-industry production reallocation after a change in factor prices. The third reflects production reallocation between industries after industry-level cost adjustments.

Finally, just like at the micro level, the macro-level elasticities of marginal costs with respect to factor prices likely deviate from those of total costs. Consistent with the definitions above, the aggregates $\overline{\Omega}_K^m$ and $\overline{\Omega}_L^m$ satisfy

$$\overline{\Omega}_K^m = \frac{\partial \log \overline{c}}{\partial \log r} = \frac{\overline{\Omega}_K \overline{\eta}_K}{\overline{\varepsilon}} \quad \text{and} \quad \overline{\Omega}_L^m = \frac{\partial \log \overline{c}}{\partial \log w} = 1 - \overline{\Omega}_K^m. \tag{A16}$$

A.4 Proofs and Derivations

A.4.1 Proof of Proposition 1

We proceed in three steps. First, we show that the optimal input choice follows a threshold rule, where tasks for which labor productivity $z_L(\nu)$ is less than some threshold z_L^* are automated. Second, we derive this threshold. Third, we calculate firm marginal costs c.

We first fix input choices $\boldsymbol{x}$. From standard arguments, task-level output is $y(\nu) =$

$\xi(\nu)c(\nu)^{-\lambda}c^{\lambda}\frac{y}{a}$, and firm-level marginal costs are

$$c^{1-\lambda} = \int_{\Upsilon} \xi(\nu)c(\nu)^{1-\lambda}d\nu. \tag{A17}$$

Consider input choices in a cell with constant marginal and fixed machine productivity, as well as task importance, $(\boldsymbol{z_K}, \xi)$ where $\boldsymbol{z_K} = (z_K, z_K^f)$. The labor marginal costs $\frac{w}{z_L}$ strictly decrease in labor productivity z_L. The optimal rule is therefore monotone in labor productivity: if a task with a productivity z_L is performed by workers, every task with higher labor productivity is also performed by workers; likewise, if a task with productivity z_L is performed by machines, every task with lower labor productivity is also automated. Otherwise, violations of this rule would weakly increase marginal costs without impacting fixed costs. The optimal rule is thus characterized by a threshold $z_L^*(\boldsymbol{z_K}, \xi)$.

The problem of minimizing total costs in Equation (2) can be reduced to choosing this cutoff and can be re-written as

$$C(w,r;y) = c\frac{y}{a} + \int \left[\int_0^{z_L^*(\boldsymbol{z_K},\xi)} d\mu_L(z_L|\boldsymbol{z_K},\xi)\right] \frac{r}{z_K^f} d\mu(\boldsymbol{z_K},\xi), \tag{A18}$$

where $\mu_L(z_L|z_K,\xi)$ denotes the distribution of labor productivity z_L conditional on cell characteristics $(\boldsymbol{z}_K,\xi)$, $\mu(\boldsymbol{z}_K,\xi)$ denotes the distribution of cells, and marginal costs c are given by Equation (6).

Within a cell indexed by $(\boldsymbol{z_K},\xi)$, suppose that the cutoff $z_L^* = z_L^*(\boldsymbol{z_K},\xi)$ is interior. Consider a perturbation of $\delta > 0$ in this cutoff, which switches the input choice from labor to machines for tasks with labor productivity z_L in the interval $[z_L^*, z_L^*+\delta]$. Denote as Δx_δ the change in x arising from the perturbation δ in z_L^*. From Equation (6), the change in the marginal cost is

$$(1-\lambda)c^{-\lambda}\Delta c_\delta = \xi \int_{z_L^*}^{z_L^*+\delta} \left[\left(\frac{r}{z_K}\right)^{1-\lambda} - \left(\frac{w}{z_L}\right)^{1-\lambda}\right] d\mu_L(z_L|\boldsymbol{z_K},\xi).$$

The change in the fixed cost is $\Delta C_\delta^f = \int_{z_L^*}^{z_L^*+\delta} \frac{r}{z_K^f} d\mu_L(z_L|\boldsymbol{z_K},\xi)$. From the definition of a derivative, the sum $\Delta C_\delta = \frac{y}{a}\Delta c_\delta + \Delta C_\delta^f$, divided by the change in the measure of tasks switched in this perturbation, should equal zero as $\delta \to 0$, at the optimal cutoff z_L^*. This yields the equation

$$\frac{1}{1-\lambda}\xi\, c^{\lambda}\frac{y}{a}\left[\left(\frac{r}{z_K}\right)^{1-\lambda} - \left(\frac{w}{z_L^*}\right)^{1-\lambda}\right] + \frac{r}{z_K^f} = 0. \tag{A19}$$

We obtain the productivity cutoff z_L^* by rearranging this expression, as defined by Equations (5) and (9). The firm automates if $z_L \le z_L^*$, which generates Equation (4). Using the task-type representation, a way to rewrite this cutoff is

$$z_{\boldsymbol{\tau}}^* = \frac{w}{r} z_{K,\boldsymbol{\tau}} S A_{\boldsymbol{\tau}}. \tag{A20}$$

□

A.4.2 Proof of Proposition 2

Total cost decomposes as the sum of variable and fixed task-level costs, $C(w,r;y) = \int_{\Upsilon} c(\nu)\, y(\nu)\, d\nu + \int_{\Upsilon} c^f(\nu)\, d\nu$. The variable component is the cost-minimum of the CES task aggregator in Equation (1) at task marginal costs $\{c(\nu)\}$, which yields $c\,(y/a)$ with $c^{1-\lambda} = \int_{\Upsilon} \xi(\nu)\, c(\nu)^{1-\lambda}\, d\nu$. Aggregating to types via Equation (10) delivers the first equality in Equation (14). Inside the type integral, the terms involving machines and labor, respectively, are

$$\xi_{\boldsymbol{\tau}} \left(\frac{r}{z_{K,\boldsymbol{\tau}}}\right)^{1-\lambda} \varphi_{K,\boldsymbol{\tau}} = r^{1-\lambda} \xi_{\boldsymbol{\tau}} z_{K,\boldsymbol{\tau}}^{\lambda-1} \varphi_{K,\boldsymbol{\tau}}, \quad \xi_{\boldsymbol{\tau}} \left(\frac{w}{\overline{Z}_{L,\boldsymbol{\tau}}}\right)^{1-\lambda} \varphi_{L,\boldsymbol{\tau}} = w^{1-\lambda} \xi_{\boldsymbol{\tau}} \overline{Z}_{L,\boldsymbol{\tau}}^{\lambda-1} \varphi_{L,\boldsymbol{\tau}},$$

where the labor term uses $\int_{z_{\boldsymbol{\tau}}^*}^{\infty} z^{\lambda-1}\, dG(z|\boldsymbol{\tau}) = \overline{Z}_{L,\boldsymbol{\tau}}^{\lambda-1} \varphi_{L,\boldsymbol{\tau}}$ from Equation (8). Integrating over types and substituting the definitions in Equation (13) delivers the second equality, $c^{1-\lambda} = \varphi_K\, r^{1-\lambda} + \varphi_L\, w^{1-\lambda}$. Pulling r outside the integral in the fixed-cost expression of Equation (14) defines k^f and gives $c^f = r\, k^f$.

For the constant-elasticity output representation, equate the two expressions for each firm-level cost share in Equation (16). Solving for the conditional input demands yields $k - k^f = \varphi_K\,(c/r)^{\lambda}\,(y/a)$ and $\ell = \varphi_L\,(c/w)^{\lambda}\,(y/a)$. Conditional on the optimized automation set, the unit-cost index $c^{1-\lambda} = \varphi_K r^{1-\lambda} + \varphi_L w^{1-\lambda}$ is dual to a constant-elasticity aggregator in $k^v = k - k^f$ and $L = \ell$. Substituting the conditional demands into $\varphi_K^{\frac{1}{\lambda}}(k - k^f)^{\frac{\lambda-1}{\lambda}} + \varphi_L^{\frac{1}{\lambda}}\, \ell^{\frac{\lambda-1}{\lambda}}$ and using $c^{1-\lambda} = \varphi_K r^{1-\lambda} + \varphi_L w^{1-\lambda}$ yields the right-hand side $(y/a)^{\frac{\lambda-1}{\lambda}}$, which rearranges to Equation (15). □

A.4.3 Proof of Proposition 3

We solve for the marginal cost scale and factor price elasticities, respectively ε^m and Ω_K^m (and Ω_L^m), in two steps. First, we derive the elasticity of two key endogenous objects, scale importance $\zeta_{\boldsymbol{\tau}}$ and the labor-productivity cutoff $z_{\boldsymbol{\tau}}^*$, to scale and factor prices. Second, we use these intermediate elasticities to derive this proposition's main results.

We derive two helpful identities from Equation (A19). The first identity connects the fixed cost ratio to the machine share of marginal costs. The fixed cost to variable cost ratio at the task type level is $\omega^f_{\boldsymbol{\tau}} = \frac{r/z^f_{K,\boldsymbol{\tau}}}{c_{\boldsymbol{\tau}} y_{\boldsymbol{\tau}}} G(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau})$. We apply standard CES results that $y_{\boldsymbol{\tau}} = \xi c_{\boldsymbol{\tau}}^{-\lambda} c^{\lambda} \frac{y}{a}$ and use the c.d.f. elasticity $\widetilde{G}(x|\boldsymbol{\tau}) \equiv \frac{\partial \log G(x|\boldsymbol{\tau})}{\partial \log x} = \frac{x g(x|\boldsymbol{\tau})}{G(x|\boldsymbol{\tau})}$. We can rearrange Equation (A19) as

$$\left(\frac{w/z^*_{\boldsymbol{\tau}}}{r/z_{K,\boldsymbol{\tau}}}\right)^{1-\lambda} = 1 + (1-\lambda)\frac{z_{K,\boldsymbol{\tau}}}{z^f_{K,\boldsymbol{\tau}}}\left(\xi\left(\frac{r/z_{K,\boldsymbol{\tau}}}{c}\right)^{-\lambda}\frac{y}{a}\right)^{-1}.$$

Using Equation (A20) and (11), the rearranged Equation (A19) becomes

$$\omega^f_{\boldsymbol{\tau}} = \frac{\zeta_{\boldsymbol{\tau}}}{1-\zeta_{\boldsymbol{\tau}}}\frac{1}{1-\lambda}\omega_{K,\boldsymbol{\tau}}. \tag{A21}$$

To derive the second identity, follow

$$\begin{aligned}(1-\zeta_{\boldsymbol{\tau}})\widetilde{G}_{\lambda-1}(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau})\omega_{L,\boldsymbol{\tau}} &= S A_{\boldsymbol{\tau}}^{1-\lambda}\left(\frac{w/\overline{Z}_{L,\boldsymbol{\tau}}}{c_{\boldsymbol{\tau}}}\right)^{1-\lambda}(z^*_{\boldsymbol{\tau}})^{\lambda}\overline{Z}_{L,\boldsymbol{\tau}}^{1-\lambda} g(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau}) \\ &= \left(\frac{r}{w}\frac{z^*_{\boldsymbol{\tau}}}{z_{K,\boldsymbol{\tau}}}\frac{w/\overline{Z}_{L,\boldsymbol{\tau}}}{c_{\boldsymbol{\tau}}}\overline{Z}_{L,\boldsymbol{\tau}}\right)^{1-\lambda}(z^*_{\boldsymbol{\tau}})^{\lambda} g(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau}) \\ &= \left(\frac{r/z_{K,\boldsymbol{\tau}}}{c_{\boldsymbol{\tau}}}\right)^{1-\lambda} z^*_{\boldsymbol{\tau}} g(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau}) \\ &= \omega_{K,\boldsymbol{\tau}}\widetilde{G}(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau}),\end{aligned} \tag{A22}$$

where the first equality uses Equations (11) and (8), the second uses Equation (A20), the third combines terms, and the fourth applies Equation (11).

Next, we derive $\frac{\partial \log c_{\boldsymbol{\tau}}}{\partial \log z^*_{\boldsymbol{\tau}}}$ and $\frac{\partial \log c_{\boldsymbol{\tau}}}{\partial \log \zeta_{\boldsymbol{\tau}}}$. First, we rearrange Equation (A20) to get the identity $\frac{1}{1-\zeta_{\boldsymbol{\tau}}} = \left(\frac{w/\overline{Z}_{L,\boldsymbol{\tau}}}{r/z_{K,\boldsymbol{\tau}}}\right)^{1-\lambda} \widehat{z}_{\boldsymbol{\tau}}^{\lambda-1}$. Therefore, Equation (10) can be rewritten as

$$c_{\boldsymbol{\tau}}^{1-\lambda} = \left(\frac{w}{\overline{Z}_{L,\boldsymbol{\tau}}}\right)^{1-\lambda}\left[G_{\lambda-1}(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau}) + (1-\zeta_{\boldsymbol{\tau}})\widehat{z}_{\boldsymbol{\tau}}^{\lambda-1} G(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau})\right].$$

One can derive the following elasticity by taking the derivative of the rewritten equation with respect to $\log z^*_{\boldsymbol{\tau}}$. We find

$$\frac{\partial \log c_{\boldsymbol{\tau}}}{\partial \log z^*_{\boldsymbol{\tau}}} = -\frac{1}{1-\lambda}\zeta_{\boldsymbol{\tau}}\omega_{L,\boldsymbol{\tau}}\widetilde{G}_{\lambda-1}(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau}) - \omega_{K,\boldsymbol{\tau}},$$

which uses the rearranged Equation (A20) to simplify. Lastly, under the identity of Equation (A22), we get

$$\frac{\partial \log c_{\boldsymbol{\tau}}}{\partial \log z^*_{\boldsymbol{\tau}}} = -\omega_{K,\boldsymbol{\tau}} - \omega^f_{\boldsymbol{\tau}} \widetilde{G}(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau}). \tag{A23}$$

Consider the next intermediate elasticity

$$\begin{aligned}
\frac{\partial \log c_{\boldsymbol{\tau}}}{\partial \log \zeta_{\boldsymbol{\tau}}} &= -\frac{1}{1-\lambda}\left[\left(\frac{w}{\overline{Z}_{L,\boldsymbol{\tau}}\widehat{z}_{\boldsymbol{\tau}}}\right)^{1-\lambda} c_{\boldsymbol{\tau}}^{\lambda-1}\zeta_{\boldsymbol{\tau}} G(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau})\right] \\
&= -\frac{1}{1-\lambda}\left(\frac{r/z_{K,\boldsymbol{\tau}}}{c_{\boldsymbol{\tau}}}\right)^{1-\lambda}\frac{\zeta_{\boldsymbol{\tau}}}{1-\zeta_{\boldsymbol{\tau}}} G(z^*_{\boldsymbol{\tau}}|\boldsymbol{\tau}) \\
&= -\frac{1}{1-\lambda}\frac{\zeta_{\boldsymbol{\tau}}}{1-\zeta_{\boldsymbol{\tau}}}\omega_{K,\boldsymbol{\tau}},
\end{aligned}$$

where the second equality uses the identity from rearranging Equation (A20) and the third uses definitions from Equation (11). Using the identity Equation (A21) yields

$$\frac{\partial \log c_{\boldsymbol{\tau}}}{\partial \log \zeta_{\boldsymbol{\tau}}} = -\omega^f_{\boldsymbol{\tau}}. \tag{A24}$$

From Equation (A20), we find $\frac{\partial \log z^*_{\boldsymbol{\tau}}}{\partial \log y} = \frac{\partial \log SA_{\boldsymbol{\tau}}}{\partial \log y}$ and $\frac{\partial \log z^*_{\boldsymbol{\tau}}}{\partial \log r} = -1 + \frac{\partial \log SA_{\boldsymbol{\tau}}}{\partial \log r}$. Examining $SA_{\boldsymbol{\tau}}$, we find the following derivatives:

$$\frac{\partial \log(SA_{\boldsymbol{\tau}}^{\lambda-1}-1)}{\partial \log y} = -1 - \lambda\frac{\partial \log c}{\partial \log y}, \qquad \frac{\partial \log(SA_{\boldsymbol{\tau}}^{\lambda-1}-1)}{\partial \log r} = \lambda\left(1 - \frac{\partial \log c}{\partial \log r}\right).$$

After solving, these equations yield $\frac{\partial \log SA_{\boldsymbol{\tau}}}{\partial \log y}$ and $\frac{\partial \log SA_{\boldsymbol{\tau}}}{\partial \log r}$. We then obtain the intermediate elasticities

$$\frac{\partial \log z^*_{\boldsymbol{\tau}}}{\partial \log y} = \frac{\zeta_{\boldsymbol{\tau}}}{1-\lambda}\left(1 + \lambda\frac{\partial \log c}{\partial \log y}\right), \qquad \frac{\partial \log z^*_{\boldsymbol{\tau}}}{\partial \log r} = -\left[1 + \frac{\lambda}{1-\lambda}\zeta_{\boldsymbol{\tau}}\left(1 - \frac{\partial \log c}{\partial \log r}\right)\right]. \tag{A25}$$

Using a similar approach and the definition of $\zeta_{\boldsymbol{\tau}}$, we know $SA_{\boldsymbol{\tau}}^{\lambda-1} - 1 = \frac{\zeta_{\boldsymbol{\tau}}}{1-\zeta_{\boldsymbol{\tau}}}$. Then we find

$$\frac{\partial \log\left(\frac{\zeta_{\boldsymbol{\tau}}}{1-\zeta_{\boldsymbol{\tau}}}\right)}{\partial \log y} = -1 - \lambda\frac{\partial \log c}{\partial \log y}, \qquad \frac{\partial \log\left(\frac{\zeta_{\boldsymbol{\tau}}}{1-\zeta_{\boldsymbol{\tau}}}\right)}{\partial \log r} = \lambda\left(1 - \frac{\partial \log c}{\partial \log r}\right).$$

Rearranging yields

$$\frac{\partial \log \zeta_{\boldsymbol{\tau}}}{\partial \log y} = -(1-\zeta_{\boldsymbol{\tau}})\left(1 + \lambda\frac{\partial \log c}{\partial \log y}\right), \qquad \frac{\partial \log \zeta_{\boldsymbol{\tau}}}{\partial \log r} = \lambda(1-\zeta_{\boldsymbol{\tau}})\left(1 - \frac{\partial \log c}{\partial \log r}\right). \tag{A26}$$

To derive firm-level scale elasticity of marginal costs $\varepsilon^m \equiv \frac{\partial \log c}{\partial \log y}$, follow:

$$\begin{aligned}
\frac{\partial \log c}{\partial \log y} &= \sum_{\tau} \frac{\partial \log c}{\partial \log z^*_{\tau}} \frac{\partial \log z^*_{\tau}}{\partial \log y} + \sum_{\tau} \frac{\partial \log c}{\partial \log \zeta_{\tau}} \frac{\partial \log \zeta_{\tau}}{\partial \log y} \\
&= \sum_{\tau} \pi_{\tau} \left[\left(-\omega_{K,\tau} - \omega^f_{\tau} \widetilde{G}(z^*_{\tau}|\boldsymbol{\tau}) \right) \frac{\zeta_{\tau}}{1-\lambda} + \omega^f_{\tau}(1-\zeta_{\tau}) \right] \left(1 + \lambda \frac{\partial \log c}{\partial \log y} \right) \\
&= -\sum_{\tau} \pi_{\tau} \widetilde{G}(z^*_{\tau}|\boldsymbol{\tau}) \frac{\zeta_{\tau}}{1-\lambda} \omega^f_{\tau} \left(1 + \lambda \frac{\partial \log c}{\partial \log y} \right) \\
&= -\frac{\omega^f}{1-\lambda} \left(1 + \lambda \frac{\partial \log c}{\partial \log y} \right) \mathbb{E}_f \left[\zeta_{\tau} \widetilde{G}(z^*_{\tau}|\boldsymbol{\tau}) \right],
\end{aligned}$$

where the second equality plugs in intermediate elasticities, the third uses the identity in Equation (A21), and the fourth collects terms. Solving for the elasticity $\frac{\partial \log c}{\partial \log y}$, we get Equation (19) in the main proposition.

Using a similar process, we examine the firm-level elasticity of marginal costs with respect to factor price $\Omega^m_K \equiv \frac{\partial \log c}{\partial \log r}$, and derive this expression as follows:

$$\begin{aligned}
\frac{\partial \log c}{\partial \log r} &= \sum_{\tau} \frac{\partial \log c}{\partial \log z^*_{\tau}} \frac{\partial \log z^*_{\tau}}{\partial \log r} + \sum_{\tau} \frac{\partial \log c}{\partial \log \zeta_{\tau}} \frac{\partial \log \zeta_{\tau}}{\partial \log r}, \\
&= \sum_{\tau} \pi_{\tau} \left(\omega_{K,\tau} + \omega^f_{\tau} \widetilde{G}(z^*_{\tau}|\boldsymbol{\tau}) \right) \left(1 + \frac{\lambda}{1-\lambda} \zeta_{\tau} \left(1 - \frac{\partial \log c}{\partial \log r} \right) \right) \\
&\qquad - \sum_{\tau} \pi_{\tau} \omega^f_{\tau} \lambda (1-\zeta_{\tau}) \left(1 - \frac{\partial \log c}{\partial \log r} \right), \\
&= \omega_K + \sum_{\tau} \pi_{\tau} \omega^f_{\tau} \widetilde{G}(z^*_{\tau}|\boldsymbol{\tau}) \left(1 + \frac{\lambda}{1-\lambda} \zeta_{\tau} \left(1 - \frac{\partial \log c}{\partial \log r} \right) \right), \\
&= \omega_K + \omega^f \left(\mathbb{E}_f \left[\widetilde{G}(z^*_{\tau}|\boldsymbol{\tau}) \right] + \frac{\lambda}{1-\lambda} \left(1 - \frac{\partial \log c}{\partial \log r} \right) \mathbb{E}_f \left[\widetilde{G}(z^*_{\tau}|\boldsymbol{\tau}) \zeta_{\tau} \right] \right)
\end{aligned}$$

where the second equality plugs in the intermediate elasticities, the third applies identity (A21) to the final term—so that $\omega^f_{\tau}(1-\zeta_{\tau}) = \frac{\zeta_{\tau}}{1-\lambda}\omega_{K,\tau}$—which cancels the $\omega_{K,\tau}$ cross-term generated by expanding the first sum, and the fourth invokes the operator $\mathbb{E}_f$. Solving for $\frac{\partial \log c}{\partial \log r}$, we get Equation (20) in the main proposition. □

A.4.4 Proof of Proposition 4

First, consider the substitution elasticity $\sigma \equiv \frac{1}{\Omega_K} \frac{\partial \log \ell}{\partial \log r}$. The firm's labor demand satisfies $w\ell = \omega_L c(y/a)$. Therefore, $\frac{\partial \log \ell}{\partial \log r} = \Omega^m_K + \frac{\partial \log \omega_L}{\partial \log r}$. Generalizing (13), the firm-level labor-share is

$$\omega_L = \left(\frac{w}{c}\right)^{1-\lambda} \int \xi_{\boldsymbol{\tau}} \overline{Z}_{L,\boldsymbol{\tau}}^{\lambda-1} G_{\lambda-1}(z_{\boldsymbol{\tau}}^*|\boldsymbol{\tau}) d\mu(\boldsymbol{z_K}, \xi).$$

From here, the elasticity is

$$\begin{aligned}
\frac{\partial \log \omega_L}{\partial \log r} &= -(1-\lambda)\Omega_K^m - \mathbb{E}_L\left[\widetilde{G}_{\lambda-1}(z_{\boldsymbol{\tau}}^*|\boldsymbol{\tau}) \frac{\partial \log z_{\boldsymbol{\tau}}^*}{\partial \log r}\right], \\
&= -(1-\lambda)\Omega_K^m + \mathbb{E}_L\left[\widetilde{G}_{\lambda-1}(z_{\boldsymbol{\tau}}^*|\boldsymbol{\tau}) \left(1 + \frac{\lambda}{1-\lambda}\zeta_{\boldsymbol{\tau}}(1-\Omega_K^m)\right)\right],
\end{aligned}$$

where the second equality comes from the intermediate elasticity Equation (A25) and the labor-cost-weighted average of a type-level object $h_{\boldsymbol{\tau}}$ is

$$\mathbb{E}_L[h_{\boldsymbol{\tau}}] = \frac{1}{\omega_L}\left(\frac{w}{c}\right)^{1-\lambda} \int \xi_{\boldsymbol{\tau}} \overline{Z}_{L,\boldsymbol{\tau}}^{\lambda-1} G_{\lambda-1}(z_{\boldsymbol{\tau}}^*|\boldsymbol{\tau}) h_{\boldsymbol{\tau}} d\mu(\boldsymbol{z_K}, \xi).$$

Putting the expressions together, the substitution elasticity is

$$\frac{\partial \log \ell}{\partial \log r} = \lambda \Omega_K^m + \mathbb{E}_L\left[\widetilde{G}_{\lambda-1}(z_{\boldsymbol{\tau}}^*|\boldsymbol{\tau})\right] + \frac{\lambda}{1-\lambda}(1-\Omega_K^m)\mathbb{E}_L\left[\widetilde{G}_{\lambda-1}(z_{\boldsymbol{\tau}}^*|\boldsymbol{\tau})\zeta_{\boldsymbol{\tau}}\right].$$

When dividing by Ω_K, this gives Equation (21) in the main text.

Next, we consider the scale elasticities of factor demand. By Shephard's lemma, $k = \frac{\partial C}{\partial r}$, so $\frac{\partial k}{\partial y} = \frac{\partial}{\partial y}\left(\frac{\partial C}{\partial r}\right) = \frac{\partial}{\partial r}\left(\frac{\partial C}{\partial y}\right)$. Since the cost function is $C = \frac{cy}{a} + c^f$, by the envelope theorem, $\frac{\partial C}{\partial y} = \frac{c}{a}$, so $\frac{\partial k}{\partial y} = \frac{c}{ar}\Omega_K^m$. Then,

$$\eta_K = \frac{\partial \log k}{\partial \log y} = \frac{c(y/a)}{rk}\Omega_K^m = \varepsilon \frac{\Omega_K^m}{\Omega_K},$$

which is Equation (22) in the main text. The proof for η_L is analogous.

We derive σ at the limit without fixed costs, where $\zeta_\tau = 0$ and $\Omega_K^m = \Omega_K = \omega_K$. Note that the extensive scale advantage margin term of Equation (21) becomes zero. Applying Equation (26), the result of Equation (28) immediately follows.

We also derive σ at the limit without marginal costs, where $\zeta_\tau = 1$ and $\Omega_K^m > \Omega_K$. Here,

$$\sigma = \left(1 + \frac{\Omega_K^m - \Omega_K}{\Omega_K}\right)\lambda + \left(\frac{1}{\Omega_K} + \frac{1-\Omega_K^m}{\Omega_K}\frac{\lambda}{1-\lambda}\right)\mathbb{E}_L\left[\widetilde{G}_{\lambda-1}(z_\tau^*|\tau)\right],$$

where Equation (29) follows from applying Equation (26) and collecting terms. □

A.4.5 Proof of Lemma 1

Fix a task τ and, for brevity, drop the corresponding subscript on $(h, \alpha, \psi, \underline{h})$ throughout the proof. The Lagrangian for the compute minimization problem in Equation (36) 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:

$$\frac{\partial \mathscr{L}}{\partial m} = 6d - \alpha_m\, \Lambda\, (\psi_m\, m)^{-\alpha_m} \frac{1}{m} = 0 \tag{A27}$$

$$\frac{\partial \mathscr{L}}{\partial d} = 6m - \alpha_d\, \Lambda\, (\psi_d\, d)^{-\alpha_d} \frac{1}{d} = 0 \tag{A28}$$

Rearrange and take the ratio of Equations (A27) and (A28):

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

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

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

Multiplying these together yields:

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

Inverting this expression delivers the fixed cost productivity in Equation (37). Finally, since each instance of inference requires computation proportional to model size, the marginal productivity of AI is the inverse of the optimal model size, $z_{K,\tau} = 1/m$, which together with the expression for m above delivers Equation (38). □

A.4.6 Firm Productivity Distribution

Across tasks, limited automation by AI occurs in 2023, so we consider the economy when $\varphi_K \to \mathbf{0}$ for all task types τ. Using Equation (14), in this limit every firm j in industry

i has the same value of $c_j = w/\overline{Z}_{L,i}$, where $\overline{Z}_{L,i}^{\lambda-1} = \sum_\tau \xi_{i\tau}\overline{Z}_{L,\tau}^{\lambda-1}$ denotes average labor productivity in an industry i. Since total variable cost is $c_j\,(y/a)$ by Proposition 2, the marginal cost of output for a firm with productivity a is c_j/a. Assuming the distribution $H_i(a)$ of firm productivity in an industry i is Pareto with shape parameter ϑ and scale parameter $\underline{a_i}$, total industry labor expenditure is $w\ell_i \equiv \int_{\underline{a_i}}^{\infty} (c_j/a)^{1-\eta}\, dH_i(a) \propto \frac{\vartheta}{1+\vartheta-\eta}\underline{a_i}^{\eta-1}$, where convergence requires $\vartheta > \eta - 1$. Labor expenditure in firms with productivity greater than a is $w\widetilde{\ell}_i^{-1}(a) \equiv \int_a^{\infty} (c_j/a')^{1-\eta}\, dH_i(a') \propto \frac{\vartheta}{1+\vartheta-\eta}\underline{a_i}^{\vartheta} a^{\eta-\vartheta-1}$. Taking the ratio yields $\frac{\widetilde{\ell}_i^{-1}(a)}{\ell_i} = \left(a/\underline{a_i}\right)^{\eta-1-\vartheta}$. From Equation (A5), at the limit $\varphi_K \to \mathbf{0}$, firm employment satisfies $\ell_j \propto (c_j/a)^{1-\eta} \propto a^{\eta-1}$, inverting gives $a \propto \ell^{\frac{1}{\eta-1}}$, so the share of labor in industry i with employment above cutoff ℓ corresponding to productivity a is

$$\frac{\widetilde{\ell}_i^{-1}(\ell)}{\ell_i} = A_i\,\ell^{1-\frac{\vartheta}{\eta-1}},$$

where A_i collects industry-level constants ($\underline{a_i}$, $\overline{Z}_{L,i}$, w, and industry demand). Taking logs, this is the OLS specification of Equation (40), with A_i absorbed by the industry fixed effects α_i. At the relevant limit, across-industry employment shares are simply the industry cost shares Ψ_i. Plugging $c_j = w/\overline{Z}_{L,i}$ into both Equations (A4) and (A7) provides an expression for uncovering industry-level productivity using data on industry employment shares.

B Appendix for Uncovering the Patterns of Productive Advantage

B.1 Mapping of Task Error Tolerance to Cross-Entropy Loss

It is easier to measure the characteristics of real production tasks in terms of error tolerance (i.e., the maximum tolerable error rate) rather than the technical measures of cross-entropy or prediction loss. Therefore, we use our experiment to build an empirical correspondence between the classification error rate and the cross-entropy loss, conditional on classification complexity, at the testing stage. Estimating this correspondence allows us to map estimated fine-tuning costs to the LLM/human-constructed measures of task-level error tolerance in Section 3.3. For a task τ with complexity χ_τ, we estimate the following mapping for a model that performs at cross-entropy h_τ and error rate e_τ:

$$\log h\left(e_\tau;\chi_\tau\right) = \beta_0^h + \beta_1^h \log e_\tau + \beta_2^h \log\chi_\tau + \beta_3^h \log e_\tau \times \log\chi_\tau$$

$$+ \beta_4^h \chi_\tau^{-1} + \beta_5^h e_\tau \log e_\tau, \tag{A29}$$

which captures multiple potential non-linearities between task error rate and complexity in their mapping to cross-entropy loss at the testing stage.

Mapping Error Rates to Cross-Entropy Loss Finally, Table A1 reports the estimated parameters for the relationship between error rates and cross-entropy loss in Equation (A29).

Table A1: Cross-Entropy Loss to Error Parameters

Parameter	β_1^h	β_2^h	β_3^h	β_0^h	β_4^h	β_5^h	**N**	$\mathbf{R^2}$
Estimate	0.282	0.210	0.126	0.505	-1.754	0.356	3435	0.933
90% CI	[0.253, 0.310]	[0.198, 0.222]	[0.119, 0.133]	[0.435, 0.573]	[-1.841, -1.664]	[0.290, 0.420]		

Note: Dependent variable is $\log h$. Sample restricted to accuracy $< 100\%$. CIs are bootstrap percentile CIs based on re-estimation of $\boldsymbol{\beta}^h$.

B.2 Scaling Law Coefficients

Table A2 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 coefficients ψ_m, ψ_d, α_m, and α_d under different task complexities χ according to Equation (39), which relates model size m and effective data size $\widehat{d}$ to cross-entropy loss h in vision classification tasks of varying complexity. Relative to data size d, which corresponds to the number of observations in training data, the effective size $\widehat{d}$ accounts both for the task complexity χ and for repeated passes over data in the training process. It is given by the equation

$$\widehat{d} = \frac{d}{\chi}\left(1 + \delta\left(1 - \exp\left(\frac{1-p}{\delta}\right)\right)\right),$$

where p is the number of training epochs and δ (estimated in Table A2) governs the diminishing returns to data reuse. 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_m 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.

Table A2: Estimated Scaling Law Coefficients

Parameter	$\beta^{\psi}_{m,0}$	$\beta^{\psi}_{m,1}$	$\beta^{\alpha}_{m,0}$	$\beta^{\alpha}_{m,1}$	$\beta^{\psi}_{d,0}$	$\beta^{\psi}_{d,1}$	$\beta^{\alpha}_{d,0}$	$\beta^{\alpha}_{d,1}$	$\log \underline{h}$	δ
Estimate	4.29	-3.88	-1.92	-0.09	-2.50	-1.10	-0.42	-0.08	-27.06	5.11
90% CI	[3.25, 5.40]	[-4.15, -3.64]	[-1.97, -1.87]	[-0.11, -0.07]	[-3.03, -2.06]	[-1.17, -1.01]	[-0.53, -0.29]	[-0.10, -0.05]	[-38.17, -24.91]	[4.68, 5.50]

Note: Dependent variable is $\log h$, where h is cross-entropy loss in Equation (35); data size enters through the effective data measure $\widehat{d}$ defined above. Sample restricted to accuracy $< 100\%$. CIs are bootstrap percentile CIs based on re-estimation of the full model using 5000 bootstrap samples.

B.3 LLM Prompts

All task-level measures in this section are generated with GPT-5. We ask the model to classify tasks by their automation type with the following prompt:

- You are an AI classification model. Given a task description and a list of automation types [a random order of (i) robotics and physical automation, (ii) analytical/predictive automation, (iii) computer vision, (iv) LLM or natural language processing automation, or (v) other or miscellaneous], assign a decimal weight to each automation type showing how much of the task it can automate. Assign nonzero weights to automation types that are central or clearly implied by the task—even if not stated word-for-word. Do not assign weights for capabilities that are irrelevant to the core task. *Output:* (1) Return only one JSON object. (2) Return a decimal number > 0 and < 1 for each automation type. (3) If a capability is not explicitly required, its weight must remain 0. (4) All weights for the task must sum to 1. (5) No extra text, comments, or explanations.

Even under identical prompts and otherwise fixed settings, LLM responses feature random variations. We ask the LLM to categorize each task ten times, randomizing the order in which we list automation types. We categorize a task as a vision task if all responses assign at least a 50% weight to computer vision as the automation type. Increasing the number of responses per task reduces dispersion in average weight to computer vision assigned to the task. Our sample size of ten corresponds to a point in which the marginal reduction in dispersion from additional sampling is small. Figure A7 illustrates the reduction in dispersion by sample size.

We ask the LLM the following prompts for vision tasks.

- **Task Frequency:** Consider the "visual" component of this task that can be automated using AI computer vision tools (e.g., image classification/segmentation).

How often does a typical worker in this occupation perform this "visual" component of the task? Answer with a single positive number representing the number of times per month (no text, no ranges).

- **Task Time Use:** Consider the "visual" component of this task that can be automated using AI computer vision tools (e.g., image classification/segmentation). Which fraction of a typical worker's total working time is actively spent performing this "visual" component of the task (excluding background or passive time)? Answer with a single decimal strictly between 0 and 1.

- **Task Accuracy Requirement:** Consider the "visual" component of this task that can be automated using AI computer vision tools (e.g., image classification/segmentation). What is the worst rate of error (e.g., in the corresponding classification/segmentation problem) that a worker could have in performing the "visual" component of this task and still be considered qualified? Answer with a single decimal number between 0 and 1 (0 = no errors, 1 = all errors).

- **Task Complexity:** Consider the "visual" component of this task that can be automated using AI computer vision tools (e.g., image classification/segmentation). How many distinct classes, categories, or segments would be present in the corresponding classification/segmentation problem? Provide the "effective" number of classes as a measure of the complexity of the classification/segmentation problem for a computer vision AI system. Answer with a single strictly positive number (no text or ranges).

We ask the LLM the following prompts for non-vision tasks:

- **Task Frequency:** Consider the "non-visual" component of this task that can be automated using AI tools other than computer vision (e.g., LLMs, robotics, predictive analytics, etc). Note that the task may be "non-visual" in its entirety. How often does a typical worker in this occupation perform the "non-visual" component of the task? Answer with a single positive number representing the number of times per month (no text, no ranges).

- **Task Time Use:** Consider the "non-visual" component of this task that can be automated using AI tools other than computer vision (e.g., LLMs, robotics, predictive analytics, etc). Note that the task may be "non-visual" in its entirety. Which fraction of a typical worker's total working time is actively spent performing this "non-visual" component of the task (excluding background or passive time)? Answer with a single decimal strictly between 0 and 1.

B.4 Concordance from Occupations to Industries

Construction of Task Weights in Industries $\xi_{i\tau}$ First, we use the 2023 Occupational Employment and Wage Statistics to construct the wage-bill share of each 6-digit SOC occupation in each 4-digit NAICS industry. The wage-bill of an occupation-industry is calculated as the product of average wage and total employment. The sum of wage-bill shares π_{io}^{WB} over occupations is one for each industry.

Second, we use the LLM-generated task frequency measures to calculate the frequency share of each task in an occupation $\bar{\bar{\xi}}_{o\tau}$. The sum of frequency shares $\bar{\bar{\xi}}_{o\tau}$ over tasks sums to one for each occupation.

Third, we use the wage-bill shares of occupations in industry π_{io}^{WB} and the frequency share of tasks in occupations $\bar{\bar{\xi}}_{o\tau}$ to construct the weighting of tasks in occupational output. To do so, we derive the wage-bill weighted sum of frequency shares across occupations for each task and industry $\bar{\bar{\xi}}_{i\tau} = \sum_o \pi_{io}^{WB} \bar{\bar{\xi}}_{o\tau}$. The measure $\bar{\bar{\xi}}_{i\tau}$ corresponds to how frequently an average employee in an industry i performs task τ as a proportion of her total tasks. These values are always less than one. Finally, we normalize $\bar{\bar{\xi}}_{i\tau}$ so that its sum over tasks amounts to one for each industry. We denote the normalized variables as $\xi_{i\tau}$.

We also construct occupation-task shares $\bar{\bar{\xi}}_{o\tau}$ based on O*NET task importance instead of frequency variables. The resulting estimates are highly correlated with the baseline: the two measures display an $R^2 = 0.83$.

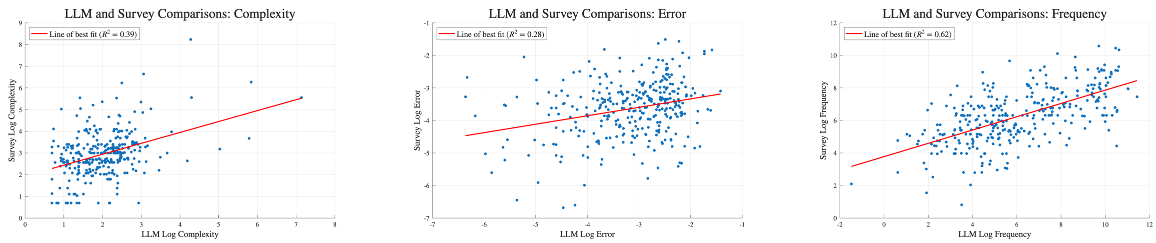
Construction of Task Complexity and Error Tolerance Measures The procedure described in Section 3.3 provides an occupation-task measure of complexity $\chi_{o\tau}$. Since every vision task in our sample maps to a single occupation $o(\tau)$ (though one occupation can map to multiple vision tasks), this measure directly defines task complexity, $\chi_\tau = \chi_{o(\tau)\tau}$; the measure of error tolerance is defined in the same manner. More generally, were a task performed by multiple occupations, we would concord these values to the industry level by considering the average complexity of the task *when it is performed in an industry*, that is, $\log \chi_{i\tau} = \sum_o \frac{\pi_{io}^{WB}\bar{\bar{\xi}}_{o\tau}}{\sum_o \pi_{io}^{WB}\bar{\bar{\xi}}_{o\tau}} \log \chi_{o\tau}$.

B.5 Correspondence Between LLM and Survey Estimates

B.6 Additional Results About Aggregate Outcomes

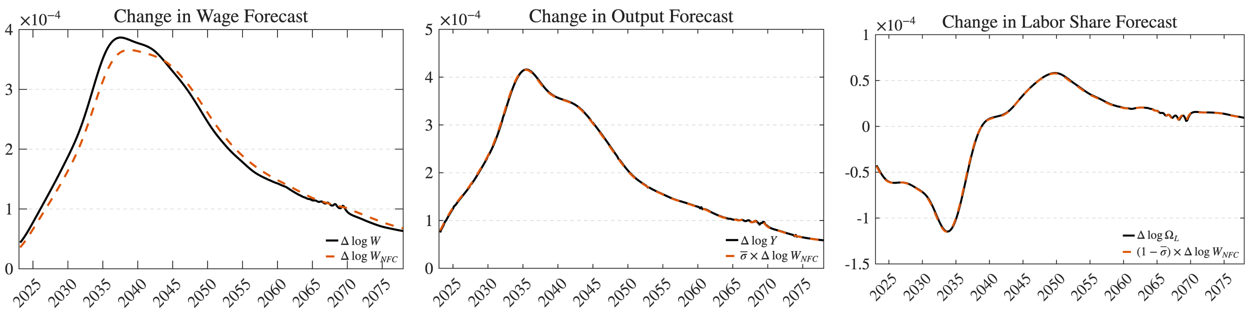
Figure A2 compares the annual changes in wages, output, and labor share with the approximation found in Equation (34). As Equation (34) shows, if we disregard scale de-

Figure A1: LLM vs. Survey Average Task Estimates



Note: Each panel compares, in logs, task-level averages of LLM-generated responses and worker-survey responses from Svanberg et al. (2024) to the same question—task complexity, error tolerance, and task frequency, respectively—for the 445 vision-related tasks with valid survey responses. Points are task items; red lines are OLS fits, with $R^2 = 0.39$, 0.28, and 0.62, respectively.

Figure A2: Changes in Aggregate Outcomes on the Automation Path



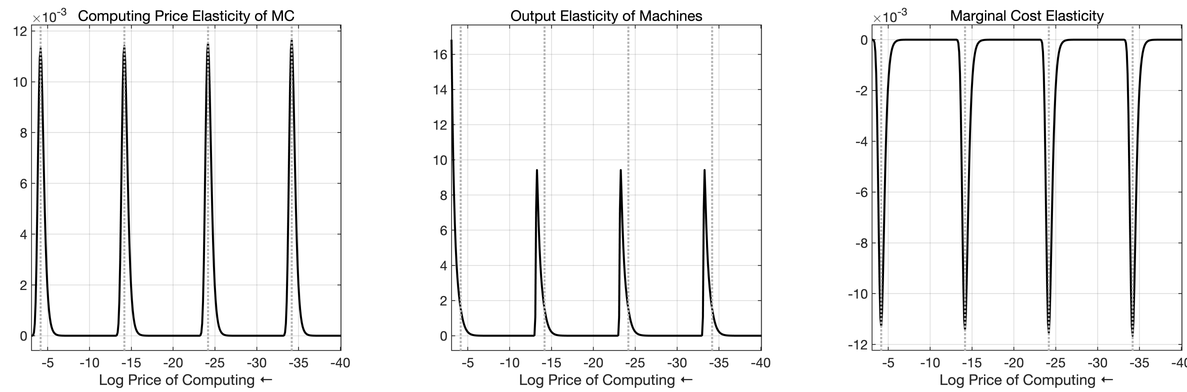
Note: Each panel plots the annual change in the log wage, log output, and log labor share along the automation path (2025 to 2075), reported on a 10^{-4} scale. Solid lines are the model outcomes; dashed lines (NFC) are the no-fixed-cost, first-order approximation of Equation (34), which sets scale dependence to zero.

pendence, to the first order of approximation the change in wages as a response to a fall in the price of AI is given by $\Delta \log w_{NFC} = -\frac{\overline{\Omega}_K}{\overline{\Omega}_L} \Delta \log r$. The leftmost panel of Figure A2 compares the annual growth in real wages predicted by our model with the value found based on this approximation, showing that the two are fairly close.

Equation (34) then shows that, disregarding scale dependence, the change in the aggregate output is proportional to the product of the change in real wages and the aggregate elasticity of substitution. The middle panel of the figure shows that this relationship is a fairly precise approximation of the one found in our model. Finally, the rightmost panel shows that the changes in labor share are also well approximated by the product of the change in real wages and $1 - \overline{\sigma}$. Since the aggregate elasticity of substitution initially exceeds unity, the labor share initially declines for a few decades until the elasticity falls below unity at which point the labor share begins to slowly rise again. The increased importance of machines relative to labor in firms' factor allocations associated with waves of AI automation is only temporary.

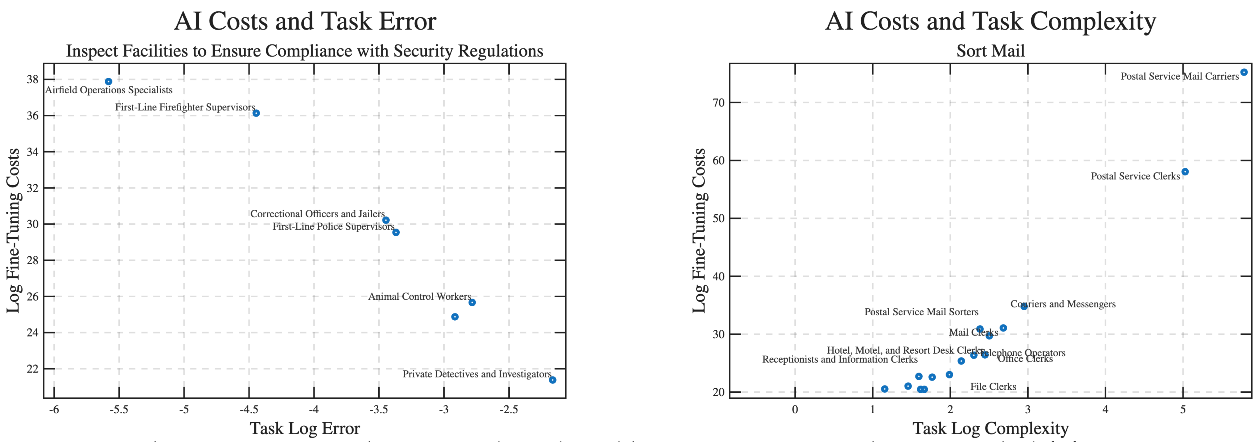
C Additional Tables and Figures

Figure A3: Scale-Dependent Elasticities in the No-Marginal-Cost Case



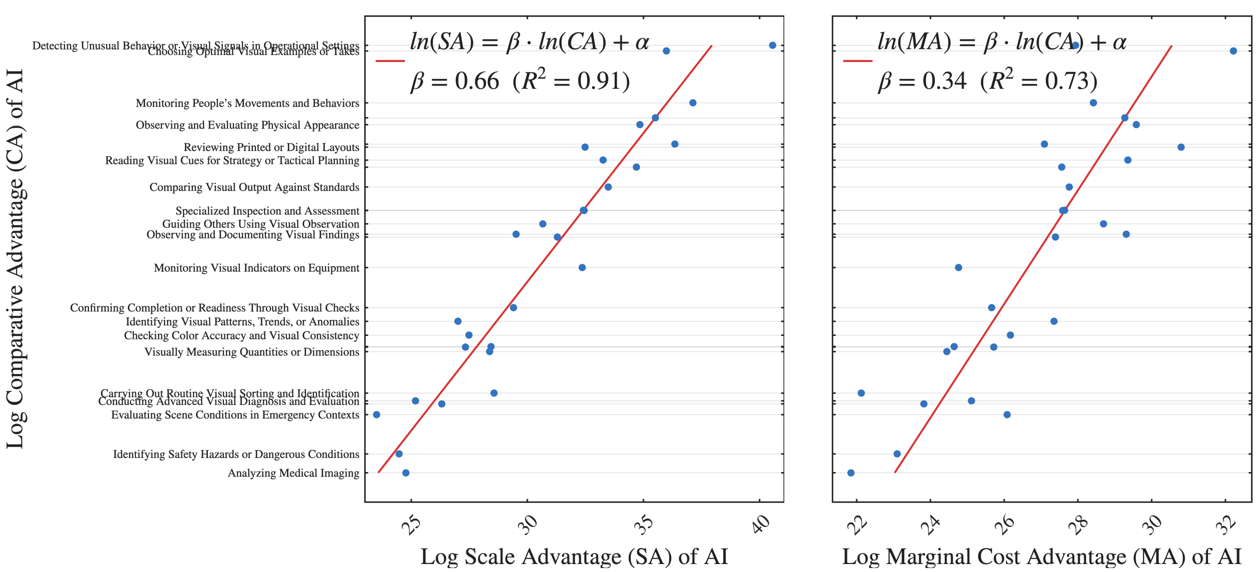
Note: This figure complements Panel B of Figure 1 by reporting firm-level elasticities that exhibit scale dependence in the no-marginal-cost case, as the price of machines r declines. The panels show, respectively, the elasticity of marginal cost to machine prices Ω_K^m (Equation (20)), the output elasticity of machine demand η_K (Equation (22)), and the scale elasticity of marginal cost ε^m (Equation (19)). Parameters are as in Panel B of Figure 1: $\lambda = 0.5$, $\theta = 1.3$, $|\mathcal{T}| = 5$, $\xi_{\boldsymbol{\tau}} = 0.01$ for $\boldsymbol{\tau} \leq 4$, $\overline{Z}_{L,\boldsymbol{\tau}} = 1$ for labor, $z_K = (\infty, \infty, \infty, \infty, 0)$, and $z_K^f = (1, e^{-10}, e^{-20}, e^{-30}, 0)$.

Figure A4: AI Costs and Task Characteristics



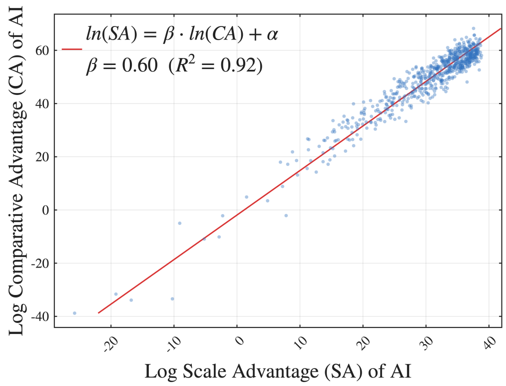
Note: Estimated AI costs increase with more complex tasks and lower maximum error tolerances. In the left figure, we examine estimated costs for the DWA category "inspect facilities to ensure compliance with security or safety regulations." This task displays substantial variation in maximum error tolerance and little variation in task complexity. In the right figure, we examine estimated costs for the DWA category "sort mail." This task displays substantial variation in task complexity but little variation in maximum error tolerance. The wide complexity spread arises because postal carriers sort mail into hundreds of delivery-sequence bins, whereas clerical sorting uses only a few categories. In the visualization, we exclude some occupations for readability, including couriers, switchboard operators, bill collectors, word processors and typists, and postal service clerks.

Figure A5: Variance Decomposition of AI Comparative Advantage Across Task Categories in 2023



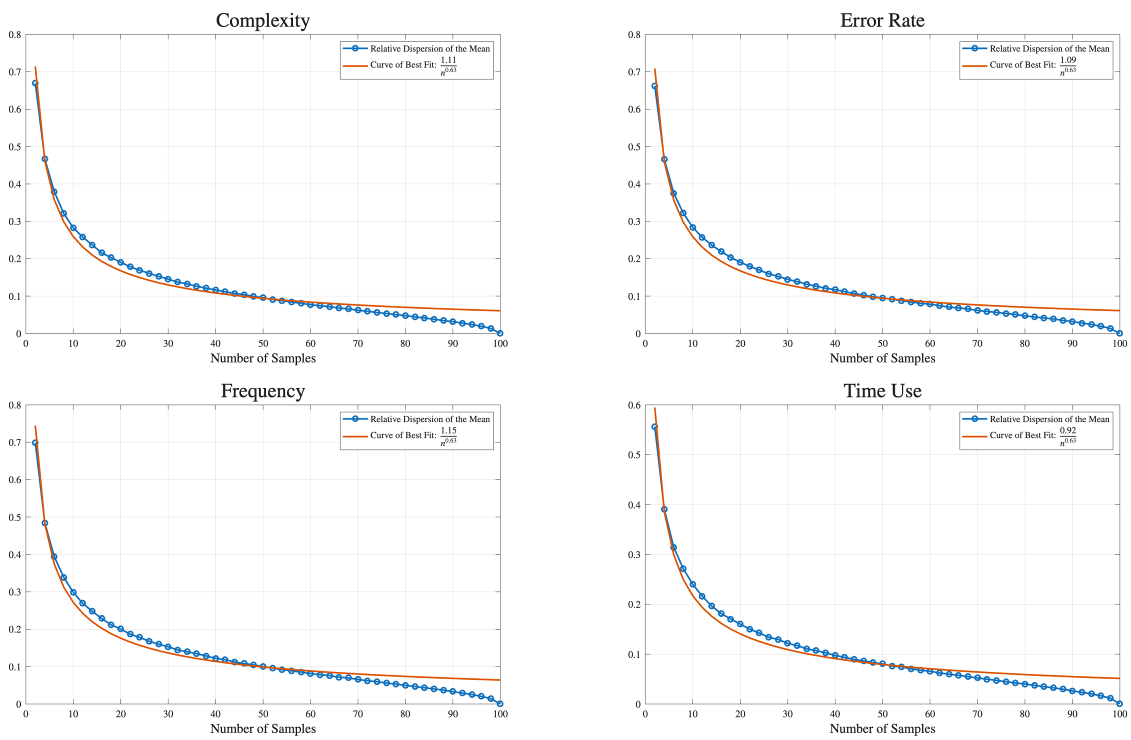
Note: This figure conducts a decomposition of variation in average comparative advantage $\overline{CA}$ in LLM-generated task categories into scale $\overline{SA}$ and marginal $\overline{MA}$ advantage components. Scale accounts for 66% of the variation.

Figure A6: Variance Decomposition of AI Comparative Advantage Across Tasks in 2050



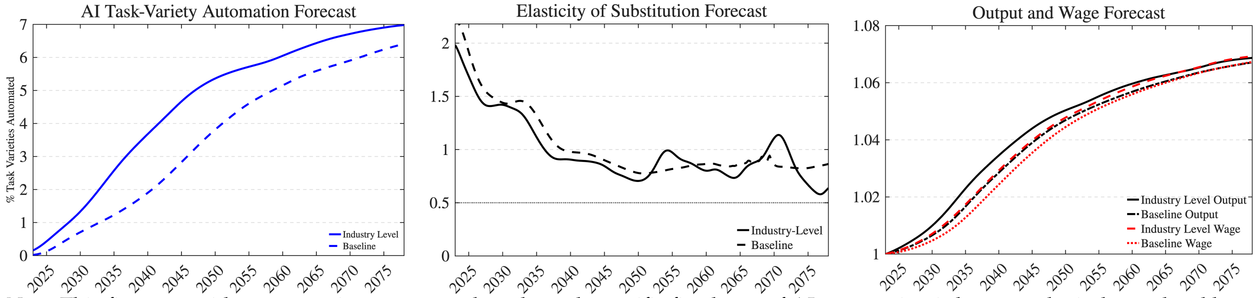
Note: This figure considers the economy when compute prices fall to their predicted 2050 levels. It decomposes across-task variation in comparative advantage CA into that attributable to scale SA and marginal MA advantage. Scale accounts for 60% of the variation.

Figure A7: Dispersion in Average Task Measures vs. Sample Size



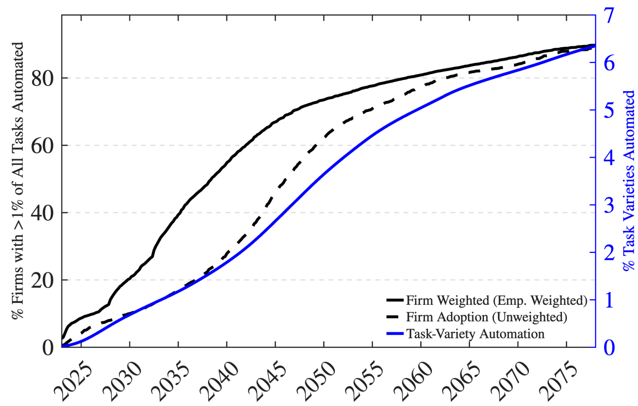
Note: This figure shows the dispersion of average LLM-generated estimates for task complexity (how many distinct classes or categories must be distinguished in the visual sub-task), error rate tolerance (the worst error rate a worker could have and still be considered qualified), frequency (how many times per month the visual sub-task is performed), and time use (what fraction of working time is spent actively performing the visual sub-task) as a function of the number of responses elicited from GPT-5. The figures show that dispersion declines predictably with sample size according to a power-law fit, meaning that additional samples steadily improve estimate stability. Our sample size of ten corresponds to a point in which the marginal reduction in dispersion from additional sampling is small. More broadly, the optimal number of iterations depends on resource budget: approximately ten samples already provide stable estimates for all variables.

Figure A8: Outcome Paths Using Industry-Level Fine-Tuning Fixed Costs



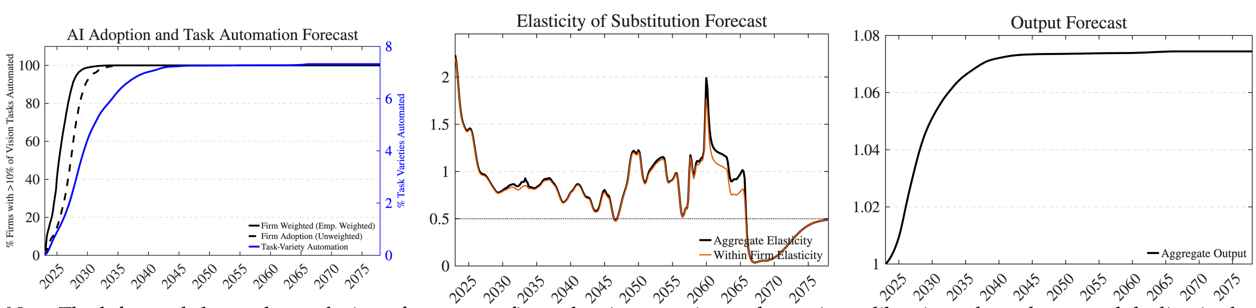
Note: This figure considers economic outcomes when the task-specific fixed cost of AI automation is borne at the industry level by a representative firm rather than at the firm level by heterogeneous firms. The left panel considers the variety automation rate, which is greater across the automation path when fixed costs are borne at the industry level relative to the baseline specification. The central panel considers the substitution elasticity between labor and AI. Like the baseline, the elasticity moves from gross substitutes (> 1) to gross complements (< 1) along the automation path, but non-monotonicities are more pronounced. The right panel considers the path of output and wages, which remains quite similar relative to the baseline.

Figure A9: Alternative Definition of Adoption in the Automation Path of AI



Note: This figure complements Figure 5 by showing the evolution of aggregate firm adoption rates using an alternative threshold of tasks automated that defines firm adoption. 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.

Figure A10: Outcome Paths Using Alternative Decline in Price of Machines



Note: The left panel shows the evolution of aggregate firm adoption rates in an alternative calibration where the annual decline in the price of machines is 74.0% in the dollar cost per FLOP (relative to 26.4% of the baseline). The central panel considers the substitution elasticity between labor and AI. Once all automatable tasks are conducted with AI, the extensive margins of automation vanish and the elasticity falls below λ: fixed training costs keep total costs more responsive than marginal costs to the machine price ($\overline{\Omega}_K^m < \overline{\Omega}_K$). The right panel considers the path of output and wages.