Research Policy 53 (2024) 105037

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Economic impacts of AI-augmented R&D

Tamay Besiroglu [a,1], Nicholas Emery-Xu [b,1,*], Neil Thompson [a]

[a] MIT FutureTech, United States of America
[b] UCLA Department of Economics, MIT FutureTech, United States of America

ARTICLE INFO

Dataset link: https://futuretech.mit.edu/community-resources

Keywords:
Artificial intelligence
Deep learning
Endogenous growth theory
Production function

ABSTRACT

Since its emergence around 2010, deep learning has rapidly become the most important technique in Artificial Intelligence (AI), producing an array of scientific firsts in areas as diverse as protein folding, drug discovery, integrated chip design, and weather prediction. As scientists and engineers adopt deep learning, it is important to consider what effect widespread deployment would have on scientific progress and, ultimately, economic growth. We assess this impact by estimating the idea production function for AI in two computer vision tasks that are considered key test-beds for deep learning and show that AI idea production is notably more capital-intensive than traditional R&D. Because increasing the capital-intensity of R&D accelerates the investments that make scientists and engineers more productive, our work suggests that AI-augmented R&D has the potential to speed up technological change and economic growth.

1. Introduction

We consider the effect of the adoption of Artificial Intelligence (AI) within science and engineering on idea production and, subsequently, on productivity and economic growth. Unlike previous work that provides only a theoretical treatment of the topic, we approach this question with microdata from deep learning, the AI paradigm responsible for nearly all landmark results in the past decade. We provide a framework for understanding the impact of two important trends: (i) the recent breakthroughs using deep learning in R&D and (ii) the rapid scaling of computation in deep learning systems. Using an endogenous growth framework, we show that if deep learning induces capital deepening in R&D, it could accelerate innovation and economic growth. Employing new computational and human capital data for deep learning papers and a novel machine learning method for estimating human capital, we estimate a production function for progress on two key computer vision tasks, obtaining estimates suggesting the technology has a substantially higher capital cost share than most R&D sectors in the U.S. Finally, we show that if deep learning is widely adopted in the U.S. R&D sector, it would induce an accumulation of computational capital that could nearly double the productivity growth rate.

Since the early 2010s, when it produced seminal breakthroughs in computer vision and speech recognition, deep learning has led to a rapid increase in the rate of progress in Artificial Intelligence (LeCun et al., 2015; Goodfellow et al., 2016, Ch. 1; Russell and Norvig, 2020, Ch. 1.3). Breakthroughs have been made in many areas, including computer vision, speech recognition, natural language processing, and game playing. Deep learning has also made inroads into parts of science largely untouched by previous AI research, including protein folding, semiconductor chip floorplanning, controlling nuclear fusion, and even discovering novel algorithms and new insights in pure mathematics. The rate at which longstanding problems have been solved and the pace at which deep learning systems have out-competed traditional algorithms have been surprisingly rapid to even some of its most seasoned practitioners.

As uses of AI proliferate, economists have sought to understand its impacts on wages, factor shares, and economic growth. A prominent line of thought asks whether deep learning has the potential to become a General Purpose Technology, one with widespread applications in a variety of industries and the ability to replace human labor across a wide variety of tasks (Goldfarb et al., 2022; Agrawal, 2022; Trajtenberg, 2018).[2]

* Corresponding author.
E-mail addresses: niemery@ucla.edu (N. Emery-Xu), neil_t@mit.edu (N. Thompson).

[1] Joint first authors.

[2] For example, Goldfarb et al. (2022) calculate the prevalence of deep learning-related job requirements in job postings within and across industries to predict whether technologies will become GPTs by whether they are in widespread use, capable of ongoing self-improvement, and enable multi-sector innovation. Across a set of 21 technologies, they find that machine learning displays the characteristics of a GPT in the deep learning era but not beforehand. In contrast, Thompson et al. (2020) question whether ongoing increases in deep learning performance will be sustainable, potentially undermining its ability to provide the long-term benefits of a GPT.

https://doi.org/10.1016/j.respol.2024.105037
Received 28 September 2023; Received in revised form 15 April 2024; Accepted 28 May 2024
Available online 8 June 2024
0048-7333/

Much of the existing research has focused on the potential of AI to impact final goods production, but it has also been pointed out (Cockburn et al., 2019) that AI also has the potential to change the innovation process itself. Such "Inventions of a Method of Invention" (IMI) can significantly affect the rate of idea production (Crafts, 2021; Cockburn et al., 2019) and, therefore, the overall rate of innovation in the economy. For example, building on the Weitzman (1998) model of recombinant technological development, Agrawal et al. (2019) argue that deep learning can improve knowledge production by effectively searching through and recombining a wider range of ideas than is possible by human scientists, potentially resulting in accelerated economic growth. Empirical testing of the impact of AI on R&D shows mixed results. Bianchini et al. (2020) find that the use of deep learning is positively correlated with the mean and variance of paper citations received, increasing the likelihood for a contribution to become an influential 'big hit.' However, they also find it is negatively correlated with the re-combinatorial novelty of ideas, measured as a function of the fraction of novel citation pairs in a given paper. Another line of research focuses on the relationship between AI and data, showing machine learning increases the returns to data and thus the rate of knowledge production for data-rich firms (Beraja et al., 2023; Abis and Veldkamp, 2024; Agrawal et al., 2018). While these insights are informative about firm-level effects, they shed less light on implications for the aggregate economy.

The impacts of AI on the innovation process deserve special attention because it has been pointed out that these, under suitable conditions, can have more dramatic permanent effects on productivity growth than those that arise from changes in final goods production. For example, in the semi-endogenous growth model of Aghion et al. (2019), the authors consider AI automation in producing final goods and in producing knowledge, and find that the latter can produce much more rapid output growth. Trammell and Korinek (2020) provide a review of the theoretical literature on AI and growth, concluding that, while a high degree of automation in final goods production can produce a one-time increase in the growth rate, a high degree of automation in the R&D sector can produce unbounded increases in economic growth.

We argue that the adoption of deep learning makes computational capital in R&D more productive, resulting in capital deepening that, if widespread, accelerates knowledge creation and economic growth. To motivate this, we derive a semi-endogenous growth model showing that a positive shock to the R&D elasticity of capital – such as might follow the widespread adoption of deep learning techniques – permanently increases the rate of idea accumulation and economic growth.

But will deep learning increase the R&D elasticity of capital? We provide supportive empirical evidence by estimating idea production functions for two relatively mature deployments of deep learning. To analyze human capital in deep learning, we develop a novel machine learning approach for estimating human capital and apply it to machine learning papers in the arXiv repository. To analyze computing resources, we augment the dataset from Thompson et al. (2020) to cover the entire universe of papers on two popular computer vision tasks.

Our estimates of the deep learning production function allow us to compare AI-augmented R&D with the R&D practiced in U.S. science and engineering fields. We find that deep learning's idea production function depends notably more on capital. This greater dependence implies that more capital will be deployed per scientist in AI-augmented R&D, boosting scientists' productivity and the economy more broadly. Our point estimates, when analyzed in our semi-endogenous growth model of the U.S. economy, suggest that AI-augmented areas of R&D would increase the rate of productivity growth by between 1.7- and 2-fold compared to the historical average rate observed over the past 70 years.

Our analysis is organized as follows. Section 2 motivates the importance of R&D capital intensity for economic growth using a semi-endogenous growth model. Section 3 describes the datasets, and Section 4 the empirical strategy we use to model and estimate idea production. A key input for these models is an estimate of the human capital of the teams working on particular AI projects. In Section 5 we develop a deep neural network for learning simple representations of human capital that outperforms other measures commonly used in scientometric literature. Our human capital estimates explain 60%–80% of variance in key publication-related outcomes, whereas standard linear models explain less than 20%. In Section 6, we present our empirical analysis, which implies that a firm in a competitive R&D sector using deep learning would be roughly 5 times more capital-intensive than current U.S. STEM R&D. In Section 7, we use our growth model to investigate the implications of higher capital intensity of R&D, and find that it implies a substantially faster productivity growth rate: 2- to 3-fold greater than the 0.8% growth rate the US saw over the last decade. In Section 8, we find that our results are robust to outliers and alternative model specifications. In Section 9, we conclude with a brief discussion of the implications of our results.

2. Capital-intensive R&D and deep learning

2.1. The role of capital in idea production

We argue that deep learning may affect the growth rate of knowledge by impacting the productivity of research capital. While capital is not usually the focus of R&D-based growth research, it has been shown to generate permanent growth effects by increasing the marginal product of labor in R&D and thus increasing investment in the R&D sector (Howitt and Aghion, 1998; Howitt, 1999). The key mechanism driving this result is that, unlike the stock of human labor, the rate of physical capital accumulation can be readily adjusted in response to a change in its productivity. A similar point is made by Aghion et al. (2019) in their study of the growth effect of AI. They show that in the classic endogenous growth case, a one-time increase in R&D automation will raise the long-run growth rate, as capital – an accumulable factor in production – becomes permanently more important.

Empirical work has validated the importance of physical capital investment in idea production. Helmers and Overman (2017) showed that the creation of the UK's Diamond Light Source synchrotron increased the research capital available to local scientists, which in turn increased their research publication output relative to UK scientists located elsewhere.

The prevalence of capital goods in U.S. R&D is documented by the National Science Foundation's Higher Education Research and Development Survey. It finds that academic institutions have spent over \$2 billion per year on capital equipment or software for R&D since 2010, representing roughly 4% of total direct R&D expenditures (National Center for Science and Engineering Statistics, 2021). In STEM fields, this share is higher — around 6% overall in 2020, with Chemistry at 14%, Material Science and Physics at 15%, and Engineering ranging between 7% and 11%. Since the overall capital intensity of academic science is 4%, there is enormous room for more capital-intensive approaches.

Compared to capital-intensive areas of science, there are reasons to suspect that R&D using deep learning might be yet more so. Recent capital-intensive examples include OpenAI's GPT-3 language model (Brown et al., 2020), DeepMind's AlphaZero game-playing system (Silver et al., 2017), and DeepMind's protein-folding system, each of which reportedly used millions of dollars worth of computing (Gibney, 2017-10-18; Jumper et al., 2021).

Computing theory also suggests reasons deep learning is capital-intensive and why it is likely to become more so. In classical statistical learning theory, there generally is a trade-off between bias and variance (Hastie et al., 2009). Once a model grows beyond a certain complexity threshold, it tends to overfit the data, worsening test performance[3] as the variance term dominates. Deep neural networks seem capable of evading this trade-off by vastly expanding the size of the network ("overparameterization"), that is by deploying more computational capital (Belkin et al., 2018; Nakkiran et al., 2021).

Surprisingly, empirical analyses have shown that the performance gains accruing to these networks with millions or billions of parameters are highly predictable. Generally, these analyses of "neural scaling laws" find that test error falls according to a power law in the scale of such models, and therefore in the amount of compute used (Kaplan et al., 2020; Hoffmann et al., 2022; Hestness et al., 2017; Sun et al., 2017; Lepikhin et al., 2020; Li et al., 2020; Jones, 2021; Bahri et al., 2021; Sharma and Kaplan, 2020).[4] Researchers are harnessing these dependencies to obtain better performance. For example, Thompson et al. (2020) shows that progress is highly dependent on computational resources across a wide range of machine learning tasks. For image classification on the ImageNet database, 71% of the variance in model performance is explained by computation used. The importance of computing resources in deep learning was elegantly summarized by Rich Sutton (Sutton, 2019), a prominent figure in the field of reinforcement learning, who wrote:

> The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin... Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation.

If deep learning is indeed more capital intensive, the investment dynamics implied by endogenous growth models would predict a rapid scale up to have occurred in the computational capital being used in AI-based R&D. Sevilla et al. (2022a) find exactly that: since the advent of deep learning, the growth in the amount of computational capital typically used in milestone models doubles roughly every 6 months, far outstripping the rate during previous eras of AI. So, while the size of capital investments made in deep learning systems are still small compared to, for example, those required for large-scale physics experiments, there are compelling reasons to believe that these models are capital intensive and will continue to become more so.

2.2. R&D capital intensity in a semi-endogenous growth model

Consider a simple semi-endogenous growth model similar to Jones (1995). There are two sectors, a final output-producing sector and an R&D sector in which additions to the stock of knowledge are made. Time is continuous and indexed by t. At time t, a fraction α_l of the labor force $L(t)$ is used in the R&D sector and fraction $1-\alpha_l$ in the goods-producing sector. Similarly, a fraction α_k of the capital stock $K(t)$ is used in R&D and the rest in goods production. For expositional clarity, we make similar simplifying assumptions as Romer (1990) that α_l and α_k are exogenous and constant. Ideas are non-rivalrous and the full stock is used equally in both sectors of the economy. For further simplicity, we assume constant returns to scale in the production of final goods. The quantity of output $Y(t)$ produced at time t is given by the Cobb–Douglas production technology:

$$Y(t) = \left[(1-\alpha_k)K(t)\right]^{\alpha}\left[A(t)(1-\alpha_l)L(t)\right]^{1-\alpha}, \quad \text{where} \quad \alpha \in (0,1), \tag{1}$$

where α denotes the returns to scale of capital. The production of new ideas depends on the quantities of capital and labor engaged in research and on the level of technology, $A(t)$. We assume there are diminishing returns in the production of new ideas from inputs that can be accumulated through investment ($\beta+\theta<1$). This assumption ensures a unique steady-state growth rate, and prevents the growth rate from exploding as the R&D inputs grow without bound. The idea stock, $A(t)$, grows as follows:

$$\dot{A}(t) = B\left[\alpha_k K(t)\right]^{\beta}\left[\alpha_l L(t)\right]^{\gamma} A(t)^{\theta}, \quad \text{where} \quad B>0, \quad \beta,\gamma,\geq 0 \quad \text{and} \quad \beta+\theta<1, \tag{2}$$

where B is a positive shift parameter and γ and β are the returns to scale from labor and capital in $R\&D$, respectively. We further make the common simplifying assumptions that there is a constant savings rate s, and that capital depreciates at a constant rate δ. We suppose that population grows at exogenous rate n. Thus, accumulation of capital and labor are described as follows:

$$\dot{K}(t) = sY(t) - \delta K(t), \quad \text{and} \quad \dot{L}(t) = nL(t), \quad \text{where} \quad s,\delta \in (0,1) \quad \text{and} \quad n>0. \tag{3}$$

Using Eqs. (1)–(3), we solve for the steady-state rates of growth in ideas and capital (denoted as g_a^* and g_k^* respectively.[5]) These are:

$$g_a^* = \frac{\beta+\gamma}{1-\beta-\theta}n, \quad \text{and} \quad g_k^* = \frac{1-\theta+\gamma}{1-\beta-\theta}n. \tag{4}$$

As is well known, when factor markets are competitive, β and γ uniquely identify capital and labor factor shares in the R&D sector. The following proposition considers the effect of a technology shock that increases the factor share of capital in R&D:

Proposition 1. *Consider a shift in the technology of R&D that, under competitive factor markets, increases the factor share of capital in R&D without reducing the returns to scale. That is, consider a shift from an R&D setting described by equation (2) to a setting described as follows:*

$$\dot{A}(t) = B\left[\alpha_k K(t)\right]^{\beta'}\left[\alpha_l L(t)\right]^{\gamma'} A(t)^{\theta}$$

where $\beta' > \beta, \quad \beta'+\gamma' \geq \beta+\gamma, \quad$ *and* $\quad \beta'+\theta<1.$

Such a shift in the technology of R&D has the following implications:

(a) the rate of idea accumulation is strictly and permanently increased, and
(b) the rate of economic growth is strictly and permanently increased.

Proof. Define $\Delta_X \equiv X' - X$ as the difference between pre- and post-values of parameter X. By assumption, $\Delta_\beta \geq -\Delta_\gamma$. The steady-state rate of growth in ideas is strictly increased if:

$$\frac{\beta'+\gamma'}{1-\beta'-\theta}n > \frac{\beta+\gamma}{1-\beta-\theta}n. \tag{5}$$

which follows from the fact that $\Delta_\beta \geq -\Delta_\gamma$. This proves part (a) of the proposition. The proof of part (b) may be found in Appendix A.2.

Proposition 1 states that if there is a shift to an R&D setting with a higher factor share of capital, the steady-state rate of technological change and economic growth will strictly and permanently increase. Further, this result is robust to shocks that also decrease the factor share of labor as long as returns to scale are not reduced. This simple result underscores the importance of the *composition* of inputs to R&D, which, as we will see later, we might have reason to expect will change if AI were widely adopted by scientists and engineers.

Why might we expect that the returns to scale in idea production will not diminish with the adoption of deep learning? One conceivable scenario is a decrease in θ, indicating a stronger 'fishing-out'

[3] Distinct from the training error, test performance is calculated on data points never seen by the network.

[4] See the "Related Works" section in Bahri et al. (2021) for an account of the state of the existing work.

[5] A complete derivation may be found in Appendix A.1.

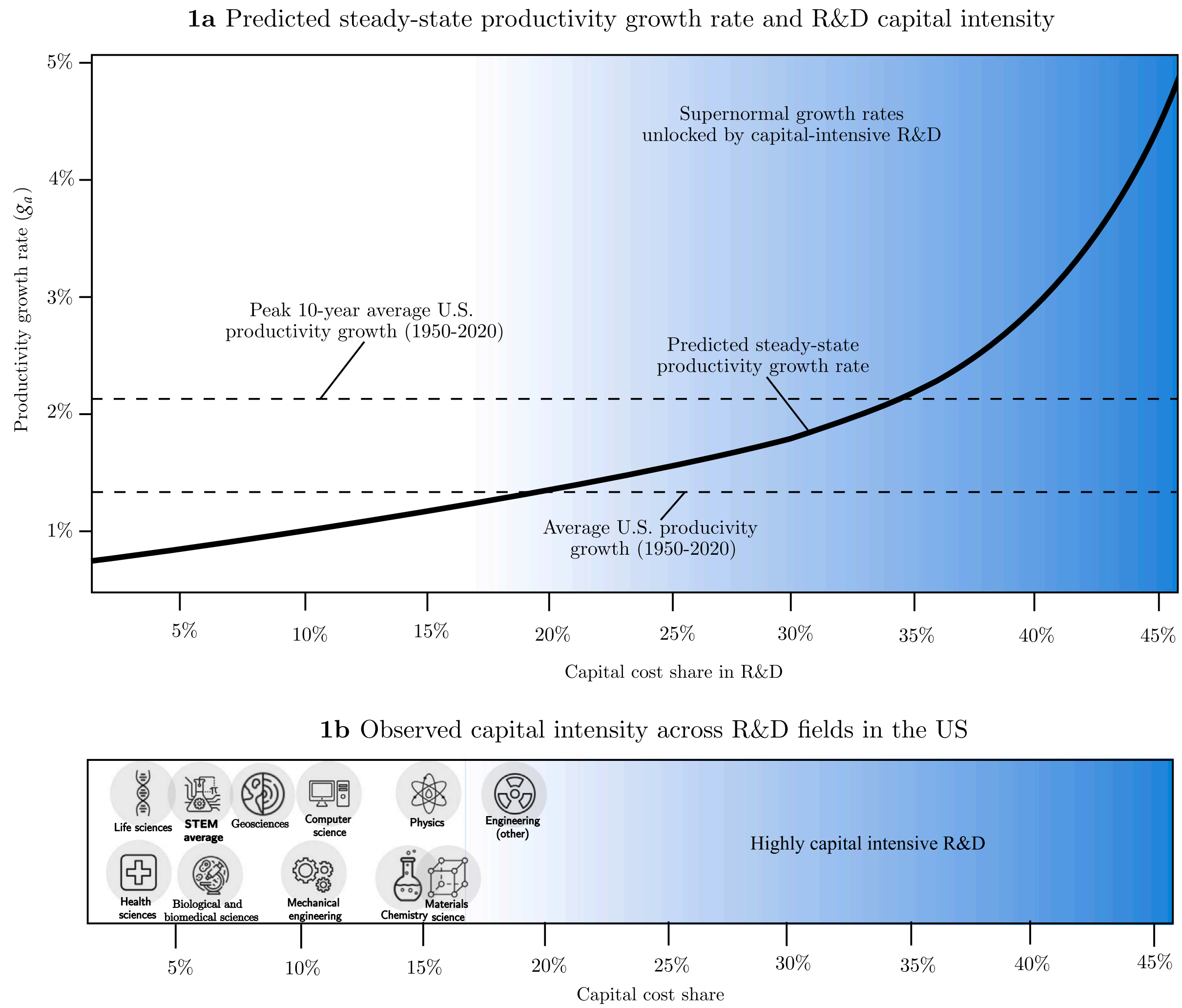


Fig. 1. Productivity growth and R&D factor share in a competitive R&D economy. Fig. 1a shows steady-state productivity growth as a function of the capital-factor share in a competitive R&D industry. Returns to scale are held constant. For this plot, we assume that the elasticity of R&D output to the stock of ideas (θ) is 1/2 and the elasticity of R&D output to labor inputs (γ) is 2/5 (consistent with our survey of existing estimates in Online Appendix B). Fig. 1b shows the share of R&D expenditure in each discipline that is spent on capital equipment, based on data from the National Science Foundation 2020 Higher Education Research and Development Survey (National Center for Science and Engineering Statistics, 2021)[6].

effect, where utilizing AI to expand our knowledge base might make it harder to generate subsequent innovations. However, without concrete evidence to support such a dynamic, we posit a technology-neutral impact from the integration of AI into knowledge creation processes and assume θ remains unchanged.

Furthermore, AI could plausibly diminish the productivity of scientists, significantly lowering γ, the returns to labor in R&D, while not providing commensurate gains through increasing the returns to capital. Strong effects in this direction are improbable in part because it is unlikely that firms will adopt AI technologies that severely impair the productivity of their human capital unless they provides offsetting gains from enhancing the productivity of physical capital. Moreover, there are reasons to believe that incorporating AI into scientific research will actually boost human productivity. This enhancement stems from AI's potential to provide new tools for reasoning and research (Carter and Nielsen, 2017; Bolanos et al., 2024), augment researchers in key tasks such as programming (Peng et al., 2023), and provide new insights that are otherwise difficult to attain (Agrawal et al., 2019; Sourati and Evans, 2023), amplifying the effectiveness of human researchers and accelerating the pace of innovation.

To illustrate how capital-intensive R&D could result in super-normal productivity growth, consider Fig. 1a. Suppose the R&D sector is competitive, such that wages and rents are equal to their marginal products. The steady-state rate of idea accumulation increases in the share of R&D expenditure dedicated to capital. Under conservative assumptions on our growth model, highly capital-intensive R&D (such as when optimizing R&D firms dedicate at least 20% of R&D expenditure to capital) would produce productivity growth rates in excess of the usual productivity growth rates observed in the U.S. By contrast, current U.S. R&D tends to be highly labor-intensive. Using 2020 survey data for NSF-supported STEM R&D (National Center for Science and Engineering Statistics, 2021) and assuming a Cobb–Douglas functional form for ideas production, we see that capital shares tend to fall between 3% and 20% (see Fig. 1b).[7,8]

In the Analysis section, we present evidence that the relative returns to capital for deep learning are higher than for other types of

[6] The underlying data and the details of the calculations used for Fig. 1b may be found here.

[7] STEM fields tend to have *higher* capital intensities than non-STEM fields.

[8] To compute the capital share, we divide capital expenditures by total R&D expenditures in each field. Our assumption and the resulting estimated cost shares are consistent with some empirical findings, such as those from Czarnitzki et al. (2009), who use a Cobb–Douglas specification and find that the capital share in R&D among Belgian firms is less than 15%.

R&D. This suggests that, if deep learning could be similarly applied to a wide range of R&D problems, its high degree of capital intensity could accelerate technological change and, as a consequence, economic growth.

3. Data

In our work, we rely primarily on two datasets. Our primary dataset covers the compute cost and performance for 136 deep learning models that were presented in publications between 2012 and 2021. The second is a bibliometric dataset of the authors of machine learning publications from 1993 to 2021, which we use to infer the human capital inputs for each deep learning model in our primary dataset.

3.1. Data on computer vision experiments

Our dataset on the compute costs and performance covers 136 models published between 2012 and 2021. This data is an augmented version of that described in Thompson et al. (2020), with additional details about the settings under which the models were trained and tested, such as whether additional training data was used or whether the training or test data was augmented.

The compute estimates are derived from the underlying papers following the procedure described by Sevilla et al. (2022b), which we summarize in Appendix C. The inclusion and exclusion criteria used to generate our datasets is described in Thompson et al. (2020). Deep learning models in this dataset span two well-known benchmarks: image classification on the ImageNet dataset and object detection on the Microsoft COCO dataset (MS COCO).

ImageNet is perhaps the most well-known and widely used computer vision dataset. It spans 1000 object classes and contains 1.28 million training images (Russakovsky et al., 2015). Some of the most important breakthroughs in deep learning have happened in ImageNet models, starting with AlexNet, a watershed moment when deep learning first outperformed other techniques on this task (Krizhevsky et al., 2017). Importantly, success on ImageNet has often proven to be general: techniques that advance its state-of-the-art have usually been found to be successful in other tasks and domains. For example, Beyer et al. (2020) documents instances in which progress on ImageNet due to architecture design or optimization have yielded corresponding gains in other modalities, such as natural language processing, audio processing, and game playing. Thus, it is plausible that our results for this benchmark could generalize to tasks and domains beyond computer vision.

The MS COCO 2017 dataset is one of the most frequently used datasets for object detection, face detection, and pose estimation, among other tasks. It contains a total of 2.5 million labeled instances in 328,000 images (Lin et al., 2014). Like ImageNet, this dataset has been used as a test bed for many influential innovations, such as He et al. (2016)'s Residual Network architecture, which has since become widely used in computer vision (Khan et al., 2020).

While these two domains of computer vision are crucial test beds for deep learning, it would be better to consider a wider range of scientific and technical domains in which these techniques were applied. Unfortunately, challenges for both inputs and outputs make this difficult. For inputs, many deep learning papers fail to report even basic details of their computational usage. For outputs, some areas of deep learning struggle to define objective measures of performance. For example, it remains unclear how to define the "correct" text summary of a picture.[9]

[9] See Thompson et al. (2020) for a more in-depth discussion.

3.2. Data on authors and publications

Our dataset on machine learning publications comes from arXiv, a pre-print server commonly used in computer science, and Scopus, Elsevier's abstract and citation database. Our dataset includes all papers on arXiv posted between 1993 and 2021 from the subfields typically associated with machine learning: Machine Learning (stat.ML), Artificial Intelligence (cs.AI), Computation and Language (cs.CL), Computer Vision and Pattern Recognition (cs.CV), and Learning (cs.LG). We match the authors of these papers to their corresponding entries in Scopus using a variety of string distance-based matching approaches. This technique allows us to match 90.1% of authors, and spot testing on 300 random matches shows that 96% were correct.

With the connection between papers and authors' publication histories, we construct a time series for each author that shows their number of publications, h-index, and citations (excluding self-citations). We supplement this data with similar time series of grant funding for each author's institution and department over time from the Dimensions grant database, institutional rankings over time from csmetrics.org, and measures of the scientific influence of computer science journals and conferences from SCImago. For full details on data collection procedures, see Online Appendix C.

4. Empirical strategy

We assume, along the lines of the semi-endogenous growth model outlined above, that idea production using deep learning depends on three factors: labor (scientists' human capital), specialized capital goods (computational capital), and total factor productivity (the extant level of technology upon which researchers build):

$$\dot{A}(t) = A(t)^{\theta} S(t)^{\gamma} C(t)^{\beta}, \quad \text{where} \quad t > 0, \text{and} \quad X(0) > 0 \quad \text{for any} \quad X \in \{A, S, C\}. \tag{6}$$

where $\dot{A}(t)$ denotes the change in the stock of technology, $S(t)$ the total human capital input of scientists, and $C(t)$ refers to the total capital inputs. To estimate this, we replace $\dot{A}(t)$ with a measure of deep learning model performance, $C(t)$ with data on computational inputs, and $S(t)$ with estimates of scientific human capital inputs.[10]

4.1. Empirical specification

Consider an economy where the level of technology grows exponentially on the balanced growth path in the way commonly assumed in growth theory models:

$$A(t) = A(0)e^{gt}, \quad A(0) > 0. \tag{7}$$

We do not observe technology directly. Instead, we observe performance on machine learning tasks. In these cases, the level of perfor mance – usually measured as predictive accuracy on the test set for a given dataset – falls on the unit interval. We assume that performance relates to technology according to the logistic function, reflecting that the most challenging parts of innovation are being able to make some initial headway with a problem and then perfecting it,

$$P(t) = \frac{A(t)}{1 + A(t)}. \tag{8}$$

This is a similar assumption used to model how effort relates to various outcomes when the outcomes are bounded, such as in contests (Vojnović, 2015; Baik, 1998), conflict interactions (Hirshleifer, 1989; Jia et al., 2012), and persuasion (Skaperdas and Vaidya, 2012). Beyond that this is a relatively standard transformation to map progress in

[10] This model implies that the elasticity of substitution between computational and human capital is unitary. In Section 8.3, we estimate the relevant elasticity of substitution to validate this assumption.

technology onto a bounded interval, there are two further motivating considerations.

First, this functional form implies a power-law between the scale of the compute deployed and the level of error achieved, which is in line with a robust finding of the relevant machine learning literature (Hoffmann et al., 2022) (see Online Appendix A). Second, this functional form enables us to construct a simple empirical counterpart for technological progress, which we derive as follows. Assuming that growth rates in adjacent periods are approximately equal (i.e., $g_t \approx g_{t-1}$) it can be shown that proportional technological growth relates to performance improvements as follows (see Appendix A.4):

$$\frac{\dot{A}(t)}{A(t)} = \underbrace{\log\left(\frac{P(t)}{P(t-1)}\right)}_{\text{Proportional increase in accuracy}} + \underbrace{\log\left(\frac{1-P(t-1)}{1-P(t)}\right)}_{\text{Proportional reduction in error}}, \quad (9)$$

which provides an easy-to-interpret decomposition of technological progress in terms of the (logs of) the proportional reduction in error rate and the proportional increase in accuracy.

Let $\tilde{g}_t$ denote the approximation of g_t given $P(t)$, i.e. $\tilde{g}_t \equiv \log\left(\frac{P(t)}{P(t-1)}\right) + \log\left(\frac{1-P(t-1)}{1-P(t)}\right)$. We can write the empirical specification of our model as follows:

$$\tilde{g}_t = A(t)^{\theta-1} S(t)^{\gamma} C(t)^{\beta}. \quad (10)$$

We thus obtain an empirical specification of $\tilde{g}_t$ that we can ground in the relevant empirical data. When relating the model to data, time becomes discrete, and experiments are produced by research groups, which are indexed by $i \in \{1, \dots, N\}$.

Assuming a multiplicative error model, we specify the empirical counterpart of (10) as follows:

$$\tilde{g}_{it} = A_t^{\theta-1} S_{it}^{\gamma} C_{it}^{\beta} \epsilon_{it}, \quad \text{where} \quad \log \epsilon_{it} \equiv u_{it} \sim N(0, \sigma^2). \quad (11)$$

Taking logs of both sides, we have:

$$\log \tilde{g}_{it} + = (\theta - 1)\log A_t + \gamma \log S_{it} + \beta \log C_{it} + u_{it}. \quad (12)$$

We estimate the following model:

$$\begin{aligned}\log \underbrace{\tilde{g}_{it}}_{\text{Extent of tech. progress by } i} &= (\theta-1)\log \underbrace{A_t}_{\text{Extant level of technology}} + \gamma \log \underbrace{S_{it}}_{i\text{'s human capital}} \\ &+ \beta \log \underbrace{C_{it}}_{i\text{'s computational capital}} + \underbrace{\boldsymbol{\alpha}\,\mathbf{X}}_{\text{Vector of controls}} \\ &+ u_{it}.\end{aligned} \quad (13)$$

which we can estimate in a pooled fashion with a time-fixed effect that captures $(\theta - 1)A_t$ for $t \in \{1, \dots, T\}$. By default, we will fix the time periods as years. In the robustness checks section, we show that shorter or longer time windows do not change our overall results. The vector $\mathbf{X}$ includes variables on whether a model was trained on data in addition to our computer vision datasets of interest and whether a model was a reimplementation of a prior model.[11]

4.1.1. Non-rivalry of AI innovations

Our estimation procedure, derived from endogenous growth theory, assumes that knowledge is non-rivalrous. This assumption merits further discussion. It implies that researchers have access to the same stock of knowledge at any given time and that the advancements made by researchers in one period become available to others in the next. These assumptions seem reasonable in the context of machine learning research, given the field's open research norms. For example, researchers often publish their work on preprint servers like arXiv, share their code on platforms like GitHub and PaperswithCode, and use common tools like PyTorch that embed model implementations. These practices facilitate the dissemination and building upon of new knowledge in the field.

While some industry AI labs are becoming increasingly secretive about their work, resulting in longer delays between invention and diffusion, we believe the assumption of non-rivalrous knowledge remains reasonable. Even if certain cutting-edge advancements are kept confidential for a period, the underlying ideas and techniques often eventually make their way into the public domain. This can occur when the model is deployed to users, as information about its capabilities and behavior may leak. Additionally, employee turnover can lead to the dissemination of knowledge as individuals move between organizations. Furthermore, open-source efforts by the broader AI community can independently reproduce or reverse-engineer proprietary technologies.

Finally, trade secrets tend to be kept in areas of particularly high human capital (Biger and Plaut, 2000; Becker, 1962). Our model implicitly treats such knowledge as an additional benefit of human capital; therefore, the existence of trade secrets would make our results about capital intensity underestimates of the true level.

4.2. Operationalizing innovations

In estimating our model, we need to operationalize proportional performance improvements in terms of observables. We measure this using our baseline data, where authors of each paper have indicated the touchstone models in the literature whose ideas they are building on. For a given task, we define relative performance gains as follows:

$$P_t / P(t-1) \equiv P_t / \text{Baseline}_t. \quad (14)$$

That is, i's innovation is defined as the proportional improvement over the performance of a model that the contemporaneous literature considered to be a relevant baseline model. This operationalization is chosen for two reasons. First, it is common practice to report these values in the machine learning literature, as the extent of innovations are often illustrated through comparisons to existing baseline levels of performance (Armstrong et al., 2009; Melis et al., 2017; Pressel et al., 2018). Second, this notion of an improvement over a model lines up well with the usual notion of the change in stock of knowledge in R&D-based growth models (e.g., Romer, 1990 and Grossman and Helpman, 1994); it represents the extent of the innovation of a new design relative to the existing stock of ideas. To find the appropriate baseline levels of performance, we survey the models that are used as baseline results in the relevant literature and take the median of their performance (see Online Appendix C).

5. Modeling human capital

The final remaining task needed to estimate the deep learning production functions for these computer vision tasks is to construct a measure of the scientific human capital used to develop each model. This measure should be predictive of outcomes that are strongly influenced by human capital and must be inferred from available data about the scientists' track records. In addition to overall predictiveness, it is important to obtain unbiased estimates for junior researchers, who contribute importantly to this young field.

Prior work has used various measures of the quality or status of scientists and engineers, including impact-based metrics such as citation counts (Azoulay et al., 2019; Jones et al., 2014; Zucker et al., 2002) or the number of high-impact citations (Azoulay et al., 2014), and bibliometric indices such as the h-index (Teplitskiy et al., 2019; Fisman et al., 2018; Breschi et al., 2014). These approaches have substantial limitations as measures of the human capital of teams of scientists and engineers. As we will see, almost all of these measures are only

[11] For full details, see Appendix G.

weakly predictive of key outcomes for which we would expect scientific and technical human capital to be important, such as the number of citations the work will receive in the future, or the quality of the journal or conference the publication in which it will be published. Moreover, impact-based metrics, such as citation counts or the *h*-index, generally assign low scores to junior researchers, as citations often take a long time to accrue following scholarly publication.

Our strategy for modeling researcher human capital is as follows. We construct a deep neural network (DNN) and train it to develop a single-dimensional representation of the total quality-adjusted research input ("human capital") that is highly predictive of key bibliometric and publication-related outcomes. Our approach implements an encoder that maps many features about the publication's authors input to a single-dimensional representation, and a decoder model (built explicitly to have the function as a linear regression) that maps this representation onto citation- and publication-related outcomes. Our approach exploits the ability of DNNs for nonlinear data compression of high-dimensional input features (Hinton and Salakhutdinov, 2006; Kramer, 1991).[12] Our approach finds human capital representations for individual papers that are highly predictive of papers' key bibliometric and publication-related outcomes, and that substantially outperform the typical approaches used in the literature.

5.1. Our machine learning approach to estimating human capital

We use our dataset of 49,251 machine learning publications to train a neural network to predict bibliometric and publication-related outcomes. The predicted outcomes include the citation trajectories for each publication and its SJR-values, a measure of the quality of the journal or conference where the work ends up being published (González-Pereira et al. (2010)). Fig. 2a provides a diagrammatic overview of our data pipeline and the training set-up used to produce our model. We provide further details of the training procedure in Appendix D, and of the input data in Appendix E.

Our architecture is constructed as follows. We first stack 15 sets hidden layers, each consisting of a 4096 or 2048 node layer, followed by a batch-norm layer (Ioffe and Szegedy, 2015). These feed into a single unit — the "human capital" unit. This layer forces the neural network to reduce the dimensionality of its representations and distill the relevant features into a single scalar. The human capital result is then concatenated with the publication date, and fed into a series of independent sub-branches, one for each output being predicted. The final layer effectively implements separate linear regressions of the sort $y_i = \alpha + \mathbf{x}\beta$, meaning that the learned human-capital representations can only be linearly re-scaled and offset to make predictions about citations or journal quality.

In other words, our approach implements an encoder that maps the input (characteristics about the publication's authors) to a representation space, and a set of decoder regression models that map the representation onto citation- and publication-related outcomes. Thus, during training, the encoder is pushed to learn single-dimensional representations that are informative of human capital of the authors of each publication.

5.2. Validating our estimates

To assess the success of our measures in evaluating human capital, we compare their predictiveness across a range of outcomes, including citations received at various points and journal quality rankings. In each case, our estimates predict more than 55% of the variation in these measures, roughly 4–5 times as much as other proxies that are commonly used (see Fig. 3). In all cases, we are predicting out of sample on a test set of a random sub-sample of 4081 publications, which was held out from the training data.[13]

These results indicate our model has learned to predict bibliometric outcomes of publications, and in doing so, it has inferred meaningful and predictive human-capital features that can be measured as activation strength. Finally, when restricting the dataset to just publications with junior researchers (defined as authors with 2 or fewer prior publications), we find that our human capital estimates are still highly predictive of each of the bibliometric and publication-related outcomes (see Online Appendix D), while the predictive power of most other proxies is limited or negligible.

Our measure has demonstrated better predictive performance relative to other commonly-used human capital predictors. However, these other predictors also have access to much less data than our measure. For a more equal comparison, we compare our measure to a Lasso regression predictor – an approach representative of linear approaches used in the literature – with access to the same inputs as our neural network. To do so, we evaluate our DNN on an out-of-sample test set. Our DNN represents a substantial improvement relative to simpler approaches found in the literature that rely on linear combinations of impact-based metrics, such as the *h*-index, received publications, or publication counts. In particular, we obtain prediction errors (measured in mean-square-error) that are at least 40% lower for each outcome compared to Lasso regressions, and thus we get much more precise predictions, as shown in Fig. 3b.

While the preceding points to clear benefits with our approach, it is important to describe the limitations of the approach. First, our framework provides little insight on what a unit of human capital is (other than its predictive ability for some incremental improvement in relevant bibliometric outcomes), and therefore the cardinality of the estimates is difficult to interpret in terms of units of some latent input. Second, it is unclear how citations and journal quality relate to actual scientific merit, novelty, or insight, as bibliometric measures such as these are well-known to be imperfect proxies of such factors. Third, scientists' observed track records are in part determined by the resources they have access to, including computational resources and data. As a result, our measure of human capital can partially absorb the effect of having access to such resources, potentially overstating the returns to human capital and understating the returns to computational resources.[14]

6. Empirical analysis

Having validated our human capital measures, we combine them with the compute data to estimate production functions for two important AI tasks: image classification and object detection. In particular, we estimate individual regression models and a pooled model described by Eq. (13). We estimate using OLS, except where a Breusch–Pagan test indicates the presence of heteroskedasticity, in which case we estimate a GLS model by Maximum Likelihood. For each of computer vision (models A1–A2) and object detection (B1–B2), we estimate a model with our human capital and compute inputs, as well as binary variables for whether training used extra data not found in ImageNet or MS COCO and whether the model was a reimplementation of a prior model. In addition, models A2 and B2 have year fixed effects (details in Appendix G).

Our results for models A1–B2 are displayed in Table 1. We also estimate a pooled model with distinct time-fixed effects, which combines

[12] Similar bottlenecks are used for a variety of feature-learning tasks, such as through under-complete auto-encoders (e.g., Bengio et al., 2013).

[13] Of these publications, incoming citations after 1 year are known for 4035 publications, incoming citations after 2 years are known for 3724 publications, and the SJR values of the publication venues are known for 1312 publications.

[14] See Appendix F for plots of computational capital over time, and scatterplots between our two capital inputs and our measures of model performance.

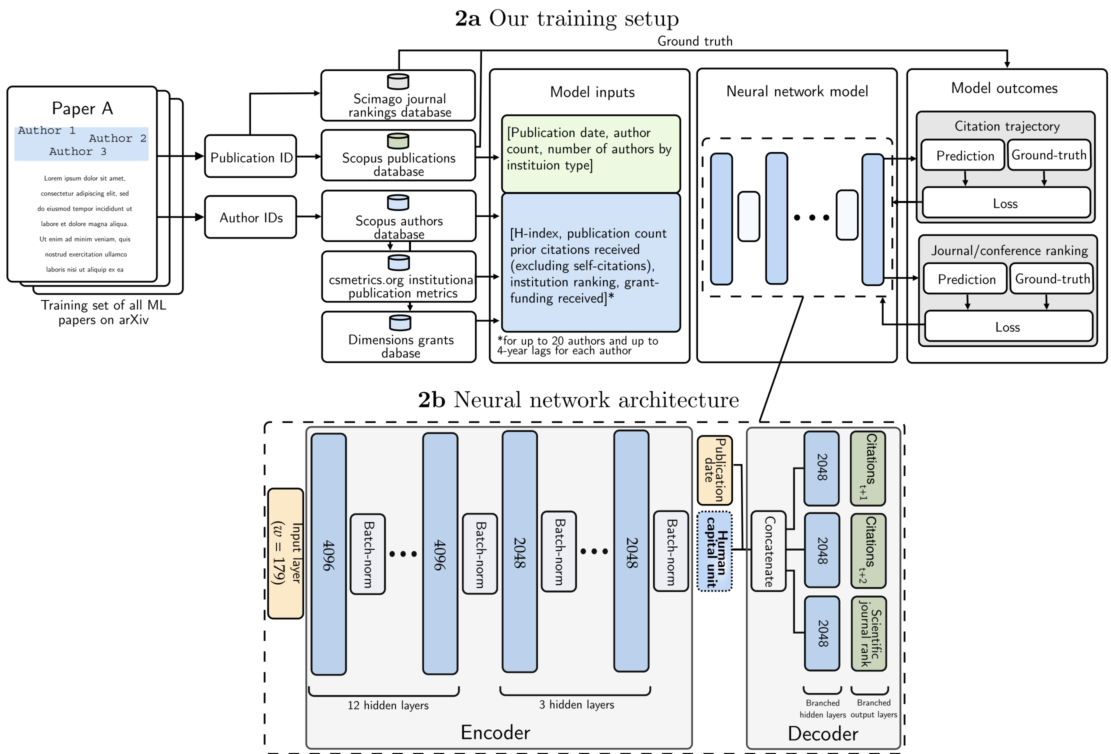
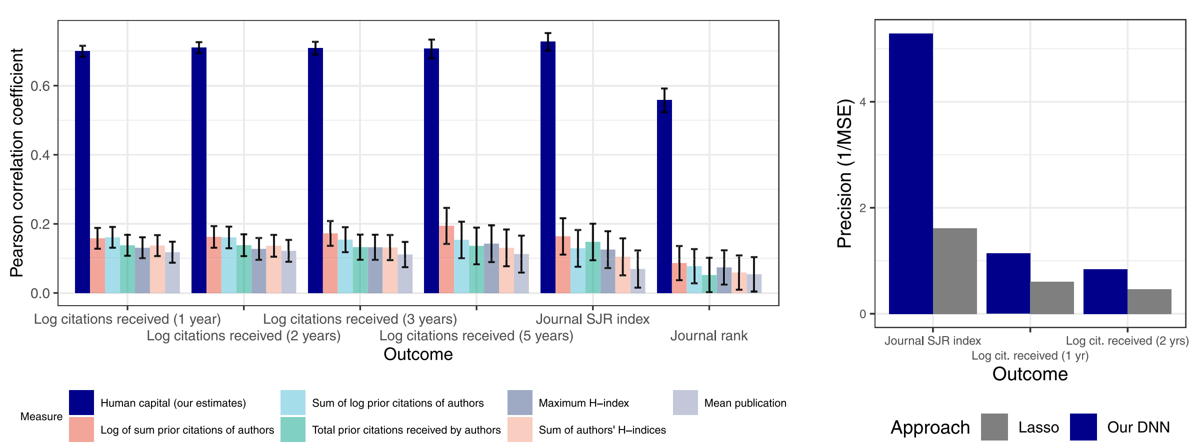


Fig. 2. Human capital estimation strategy. Fig. 2a presents our setup for learning human capital representations for machine learning publications. Fig. 2b shows our neural network architecture. Highlighted is the human capital unit, whose activations are strongly related to the quality of the research team. The numbers on each layer represent the number of units in that layer (for the human capital unit, this is just 1).



(a) Our human capital estimates predicts key outcomes much better than commonly used indicators

(b) Our DNN achieves better accuracy compared to separate lasso regressions

Fig. 3. Evaluating our human capital measure. Fig. 3(a). Correlations between human capital measures and publication outcomes for a hold-out set of 4081 publications. Error bars indicate the 95 confidence interval. Prior citations are the cumulative total citations received by the authors up until the year prior to publishing the relevant publication (excluding self-citations). H-indices and publication counts are evaluated at the year of publication for each author. Journal rank here represents the ordinal value of each journal in descending order of SJR-value. Fig. 3(b) shows the precision (defined as 1/MSE) of predictions of our DNN-based model and Lasso regressions on the same hold-out set of 4081 publications estimated separately for each outcome.

data across both computer vision tasks. A likelihood ratio test indicates that the pooled model fits the data better than separately estimated models. The estimates of models C1–C2 are displayed in Table 2.

For image classification, we estimate the R&D elasticity of capital (β) to be 0.111 (model A1) and 0.140 (model A2), as shown in Table 1. This means that a 1% increase in the computational capital used for this type of R&D is associated with a 0.111–0.140% increase in the rate of technological change. For object detection, we estimate β is 0.246 (for both model B1 and B2), considerably higher than for image classification. For our pooled estimates, we obtain estimates between 0.145 and 0.176 (see Table 2). All of these results are statistically significant at the 5% level; most are also significant at the 0.1% level.

Our estimates of the R&D elasticity of human capital (γ) show less variation between the two deep learning tasks. Our human capital

Table 1
Deep learning production function estimates. Estimation results for image classification ($n = 96$) and object detection ($n = 40$). *, **, ***denote p<0.05, p<0.01, p<0.001 respectively.

Data	Model	Estimates			Log likelihood
		R&D elasticity to capital (β)	R&D elasticity to human capital (γ)	Trend	
Image classification	A1	0.111 *** (0.021)	0.246 *** (0.086)	–	−48.798
	A2	0.140 *** (0.029)	0.350 *** (0.085)	0.051 *** (0.014)	−39.787
Object Detection	B1	0.246 * (0.106)	0.352 * (0.165)	–	−43.256
	B2	0.253 ** (0.090)	0.319 (0.158)	0.013 (0.030)	−43.187

Table 2
Pooled deep learning production function estimates. Estimation results for pooled computer vision experiments (n = 136). **, ***denote p<0.05, p<0.01, p<0.001 respectively.

Data	Model	Estimates			Log likelihood
		R&D elasticity to capital (β)	R&D elasticity to human capital (γ)	Trend	
Computer Vision (pooled)	C1	0.145 *** (0.022)	0.278 ** (0.088)	–	−106.552
	C2	0.176 *** (0.017)	0.298 *** (0.076)	0.032 *** (0.004)	−95.586

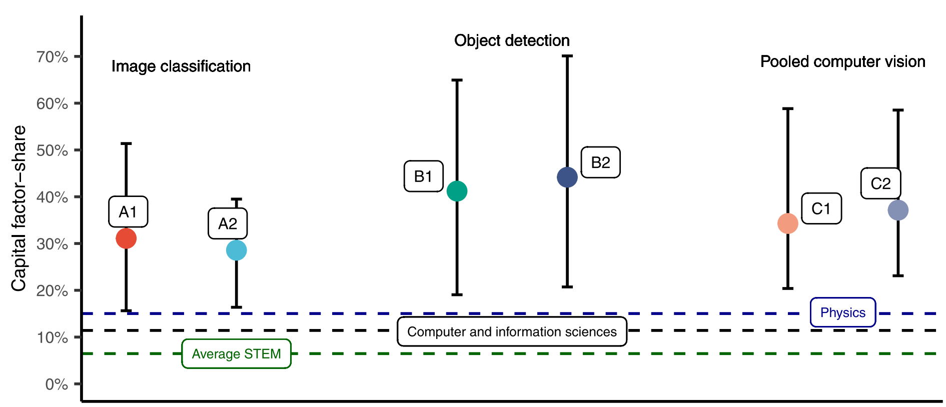


Fig. 4. Implied optimal R&D expenditure breakdown. Implied capital-cost shares given the estimates presented in tables Table 1 and Table 2, computed as $\hat{\beta}/(\hat{\beta}+\hat{\gamma})$. Error bars represent 90% confidence intervals generated by bootstrapping 10,000 iterations. We use the bias-corrected percentile method for bootstrapping confidence intervals for ratios outlined in Campbell and Torgerson (1999).

elasticity estimates for image classification are 0.246 (model A1) and 0.350 (model A2), both significant at the 0.1% level. These estimates are just over twice as high as those for computational capital. For object detection tasks, the estimates of γ are similar at 0.352 (model B1) and 0.319 (model B2), though only the former of these two estimates is significant at the 5% level. For our pooled model, we find estimates of the R&D elasticity of human capital is 0.278 (model C1) and 0.298 (model C2), each statistically significant at the 1% level.

Using our growth model from Section 2.2, we can directly infer the equilibrium cost shares dedicated to capital and labor from the relevant elasticities (assuming a competitive market).[15] We find that the implied capital-cost estimates range from 0.29 and 0.44 (see Fig. 4). Our estimates indicate that a firm in the R&D sector should allocate between 29% and 44% of its total expenditure on computational capital. Using our confidence intervals generated by bootstrapping, we find that implied capital cost estimates that are statistically significantly greater than 0.15 at the 5% level for all models.

We find that the implied capital cost share for object detection is higher than for image classification by a margin of roughly 10 percentage points. However, this difference is not statistically significant. Overall, we see that the implied R&D capital shares for AI are substantially higher than in other areas of U.S. science and engineering Section 2.2, in which the capital share generally falls below 20%.

7. R&D with deep learning

Having estimated the capital intensity of R&D that is augmented with AI, we analyze the potential productivity effects that the widespread adoption of deep learning would have on economic productivity and growth. To do so, we suppose that the arrival of deep learning would act as a one-time shock, raising the capital intensity of knowledge production in the economy to the levels we estimated in computer vision.

Along the balanced growth path, the steady-state growth rate in the stock of knowledge is described by Eq. (4). We substitute in the parameter values given by our empirical estimates above into our semi-endogenous growth model and compute the predicted change in R&D productivity growth conditional on the widespread of adoption of deep learning. As shown in Fig. 5, depending on model specification, the results from image classification imply a productivity growth rate between 1.6% and 1.8%, whereas those from object detection imply a rate between 3.1% and 3.9%. Our preferred estimate is the pooled

[15] See Appendix A.3.

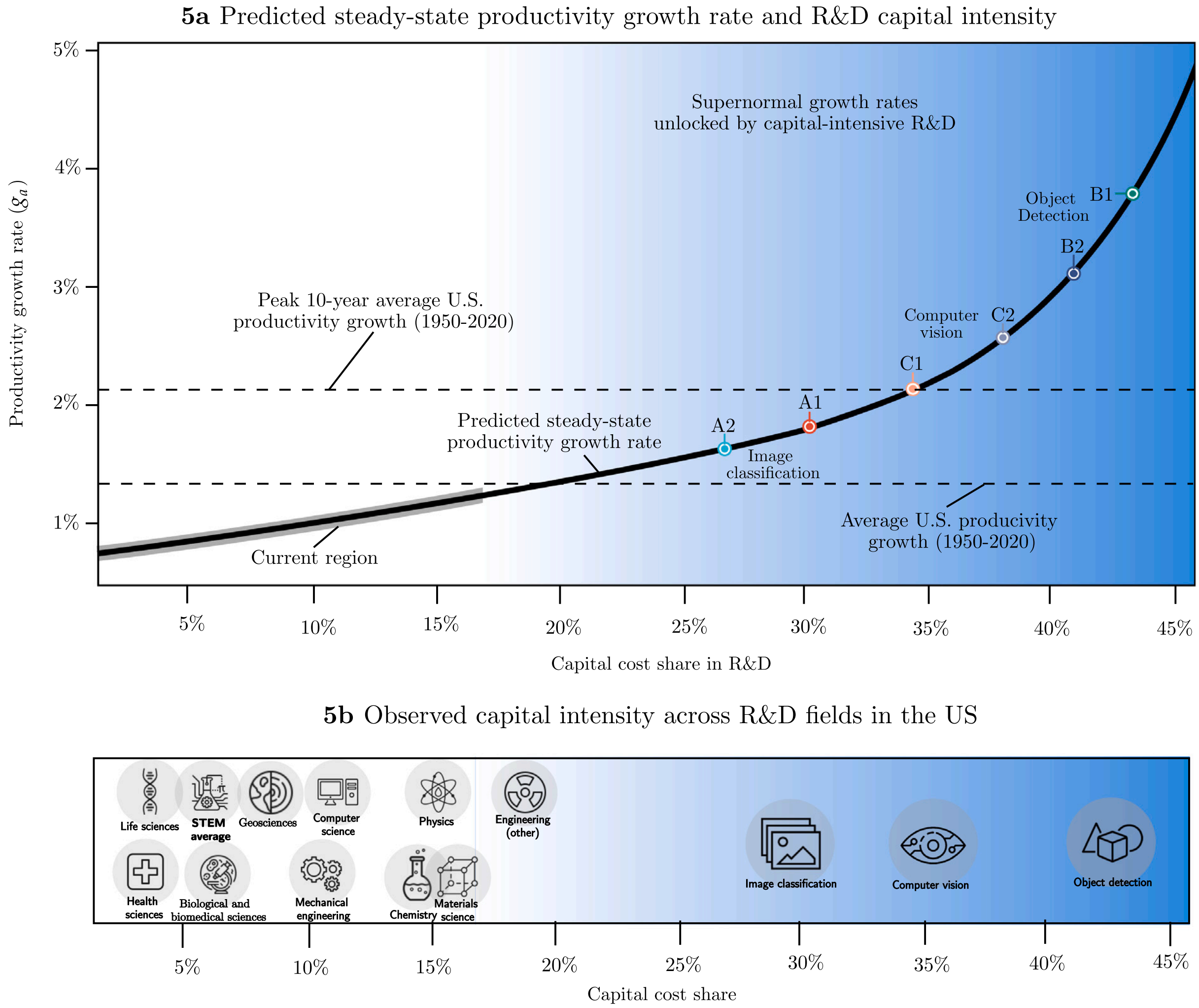


Fig. 5. Figure 5. Predicted productivity growth under widespread deployment of AI in R&D. Steady-state productivity growth as a function of the implied capital factor share in the R&D sector with competitive factor markets. Returns to scale are held constant. For this plot, we assume that the elasticity of R&D output to the stock of ideas (θ) is 1/2 and the elasticity of R&D output to labor inputs (γ) is 2/5. Markers indicate point estimates of implied optimal R&D expenditure with deep learning according to models A1–C2 as estimated in Section 6. "Current region" indicates the current level of capital intensity of R&D according to NSF data, which our semi-endogenous growth model predicts to result in 0.5% to 1.3% productivity growth, a level consistent with observed recent US productivity growth.

estimate for computer vision, as it is informed by more data and provides more precise estimates. With the widespread adoption of deep learning raising ideas production in the economy to this level of capital intensity, we would expect the productivity growth rate to rise to between 2.1% and 2.4%. In context, this would amount to an increase between 1.7 and 2-fold relative to the 1.2% average U.S. productivity growth from 1948 to 2021, and a 2.6 to 3-fold increase compared to the post-2000 0.8% growth (San Francisco Fed, 2022). Thus, our results indicate that if adopting deep learning in other areas of R&D allows those areas to leverage capital better in the same way that computer vision has, it will represent a substantial acceleration of scientific progress.[16]

[16] While our estimates are for fundamental computer vision research, applied areas of deep learning use the same procedure to train models (Reed et al., 2022). While production function estimates for downstream tasks are scarce, a recent study of machine learning in finance (Abis and Veldkamp, 2024) finds that an even smaller share of R&D progress is driven by human capital.

8. Discussion

In this section, we show that the results in Section 7 are robust to outliers and the granularity of time periods. Moreover, we show our assumption about the substitutability of labor and capital is consistent with our data, supporting the calculation of the capital intensity of deep learning derived from our estimates. Finally, we discuss the generalizability of our estimates from computer vision to other R&D tasks and the validity of our assumption about competitive factor markets.

8.1. Sensitivity of estimates to outliers

Certain empirical results from high-profile studies can be reversed by removing less than 1% of the sample even when standard errors are small (Broderick et al., 2020). In this section, we assess the sensitivity of our results to outliers by showing that the removal of a small fraction of the data is not determinative of our findings. To test the robustness of our results to the removing of samples, we re-run our analysis between around 10^4 and 10^6 times (depending on which model is re-estimated)

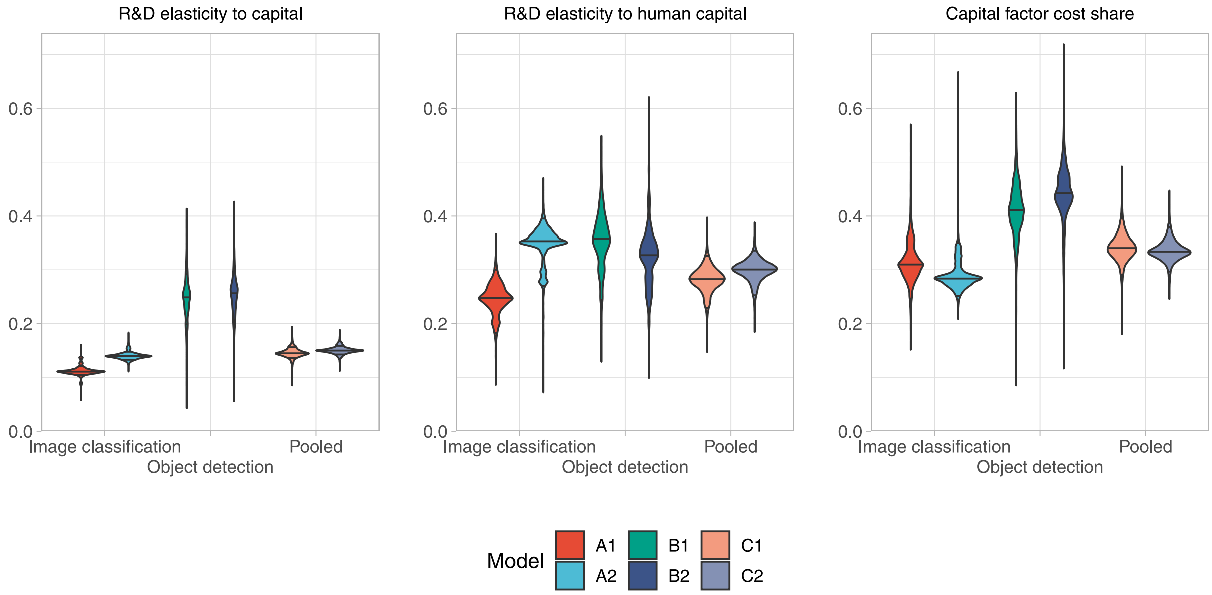


Fig. 6. Estimates after removing a random 5% of dataset Median estimates when taking a random sub-sample of our dataset that excludes 1.6% to 5% of the total number of observations. Results are displayed as violin plots using kernel density estimation to create the distributions. Inside the violins, the box plots show median and interquartile ranges.

on random sub-samples of our datasets that excludes a fraction of our observations. The point estimates are plotted in Fig. 6.

For our dataset on object detection, we estimated models on all possible sub-samples that exclude 3 observations (of which there are $\frac{40!}{37!3!} = 9880$). For our dataset on image classification and for our pooled dataset, we estimated models on 10^6 random sub-samples that exclude 5 observations. The sub-samples considered covers around 1.6% of all possible sub-samples for the image classification dataset and 0.3% of all sub-samples for the pooled dataset and is, therefore, a non-trivial fraction of all possible permutations.

We find that the point estimates are largely robust to the removal of any small subset of observations, as we see that most estimates are tightly clustered around their median value, particularly for the estimates for which our datasets are the largest (image classification and our pooled dataset). Moreover, the point estimates are consistent with the estimates found in our baseline empirical analysis presented in Section 6, providing evidence that our estimates do not result from a small number of outliers.

8.2. Alternative model specifications

Our estimation strategy for A_t (the stock of knowledge) takes advantage of cross-sectional variation at each time point. That is, we effectively pool publications into groups of contemporaries published around the same time. We then estimate the variation in performance due to changes in inputs among these contemporaries, supposing that this variation is due to inputs rather than changes in A_t. However, papers are published continuously over time, so this necessarily involves a bias–variance trade-off: specifying more granular time periods (months instead of years) increases variance, but each interval is estimated with fewer data points, making it noisier and more prone to over-fitting. In our empirical analysis, we balanced this trade-off by fixing time periods as yearly intervals. Now we show that our conclusions are robust to different reasonable choices of the granularity of periods.

We re-estimate models A1–C2 with window lengths 6 and 18 months and compare our estimates to those obtained in Section 6 (12 months). We find that the estimates are largely similar for all relevant datasets, as shown in Table 3 for estimates for models A1–C1 and Appendix H for re-estimates of models A2–C2. Moreover, similar patterns remain: estimates of the R&D elasticity to capital (β) are relatively lower for image classification tasks than for object detection tasks. These estimates strongly suggest that our key estimates are robust to the specification of time window length.

8.3. Sensitivity to model assumptions about the substitutability of human scientists and competitive factor markets

In our semi-endogenous growth model, we assume production is Cobb–Douglas, so that the elasticity of substitution equals 1. Hence, we assume a substantial level of substitutability of human scientists for compute. In this section, we test whether our data is consistent with a higher or lower level of substitutability.

The substitutability assumption is important because our semi-endogenous growth model implies that compute stock will grow faster than the stock of scientists. To see this, recall that from Eq. (4), the steady-state growth of capital dedicated to R&D is $\frac{1-\theta+\gamma}{1-\beta-\theta}n$. With the estimates from , this would be $\sim 2n$, while the stock of scientists grows just at rate n. Hence, we should expect that, in the long run, the stock of specialized capital goods, $C(t)$, will be substantially greater than that of scientists' human capital $S(t)$.[17]

If scientists were more difficult to substitute with compute than we have assumed, the optimal investment path might involve larger investments in scientists. To see this, suppose the idea production function instead followed a more general constant elasticity of substitution (CES) production function:

$$\dot{A}(t) = A(t)^{\theta}[\gamma S(t)^{\frac{\sigma-1}{\sigma}} + \beta C(t)^{\frac{\sigma-1}{\sigma}}]^{\frac{\sigma}{\sigma-1}}, \tag{15}$$

where $\sigma \in \mathbb{R}_+$ denotes the elasticity of substitution between compute and scientists. In this framework, assuming a competitive R&D sector, the share of expenditure dedicated to compute (which we will denote by f) is given by:

$$f = \frac{\beta C(t)^{\frac{\sigma-1}{\sigma}}}{\beta C(t)^{\frac{\sigma-1}{\sigma}} + \gamma S(t)^{\frac{\sigma-1}{\sigma}}}. \tag{16}$$

From this expression, we can see that in the Cobb–Douglas case (where $\sigma = 1$), $f = \beta/(\beta+\gamma)$. Note moreover, that $f \geq \beta/(\beta+\gamma)$ if and only if

[17] This conclusion is consistent with what we see in our empirical data as well as what has been found in other areas of computing (Thompson et al., 2022).

Table 3
Estimation results for separate models with alternative window lengths. Estimates of models A1, B1, and C1 with different window lengths. Specifications are the same as in the main analysis in Section 6.

Data	Time-period length	Estimates: R&D elasticity to capital (β)	Estimates: R&D elasticity to human capital (γ)	Log likelihood
Image classification	6	0.099 *** (0.017)	0.250 ** (0.088)	−47.379
	12	0.111 *** (0.021)	0.246 * (0.086)	−48.799
	18	0.155 *** (0.029)	0.332 *** (0.089)	−75.380
Object detection	6	0.215 * (0.095)	0.400 ** (0.139)	−43.256
	12	0.246 * (0.096)	0.352 * (0.165)	−43.256
	18	0.214 (0.111)	0.348 (0.204)	−44.030
Computer vision (pooled)	6	0.150 *** (0.016)	0.294 *** (0.072)	−74.276
	12	0.132 *** (0.020)	0.245 ** (0.075)	−87.288
	18	0.142 *** (0.020)	0.218 ** (0.082)	−95.839

$\sigma \geq 1$. Hence, if scientists are less easily substitutable than we supposed, the capital intensity of R&D will be lower than our estimates imply.

We investigate our assumption that $\sigma = 1$ by estimating Eq. (15). Note that we can rewrite (15) by dividing by $A(t)$ as follows:

$$g_t = A_t^{\theta-1}[\gamma S_t^{\rho} + \beta C_t^{\rho}]^{\frac{1}{\rho}}, \tag{17}$$

where $\rho \equiv \frac{\sigma-1}{\sigma}$. To simplify the estimation procedure, we approximate this expression using the second-order McLaurin expansion (i.e., the Taylor series evaluated at $\sigma = 1$),

$$\log g_t \approx (\theta - 1)\log A_t + \gamma \log S_t + \beta \log C_t + \frac{1}{2}\rho\gamma\beta[\log S_t - \log C_t]^2 \tag{18}$$

This is the translog production function, which has the well-known advantage that it is linear in its parameters and, therefore, able to be estimated using OLS. Because of this, this approximation is widely used in similar settings (Guilkey et al., 1983; Berndt and Christensen, 1973). In our case, the empirical model we estimate becomes:

$$\log \tilde{g}_{it} = (\theta-1)\log A_t + \gamma \log S_{it} + \beta \log C_{it} + \frac{1}{2}\rho\gamma\beta[\log S_{it} - \log C_{it}]^2 + \epsilon_{it}, \tag{19}$$

where each variable has its usual meanings it had in Section 4. Fig. 7 plots the distributions generated by bootstrapping estimates of σ in Eq. (19) for each of our models A1–C2. Our estimates of σ are tightly clustered around unity. These findings reinforce our assumption that $\sigma = 1$ is reasonable and, therefore, our inferences about capital intensity are consistent with the data. Moreover, for the tasks for which we have the most data, we observe that our estimates σ are slightly above 1. Hence, for these tasks, if anything, our inferred level of capital intensity of deep learning-based R&D is an underestimate of the true level after adequately accounting for the degree of substitutability of scientists.

8.4. Limits to external validity stemming from our choice of domain

Our analysis focuses on two computer vision tasks. Although this is a small segment of scientific and engineering problems that deep learning might be applied to, there are good reasons to expect that the key insights gained from studying a wide range of architectures for computer vision could apply to a wider range of R&D problems.

One reason is that deep learning often builds on common techniques, algorithms, and architectures across different subfields (see, e.g., Goodfellow et al., 2016, Ch. 3). Across many domains, deep learning systems are based on similar ideas and algorithms. Almost all modern deep learning systems, regardless of modality or task, are some "deep" computational graph with many parameters that are learned through gradient descent along gradients of some loss function computed by backpropagation. Indeed it is widely considered (e.g., by Alom et al., 2018) that one of the first models to utilize all of these key features of deep learning AlexNet in 2012, a model in our dataset.

There exist some pronounced architectural divides across domains. For example, convolutional neural networks are widespread in computer vision, while transformer-based models are ubiquitous in machine translation. However, these architectural differences are also represented in our data, which includes convolutional neural networks, vision transformers, and architectures based purely on multi-layer perceptrons. Hence, our data reflects much of the variation between modalities and tasks in modern deep learning.

In addition, some neural network architectures, like the transformer, have been shown to operate effectively across domains and tasks (see, e.g., Reed et al., 2022), including as state-of-the-art techniques across many R&D-adjacent tasks like code generation (Li et al., 2022), cheminformatics (Irwin et al., 2022), and bioinformatics (Elnaggar et al., 2020). This suggests that our findings may generalize to domains outside of computer vision.

Moreover, work on neural scaling laws suggests that the relation between the model's size and performance scales according to the usual power-law independently of the domain the model is trained to handle. Henighan et al. (2020) find that, for transformer models, there is a remarkable near-universal relation between the optimal model size and compute budget across a range of domains including images, language, mathematics, and video, which supports the notion the role of compute depends strongly on technique or architecture, not the domain in which these techniques are applied. For these reasons, we expect the key insights gained from studying a range of architectures for computer vision to broadly carry over to a wider range of R&D problems to which deep learning is applied.

8.5. How data availability influences our estimates

Another reason to expect that our results may fail to apply to a broader set of problems is that we study a set of problems within computer vision with a relative abundance of quality labeled data. In some domains – such as the problem of protein folding, where the crystal structures of proteins are expensive and arduous to generate – high-quality data might be less readily available or expensive to generate. As a result of the relative abundance of data in computer vision, the returns to compute might be higher than they would be in lower-data-abundance regimes. To illustrate this, consider a simple model where deep learning system performance can be described as a

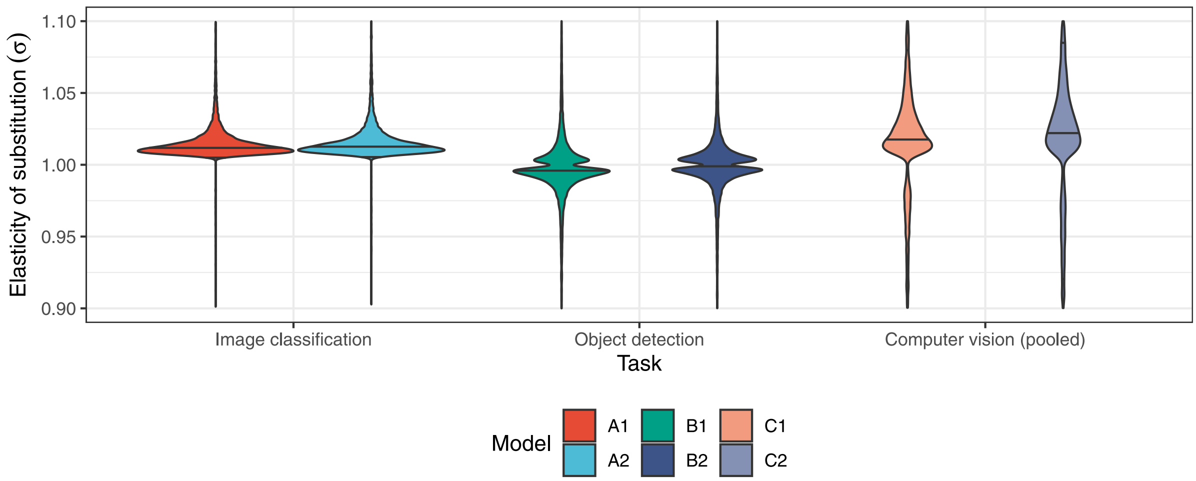


Fig. 7. Elasticity of substitution estimates Kernel densities of the estimates of σ across each of our main models generated by bootstrapping 10,000 iterations.

CES production function in data D and compute C:

$$P = \left[\zeta D^{\frac{\sigma-1}{\sigma}} + \eta C^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{1-\sigma}}, \quad \text{where} \quad \sigma, \quad \zeta, \quad \eta > 0. \tag{20}$$

where $\sigma \in \mathbb{R}_+$ is the elasticity of substitution between the inputs, and ζ and η are the returns to scale of data and compute, respectively. It becomes clear whenever data and computation are gross complements ($\sigma < 1$), then, the returns to compute will be lower in low-data regimes than in high-data regimes:

$$\begin{aligned}\left.\frac{\partial P}{\partial C}\right|_{D_{\text{Low}}} &= \eta C^{\frac{\sigma-1}{\sigma}-1}\left[\zeta D_{\text{Low}}^{\frac{\sigma-1}{\sigma}} + \eta C^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{1-\sigma}-1} \leq \left.\frac{\partial P}{\partial C}\right|_{D_{\text{High}}} \\ &= \eta C^{\frac{\sigma-1}{\sigma}-1}\left[\zeta D_{\text{High}}^{\frac{\sigma-1}{\sigma}} + \eta C^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{1-\sigma}-1}.\end{aligned} \tag{21}$$

This suggests the returns to compute in high-data regimes can be unusually high. Moreover, empirical evidence has shown that AI-based ideas production is rapidly expanding in data-rich sectors like investment management (Abis and Veldkamp, 2024), and that computer vision firms with access to additional data are more innovative (Beraja et al., 2023). Might our results fail to generalize to low-data regimes? We argue they will likely generalize, particularly for economically important R&D tasks.

One piece of evidence comes from work on neural scaling laws for language modeling. Hoffmann et al. (2022) derive a parametric loss function in which the amount of data and the model size enters additively separably. Although this result comes from language modeling, where data is abundant, that firms try to hold training data as trade secrets suggests they believe data can substitute for compute to train improved models.

Moreover, we might expect that, for economically important R&D tasks, complementary investments in generating the necessary datasets to train machine learning models will be made, including in high-fidelity physical simulations, high-quality synthetic datasets, and additional sensors. Hence, for economically-important R&D tasks, we anticipate low-data regimes to be short-lived.

Finally, while increasingly large compute and data budgets might increase barriers to deep learning's adoption for smaller or less productive firms, this barrier has been declining. Algorithmic progress in many areas of deep learning has been rapidly advancing (Ho et al., 2024), driving down the cost of capital necessary to achieve a given level of task performance. If enough of these advances are widely shared or easily replicable, the fixed costs of deep learning should decline over time, increasing adoption.

8.6. *Lack of competitive factor markets*

A strong assumption our model makes is that input markets are competitive. While a common assumption in macroeconomic production function estimation (e.g., Doraszelski and Jaumandreu, 2018), in the short run, it is unlikely to hold in the GPU market. From 2019 to 2023, for example, NVIDIA had a market share above 80% for datacenter GPUs, indicating a significant case of market power (Mujtaba, 2023). As a result, additional demand for GPUs could result in higher compute prices, biasing our estimates and lowering the proliferation of deep learning.

While this is an important area for future research, we argue that our results should obtain in the long run, as the higher-than-average returns obtained by GPU firms should induce new firms to enter, driving prices closer to marginal cost. This is the logic behind Schumpeterian creative destruction (Schumpeter, 1942) and underlies models of endogenous growth (Grossman and Helpman, 1991; Romer, 1990).[18]

9. Conclusion

Our main contributions are threefold. First, we develop a framework to analyze the impact of two key trends: recent breakthroughs in applying deep learning to R&D, and the rapid increase in computational resources used by deep learning systems. We demonstrate that if deep learning is widely adopted in R&D and significantly increases the returns to computational capital, then under certain conditions, the rate of technological progress will permanently increase.

Second, we introduce a novel machine learning approach for estimating human capital from scientific publications. Our method employs an encoder to compress information about authors into a low-dimensional latent representation. The encoder is trained to learn human capital representations that best predict publication and citation metrics, which we argue are indicative of the quality of human capital. Our human capital estimates predict key outcomes 4–5 times more accurately than commonly used methods in the literature.

Third, using data from two crucial computer vision benchmarks, we estimate that deep learning R&D is more capital-intensive than other forms of R&D, suggesting that a profit-maximizing firm would allocate 29%–44% of its total R&D budget to computational capital. This finding implies that if deep learning becomes prevalent in R&D and enhances

[18] We discuss policy implications of violations of our assumptions in Online Appendix E.

the returns to computational capital, the pace of technological advancement will permanently accelerate, causing a 1.7 to 2-fold increase in the rate of economic growth.

CRediT authorship contribution statement

Tamay Besiroglu: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Nicholas Emery-Xu:** Conceptualization, Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Neil Thompson:** Conceptualization, Data curation, Methodology, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

https://futuretech.mit.edu/community-resources

Acknowledgments

The authors are grateful to Open Philanthropy for financial support. Nicholas Emery-Xu is also grateful to the UCLA Graduate Division and the Future of Humanity Institute for financial support. For helpful comments we are grateful to Philip Trammell, Basil Halperin, Matt Clancy, Charlotte Siegmann, Ege Erdil, Gabriel Filipe, Wensu Li, Tom Davidson, Jaime Sevilla, and Nur Ahmed.

Appendix A. Derivations and mathematical results

A.1. Deriving steady-state growth rates

Using Eqs. (1)–(3), we can derive the rates at which capital and ideas grow:

$$g_k(t) \equiv \frac{\dot{K}(t)}{K(t)} = c_k \left[\frac{A(t)L(t)}{K(t)}\right]^{1-\alpha} - \delta, \quad \text{where} \quad c_k \equiv s(1-\alpha_k)^{\alpha}(1-\alpha_l)^{1-\alpha}, \tag{22}$$

$$g_a(t) \equiv \frac{\dot{A}(t)}{A(t)} = c_a K(t)^{\beta} L(t)^{\gamma} A(t)^{\theta-1}, \quad \text{where} \quad c_a \equiv B\alpha_k^{\beta}\alpha_l^{\gamma}. \tag{23}$$

Along the balanced growth path (defined as an equilibrium path where $Y(t)$, $K(t)$, $A(t)$ and $L(t)$ grow at a constant rate), it can be shown that:

$$\tilde{g}_k(t) \equiv \frac{\dot{g}_k(t)}{g_k(t)} = (1-\alpha)(g_a+n-g_k), \quad \text{and} \quad \tilde{g}_a(t) \equiv \frac{\dot{g}_a(t)}{g_a(t)} = \beta g_k + \gamma n - (1-\theta) g_a. \tag{24}$$

The steady-state rates of growth in ideas and capital can then simply be found by solving for $g_k(t)$ and $g_a(t)$ that solves $\tilde{g}_k(t) = \tilde{g}_a(t) = 0$. Solving this system yields the following equilibrium growth rates (equilibrium growth rates are marked with the * superscript):

$$g_a^* = \frac{\beta+\gamma}{1-\beta-\theta}n, \quad \text{and} \quad g_k^* = \frac{1-\theta+\gamma}{1-\beta-\theta}n, \tag{25}$$

which can be shown to be unique and stable.[19]

A.2. Proof of Proposition 1b

For part (b) of Proposition 1, we wish to show that the steady-state rate of economic growth is strictly increased under the relevant shift. This happens if both the steady-state rates of idea and capital accumulation are permanently increased when $\Delta_\beta \geq -\Delta_\gamma$. Hence, to show this, it remains to show that the steady-state rate of capital accumulation is permanently increased. This is true if:

$$\frac{1-\theta+\gamma+\Delta_\gamma}{1-\beta'-\Delta_\beta-\theta}n > \frac{1-\theta+\gamma}{1-\beta-\theta}n \tag{26}$$

Rearranging, this holds when:

$$-\Delta_\gamma < \left(\frac{1-\theta+\gamma}{1-\beta-\theta}\right)\Delta_\beta \tag{27}$$

Since $\frac{1-\theta+\gamma}{1-\beta-\theta} > 1$, this holds whenever $\Delta_\beta \geq -\Delta_\gamma > 0$, yielding the desired result. ■

A.3. Capital cost share in competitive R&D sector

Given perfect competition in the R&D sector, total expenditure on wages and rents (Lw and Kr, respectively) are given by:

$$Lw = L\frac{\partial \dot{A}(t)}{\partial L} = \gamma \dot{A}(t), \quad \text{and} \quad Kr = K\frac{\partial \dot{A}(t)}{\partial L} = \beta \dot{A}(t). \tag{28}$$

It follows that the capital cost share $Kr/(Kr+Lw)$ is given by $\beta/(\beta+\gamma)$.

A.4. Empirical specification

To see that:

$$\frac{\dot{A}(t)}{A(t)} \approx \log\left(\frac{P(t)}{P(t-1)}\right) + \log\left(\frac{1-P(t-1)}{1-P(t)}\right), \tag{29}$$

note first that $\dot{A}(t)/A(t) = g_t$. Moreover note that:

$$\frac{P(t)}{1-P(t)}\frac{1-P(t-1)}{P(t-1)} = \frac{\exp(g_t t)}{\exp(g_{t-1}(t-1))} \approx \exp(g_t). \tag{30}$$

Hence $\log\left(\frac{P(t)}{P(t-1)}\right) + \log\left(\frac{1-P(t-1)}{1-P(t)}\right) \approx g_t$ whenever $g_t \approx g_{t-1}$.

Appendix B. Background on benchmarks in machine learning

Experimental benchmarks are a core feature of machine learning research. Benchmarks, particular combinations of a dataset or sets of datasets, have long been used to perform experimental validation of new techniques (Hothorn et al., 2005). It is common that the significance of new techniques are explicitly linked to benchmark performance, which often proxy the amount of progress made (Martinez-Plumed et al., 2021; Raji et al., 2021).

Benchmarks are a tool that enables the comparative assessment of machine learning techniques. Consider the well-known computer vision benchmark ImageNet (Deng et al., 2009). ImageNet is a set of over 1.2 million images that belong to 1000 mutually exclusive classes.[20] On the ImageNet Image Classification benchmark, deep learning models are trained to predict probability distributions over the classes of each image. Techniques are evaluated on a distinct test set consisting of 100,000 images, and typically evaluated on the basis of some top-k error (the rate at which the 'true' classes are not among the top k highest-probability classes predicted by the model). A reduction in the error rate, then, represents some notion of progress on this particular task.[21]

[19] This is a simple extension of the usual results for semi-endogenous growth models (Romer (2012, Chapter 3)).

[20] By "ImageNet", we refer to ImageNet-1k, the most highly-used subset of the Scale Visual Recognition Challenge (ILSVRC) image classification and localization dataset, not the larger ImageNet-21K dataset.

[21] We note that benchmark experiments have important limitations in assessing innovations, as discussed in the machine learning literature (Raji et al., 2021; Recht et al., 2019; Picard, 2021).

Appendix C. Estimating compute

We use two methods from Sevilla et al. (2022b) to infer the amount of compute used to train an AI system. These methods (one based on the architecture of the network and number of training batches processed, and another based on the hardware setup and amount of training time) are outlined below.

C.1. Method 1 — counting operations in the model

The first method can be summarized as:

$$\text{Training compute} = (\text{FLOP per forward pass} + \text{FLOP per backward pass}) \times \text{Nr. of passes}, \tag{31}$$

where "FLOP per forward pass" is the number of floating point operations in a forward pass, "FLOP per backward pass" is the number of operations in a backward pass and "Nr. of passes" is the number of full passes (a full pass includes both the backward and forward pass) made during training.

Moreover, we use the facts that:

$$\text{Nr. of passes} = \text{Nr. of epochs} \times \text{Nr. of examples}, \tag{32}$$

$$\text{FLOP per forward pass} + \text{FLOP per backward pass} \approx 3 \times \text{FLOP per forward pass}, \tag{33}$$

where the latter is implied by the fact that computing the backward pass requires each layer to compute the gradient with respect to the weights and the error gradient of each neuron with respect to the layer input to backpropagate. Each of these operations requires compute roughly equal to the amount of operations in the forward pass of the layer. Hence, the total number of FLOP per forward and backward pass is roughly 3-fold the number of FLOP per forward pass. If the number of examples is not directly stated, it can be computed as the number of batches per epoch times the size of each batch:

$$\text{Nr. of examples} = \text{Nr. of batches} \times \text{batch size}. \tag{34}$$

C.2. Method 2—GPU time

Second, we can use the reported training time and GPU model performance to estimate training compute. For example, if the training lasted 2 days and used a total of 5 GPUs, that equals 10 GPU-days. By multiplying the number of GPU-days by the performance of the GPU, we can infer the amount of compute in FLOP needed to train the model. In particular, we use the formula:

$$\text{Training compute} = \text{Training time (in seconds)} \times \text{Nr. of cores} \times \text{Peak FLOP/s} \times \text{utilization rate}, \tag{35}$$

where peak performance in FLOP/s is found in the relevant GPU product datasheets, and the utilization rate corrects for imperfect hardware utilization, for which 30% is often a reasonable baseline (Sevilla et al., 2022b).

Appendix D. Training procedure and hyper-parameter settings

The weights in the first 12 layers are frozen, and the weights in the remaining components are randomly initialized and trained on a subset of publications for which the 3 output features are known (∼13k publications). Finally, we transfer this model to our dataset of publications of 136 machine learning models, which are excluded from our training set, and inspect the activations of the human capital unit.

To train our DNN, we use the adamax optimizer (Kingma and Ba, 2014) with $\beta_1 = 0.9$, $\beta_2 = 0.999$. Throughout, use the GELU activation function (Hendrycks and Gimpel, 2016), with the exception of the Human Capital unit, which is ReLU-activated (see Table 4).

Appendix E. Input data

We train our DNN to predict a variety of bibliometric-related and publication-related outcomes based on a large number of author- and publication-level features, which were encoded as a 269 × 1 input vector. The input-layer features are described in Table 5.

If a publication has more than 20 authors, we select data for the first 15 authors and the last 5 authors, as authors in computer science are commonly ordered in descending order of share of contributions, with the exception of supervisory or senior authors, who are listed last. This ensures that we include both the subset of authors who might have made the largest contributions and those who had a supervisory role.

Appendix F. Inputs and performance

The computer vision models in our dataset are trained on a wide range of compute budgets: 4 orders of magnitude for image classification models and 2 orders of magnitude for object detection models. Training compute is highly variable yet displays steady growth in the most compute-intensive models (see Fig. 8). The estimated doubling times of computation used in the most compute-intensive models not trained on extra data are 8.72 months [95% CI: 2.22 to 34.33 months] for image classification models and 8.96 months [95% CI: 2.48 to 32.35 months] for object detection models. This growth rate in compute intensity is consistent with previous estimates of recent compute trends (Sevilla et al., 2022a).

Fig. 9 plots model performance against both training compute and human capital, both of which we see are associated with improved model performance. For each machine learning task, we find roughly a power-law relationship between compute and error rates, in line with the experimental results in the literature (Kaplan et al., 2020; Hestness et al., 2017). Additionally, there is positive correlation between performance and the estimated amount of human capital, though it appears much weaker than that between computational capital and performance.

Appendix G. Regression model specifications and estimation procedures

Recall that our empirical estimation is given by:

$$\log \tilde{g}_i = \boldsymbol{\phi}'\mathbf{1} + \gamma \log S_i + \beta \log C_i + \boldsymbol{\alpha}\mathbf{X} + u_i, \quad \text{where} \quad \boldsymbol{\phi} \equiv (1-\theta)\begin{bmatrix} \log A_1 \\ \log A_2 \\ \vdots \\ \log A_T \end{bmatrix}.$$

Table 6 specifies each of the models we estimate in Section 6. For each, if a Breusch–Pagan test detects heteroskedasticity, we employ the Generalized Least Squares (GLS) method. GLS adjusts the weight of each observation according to its variance, enhancing the reliability of parameter estimates. To accurately estimate variance components and ensure the robustness of our regression model, we utilize the Restricted Maximum Likelihood (REML) approach (see Table 7).

Appendix H. Supporting empirical results

We re-estimate models A2, B2, and C2 with window lengths 6 and 18 months, and compare our estimates to those obtained in Section 6 (12 months). We find the estimates are mostly similar for all datasets (see Table 8).

Table 4
Training steps and settings. Output layer refers to the outcomes that the neural network was trained to predict ($t+k$ refers to citations k-years post-publication, and SJR refers to the publication's SJR value). N refers to the number of observations that we trained on for each step. which is equal to the number of examples for which we had all relevant observations for in our training set.

Pre-training step	Output layer	Batch size	Learning rate	Epochs	N
1	t+1	1024	5×10^{-4}	90	44 562
2	t+2	1024	5×10^{-4}	90	41 120
3	t+1, t+2, t+3, t+5	248	8×10^{-4}	90	14 217
4	t+1, t+2, t+3, SJR	248	5×10^{-4}	90	8978
5	t+1	500	5×10^{-4}	90	44 562
6	t+1, t+2, SJR	500	8×10^{-4}	90	13 157
7	t+1, t+2	500	8×10^{-4}	90	41 120
8	t+1, t+2, SJR	500	8×10^{-4}	90	13 157

Table 5
Features in input layer. Variables that make up our input vector. All variables are evaluated at the time of publication. Each input vector is of the dimension 1 × 179.

Variable	Included for which authors	Lags
h-index	Up to 20 authors	4 years
Number of publications	Up to 20 authors	4 years
Number of citations received (excluding self-citations)	Up to 20 authors	4 years
Total grant funding received by authors' institution	All	5 years
Ranking of institution	Up to 20 authors	–
Number of authors by institution type (academic, govm, company)	All	–
Number of authors on publication	–	–
Publication date	–	–

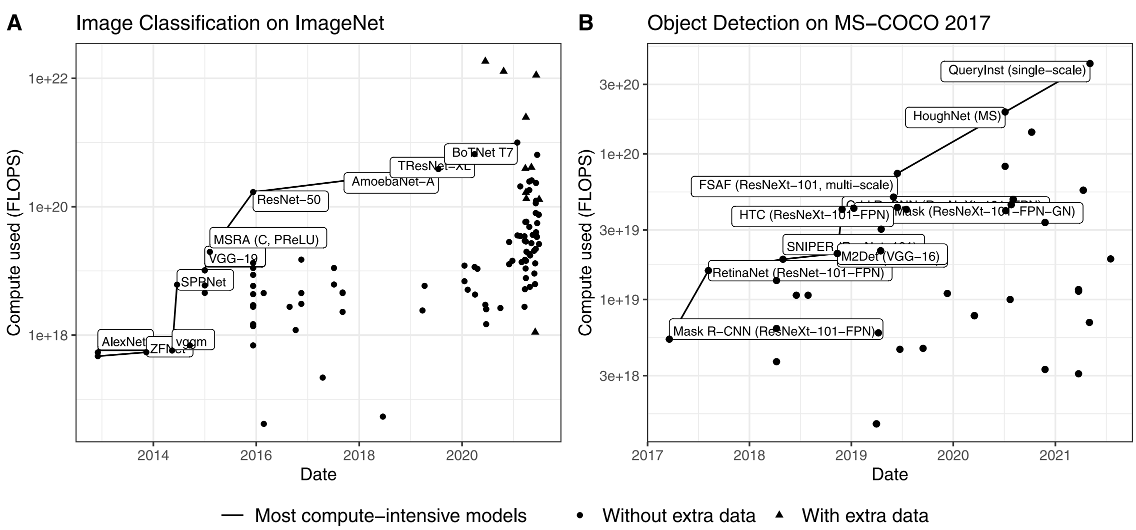


Fig. 8. Compute intensity of training runs over time for each benchmark. Note that the y-axis is logarithmic ($\log_{10}$). Solid line traces the most compute-intensive models that were trained without extra data beyond the usual training set.

Table 6
Full model specifications. ϕ' denotes the transpose of ϕ. reimplementation$_i$ is equal to 1 if the model i is a re-implementation of prior work, and a 0 otherwise. extra data$_i$ is equal to 1 if model i was trained using extra data (other than the ImageNet training set), and 0 otherwise (this includes any pre-training of the model, including any of its backbones). The trend coefficients reported in our empirical results correspond to the estimates of the coefficient denoted α_3 in this table.

Model	Dataset	Specification
A1	Image classification	$\log \tilde{g}_i = \boldsymbol{\phi}'\mathbf{1} + \gamma \log S_i + \beta \log C_i + \alpha_1 \text{extra data}_i + \alpha_2 \text{reimplementation}_i$
A2	Image classification	$\log \tilde{g}_i = \boldsymbol{\phi}'\mathbf{1} + \gamma \log S_i + \beta \log C_i + \alpha_1 \text{extra data}_i + \alpha_2 \text{reimplementation}_i + \alpha_3 \text{years from 2012}_i$
B1	Object detection	$\log \tilde{g}_i = \boldsymbol{\phi}'\mathbf{1} + \gamma \log S_i + \beta \log C_i + \alpha_1 \text{extra data}_i + \alpha_2 \text{reimplementation}_i$
B2	Object detection	$\log \tilde{g}_i = \boldsymbol{\phi}'\mathbf{1} + \gamma \log S_i + \beta \log C_i + \alpha_1 \text{extra data}_i + \alpha_2 \text{reimplementation}_i + \alpha_3 \text{years from 2012}_i$
C1	All computer vision	$\log \tilde{g}_i = \boldsymbol{\phi}'\mathbf{1} + \gamma \log S_i + \beta \log C_i + \alpha_1 \text{extra data}_i + \alpha_2 \text{reimplementation}_i$
C2	All computer vision	$\log \tilde{g}_i = \boldsymbol{\phi}'\mathbf{1} + \gamma \log S_i + \beta \log C_i + \alpha_1 \text{extra data}_i + \alpha_2 \text{reimplementation}_i + \alpha_3 \text{years from 2012}_i$

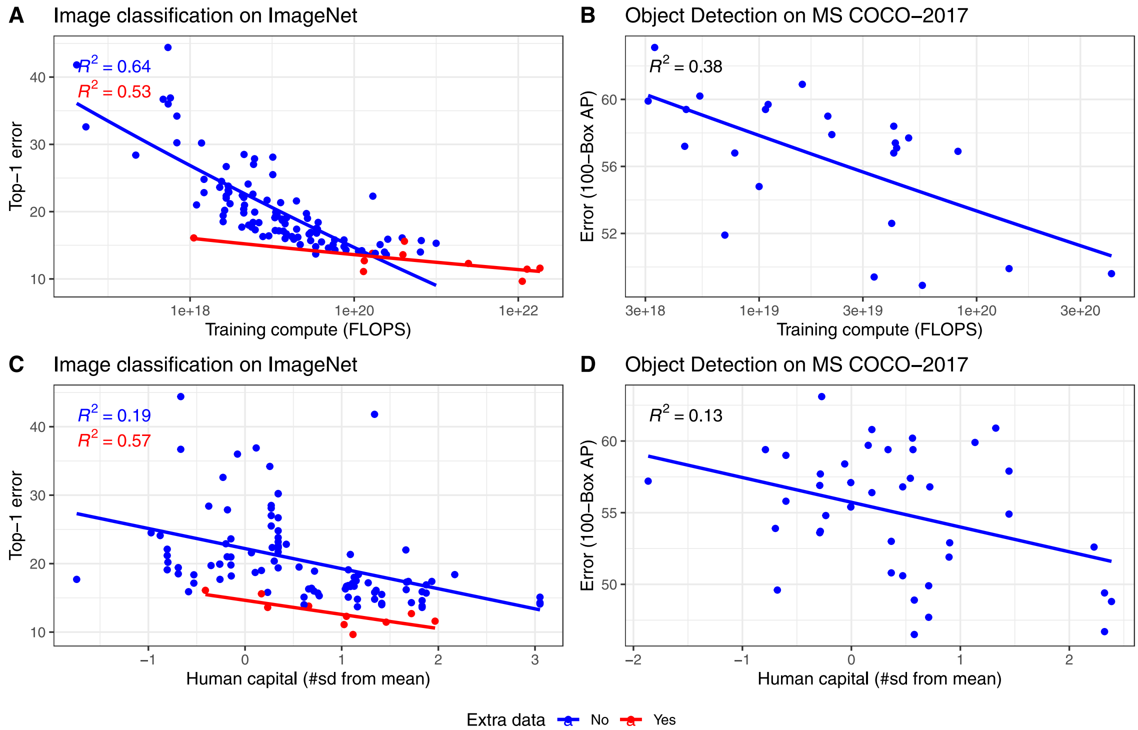


Fig. 9. Associations between performance and training compute (top panel) and human capital (bottom panel).

Table 7
Estimation techniques used. GLS refers to Generalized Least Squares, and OLS (HC) refers to Ordinary Least Squares, with robust covariance matrix estimators with a degrees of freedom correction $(n-1)/(n-k)$ where n is the number of observations and k is the number of explanatory or predictor variables in the model. We use GLS when, after performing a Breusch–Pagan test for heteroskedastic errors, we reject the null of no heteroskedasticity at a 5% significance level.

Time-period length	Model					
	A1	A2	B1	B2	C1	C2
6	GLS Table 6	GLS Table 12	OLS (HC) Table 6	OLS (HC) Table 12	GLS Table 6	OLS (HC) Table 12
12	GLS Tables 4, 6	OLS (HC) Tables 4, 6	OLS (HC) Tables 4, 6	OLS (HC) Tables 4, 6	OLS (HC) Tables 5, 6	OLS (HC) Tables 5, 6
stmrowsep[3pt] 18	OLS (HC) Table 6	OLS (HC) Table 12	OLS (HC) Table 6	OLS (HC) Table 12	OLS (HC) Table 6	OLS (HC) Table 12

Table 8
Estimation results for separate models with alternative window lengths. Estimates of models A2, B2, and C2 with different window-lengths. Specifications are the same as in the main analysis in Section 6.

Data	Time-period length	Estimates	Estimates		Log likelihood
		R&D elasticity to capital (β)	R&D elasticity to human capital (γ)	Trend	
Image classification	6	0.100 *** (0.017)	0.211 * (0.087)	0.065 (0.029)	−47.958
	12	0.140 *** (0.025)	0.350 ** (0.114)	0.052 *** (0.014)	−39.786
	18	0.154 *** (0.021)	0.184 (0.121)	0.042 ** (0.015)	−68.709
Object detection	6	0.210 * (0.089)	0.424 * (0.159)	−0.014 (0.048)	−27.169
	12	0.253 ** (0.090)	0.319 (0.158)	0.013 (0.030)	−43.187
	18	0.236 * (0.109)	0.376 * (0.197)	0.054 * (0.022)	−41.862
Computer vision (pooled)	6	0.130 *** (0.015)	0.274 *** (0.072)	0.050 * (0.020)	−76.372
	12	0.142 *** (0.016)	0.263 *** (0.068)	0.030 *** (0.005)	−80.164
	18	0.144 *** (0.018)	0.207 ** (0.079)	0.020 *** (0.004)	−98.355

Appendix I. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.respol.2024.105037.

References

Abis, S., Veldkamp, L., 2024. The changing economics of knowledge production. Rev. Financ. Stud. 37 (1), 89–118.

Aghion, P., Jones, B.F., Jones, C.I., 2019. Artificial intelligence and economic growth. In: The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, pp. 237–282.

Agrawal, A., 2022. Introduction by ajay agrawal. In: NBER Economics of AI Conference 2017. URL https://www.economicsofai.com/nber-conference-toronto-2017.

Agrawal, A., Gans, J., Goldfarb, A., 2018. Prediction machines: the simple economics of artificial intelligence. Harvard Business Press.

Agrawal, A., McHale, J., Oettl, A., 2019. Finding needles in haystacks: Artificial intelligence and recombinant growth. In: The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, pp. 149–174.

Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Van Esesn, B.C., Awwal, A.A.S., Asari, V.K., 2018. The history began from AlexNet: A comprehensive survey on deep learning approaches. arXiv preprint arXiv:1803.01164.

Armstrong, T.G., Zobel, J., Webber, W., Moffat, A., 2009. Relative significance is insufficient: Baselines matter too. In: Proceedings of the SIGIR 2009 Workshop on the Future of IR Evaluation. pp. 25–26.

Azoulay, P., Fons-Rosen, C., Graff Zivin, J.S., 2019. Does science advance one funeral at a time? Amer. Econ. Rev. 109 (8), 2889–2920.

Azoulay, P., Stuart, T., Wang, Y., 2014. Matthew: Effect or fable? Manage. Sci. 60 (1), 92–109.

Bahri, Y., Dyer, E., Kaplan, J., Lee, J., Sharma, U., 2021. Explaining neural scaling laws. arXiv preprint arXiv:2102.06701.

Baik, K.H., 1998. Difference-form contest success functions and effort levels in contests. Eur. J. Political Econ. 14 (4), 685–701.

Becker, G.S., 1962. Investment in human capital: A theoretical analysis. J. Political Econ. 70 (5, Part 2), 9–49.

Belkin, M., Hsu, D., Ma, S., Mandal, S., 2018. Reconciling modern machine learning practice and the bias-variance trade-off. arXiv preprint arXiv:1812.11118.

Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. 35 (8), 1798–1828.

Beraja, M., Yang, D.Y., Yuchtman, N., 2023. Data-intensive innovation and the state: Evidence from AI firms in China. Rev. Econ. Stud. 90 (4), 1701–1723.

Berndt, E.R., Christensen, L.R., 1973. The translog function and the substitution of equipment, structures, and labor in U.S. manufacturing 1929-68. J. Econometrics 1 (1), 81–113. http://dx.doi.org/10.1016/0304-4076(73)90007-9, URL https://www.sciencedirect.com/science/article/pii/0304407673900079.

Beyer, L., Hénaff, O.J., Kolesnikov, A., Zhai, X., Oord, A.v.d., 2020. Are we done with imagenet? arXiv preprint arXiv:2006.07159.

Bianchini, S., Muller, M., Pelletier, P., 2020. Deep learning in science. arXiv:2009.01575[cs, econ].

Biger, N., Plaut, S.E., 2000. Trade secrets, firm-specific human capital, and optimal contracting. Eur. J. Law Econ. 9, 51–73.

Bolanos, F., Salatino, A., Osborne, F., Motta, E., 2024. Artificial intelligence for literature reviews: Opportunities and challenges. arXiv preprint arXiv:2402.08565.

Breschi, S., Lissoni, F., Tarasconi, G., et al., 2014. Inventor data for research on migration and innovation: a survey and a pilot, vol. 17, WIPO.

Broderick, T., Giordano, R., Meager, R., 2020. An automatic finite-sample robustness metric: When can dropping a little data make a big difference? arXiv preprint arXiv:2011.14999.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al., 2020. Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901.

Campbell, M.K., Torgerson, D.J., 1999. Bootstrapping: estimating confidence intervals for cost-effectiveness ratios. Qjm 92 (3), 177–182.

Carter, S., Nielsen, M., 2017. Using artificial intelligence to augment human intelligence. Distill http://dx.doi.org/10.23915/distill.00009, https://distill.pub/2017/aia.

Cockburn, I.M., Henderson, R., Stern, S., 2019. The impact of artificial intelligence on innovation: An exploratory analysis. In: The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, pp. 115–148. http://dx.doi.org/10.7208/9780226613475-006.

Crafts, N., 2021. Artificial intelligence as a general-purpose technology: an historical perspective. Oxford Rev. Econ. Policy 37 (3), 521–536. http://dx.doi.org/10.1093/oxrep/grab012.

Czarnitzki, D., Kraft, K., Thorwarth, S., 2009. The knowledge production of 'R' and 'D'. Econom. Lett. 105 (1), 141–143. http://dx.doi.org/10.1016/j.econlet.2009.06.020, URL https://www.sciencedirect.com/science/article/pii/S0165176509002298.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 248–255.

Doraszelski, U., Jaumandreu, J., 2018. Measuring the bias of technological change. J. Polit. Econ. 126 (3), 1027–1084.

Elnaggar, A., Heinzinger, M., Dallago, C., Rihawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., Steinegger, M., et al., 2020. ProtTrans: towards cracking the language of life's code through self-supervised deep learning and high performance computing. arXiv preprint arXiv:2007.06225.

Fisman, R., Shi, J., Wang, Y., Xu, R., 2018. Social ties and favoritism in Chinese science. J. Polit. Econ. 126 (3), 1134–1171.

Gibney, E., 2017-10-18. Self-taught AI is best yet at strategy game go. Nature http://dx.doi.org/10.1038/nature.2017.22858, URL https://www.nature.com/articles/nature.2017.22858, Publisher: Nature Publishing Group.

Goldfarb, A., Taska, B., Teodoridis, F., 2022. Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings. NBER working paper, (w29767).

González-Pereira, B., Guerrero-Bote, V.P., Moya-Anegón, F., 2010. A new approach to the metric of journals' scientific prestige: The SJR indicator. J. Inf. 4 (3), 379–391.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT Press.

Grossman, G.M., Helpman, E., 1991. Quality ladders in the theory of growth. Rev. Econ. Stud. 58 (1), 43–61.

Grossman, G.M., Helpman, E., 1994. Endogenous innovation in the theory of growth. J. Econ. Perspect. 8 (1), 23–44.

Guilkey, D.K., Lovell, C.A.K., Sickles, R.C., 1983. A comparison of the performance of three flexible functional forms. Internat. Econom. Rev. 24 (3), 591–616. http://dx.doi.org/10.2307/2648788, URL https://www.jstor.org/stable/2648788.

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction, vol. 2, Springer.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Helmers, C., Overman, H.G., 2017. My precious! the location and diffusion of scientific research: evidence from the synchrotron diamond light source. Econ. J. 127 (604), 2006–2040.

Hendrycks, D., Gimpel, K., 2016. Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415.

Henighan, T., Kaplan, J., Katz, M., Chen, M., Hesse, C., Jackson, J., Jun, H., Brown, T.B., Dhariwal, P., Gray, S., et al., 2020. Scaling laws for autoregressive generative modeling. arXiv preprint arXiv:2010.14701.

Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., Patwary, M., Ali, M., Yang, Y., Zhou, Y., 2017. Deep learning scaling is predictable, empirically. arXiv preprint arXiv:1712.00409.

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313 (5786), 504–507.

Hirshleifer, J., 1989. Conflict and rent-seeking success functions: Ratio vs. difference models of relative success. Public Choice 63 (2), 101–112.

Ho, A., Besiroglu, T., Erdil, E., Owen, D., Rahman, R., Guo, Z.C., Atkinson, D., Thompson, N., Sevilla, J., 2024. Algorithmic progress in language models. arXiv preprint arXiv:2403.05812.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D.d.L., Hendricks, L.A., Welbl, J., Clark, A., et al., 2022. Training compute-optimal large language models. arXiv preprint arXiv:2203.15556.

Hothorn, T., Leisch, F., Zeileis, A., Hornik, K., 2005. The design and analysis of benchmark experiments. J. Comput. Graph. Statist. 14 (3), 675–699.

Howitt, P., 1999. Steady endogenous growth with population and R & D inputs growing. J. Polit. Econ. 107 (4), 715–730. http://dx.doi.org/10.1086/250076, URL https://www.journals.uchicago.edu/doi/abs/10.1086/250076.

Howitt, P., Aghion, P., 1998. Capital accumulation and innovation as complementary factors in long-run growth. J. Econ. Growth 3 (2), 111–130. http://dx.doi.org/10.1023/A:1009769717601.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. PMLR, pp. 448–456.

Irwin, R., Dimitriadis, S., He, J., Bjerrum, E.J., 2022. Chemformer: a pre-trained transformer for computational chemistry. Mach. Learn.: Sci. Technol. 3 (1), 015022.

Jia, H., Skaperdas, S., et al., 2012. Technologies of conflict. In: The Oxford Handbook of the Economics of Peace and Conflict. Oxford University Press New York, pp. 449–472.

Jones, C.I., 1995. R & D-based models of economic growth. J. Polit. Econ. 103 (4), 759–784. http://dx.doi.org/10.1086/262002, Publisher: The University of Chicago Press, URL https://www.journals.uchicago.edu/doi/abs/10.1086/262002.

Jones, A.L., 2021. Scaling scaling laws with board games. arXiv preprint arXiv:2104.03113.

Jones, B., Reedy, E., Weinberg, B.A., 2014. Age and scientific genius. tech. rep., National Bureau of Economic Research.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., ÅœÃdek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596 (7873), 583–589. http://dx.doi.org/10.1038/s41586-021-03819-2, URL https://www.nature.com/articles/s41586-021-03819-2.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., Amodei, D., 2020. Scaling laws for neural language models. arXiv:2001.08361.

Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S., 2020. A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. 53 (8), 5455–5516.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kramer, M.A., 1991. Nonlinear principal component analysis using autoassociative neural networks. AIChE J. 37 (2), 233–243.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60 (6), 84–90. http://dx.doi.org/10.1145/3065386, URL https://dl.acm.org/doi/10.1145/3065386.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lepikhin, D., Lee, H., Xu, Y., Chen, D., Firat, O., Huang, Y., Krikun, M., Shazeer, N., Chen, Z., 2020. Gshard: Scaling giant models with conditional computation and automatic sharding. arXiv preprint arXiv:2006.16668.

Li, Y., Choi, D., Chung, J., Kushman, N., Schrittwieser, J., Leblond, R., Eccles, T., Keeling, J., Gimeno, F., Lago, A.D., et al., 2022. Competition-level code generation with alphacode. arXiv preprint arXiv:2203.07814.

Li, Z., Wallace, E., Shen, S., Lin, K., Keutzer, K., Klein, D., Gonzalez, J.E., 2020. Train large, then compress: Rethinking model size for efficient training and inference of transformers. arXiv preprint arXiv:2002.11794.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In: European Conference on Computer Vision. Springer, pp. 740–755.

Martinez-Plumed, F., Barredo, P., Heigeartaigh, S.O., Hernandez-Orallo, J., 2021. Research community dynamics behind popular AI benchmarks. Nat. Mach. Intell. 3 (7), 581–589.

Melis, G., Dyer, C., Blunsom, P., 2017. On the state of the art of evaluation in neural language models. arXiv preprint arXiv:1707.05589.

Mujtaba, H., 2023. GPU market rebounds in Q2 2023: AMD, NVIDIA, and intel see increased shipments, discrete GPU up by 12.4%. URL https://wccftech.com/gpu-market-rebounds-q2-2023-amd-nvidia-intel-increased-shipments-discrete-gpus-up/.

Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., Sutskever, I., 2021. Deep double descent: Where bigger models and more data hurt. J. Stat. Mech. Theory Exp. 2021 (12), 124003.

National Center for Science and Engineering Statistics, 2021. Higher education research and development: Fiscal year 2020: Data tables. Higher Education Research and Development: Fiscal Year 2020: Data Tables.

Peng, S., Kalliamvakou, E., Cihon, P., Demirer, M., 2023. The impact of ai on developer productivity: Evidence from github copilot. arXiv preprint arXiv:2302.06590.

Picard, D., 2021. Torch. manual_seed (3407) is all you need: On the influence of random seeds in deep learning architectures for computer vision. arXiv preprint arXiv:2109.08203.

Pressel, D., Lester, B., Choudhury, S.R., Barta, M., Zhao, Y., Hemmeter, A., 2018. Baseline: Strong, extensible, reproducible, deep learning baselines for NLP.

Raji, I.D., Bender, E.M., Paullada, A., Denton, E., Hanna, A., 2021. AI and the everything in the whole wide world benchmark. arXiv preprint arXiv:2111.15366.

Recht, B., Roelofs, R., Schmidt, L., Shankar, V., 2019. Do imagenet classifiers generalize to imagenet? In: International Conference on Machine Learning. PMLR, pp. 5389–5400.

Reed, S., Zolna, K., Parisotto, E., Colmenarejo, S.G., Novikov, A., Barth-Maron, G., Gimenez, M., Sulsky, Y., Kay, J., Springenberg, J.T., et al., 2022. A generalist agent. arXiv preprint arXiv:2205.06175.

Romer, P.M., 1990. Endogenous technological change. J. Polit. Econ. 98 (5), S71–S102. http://dx.doi.org/10.1086/261725, Publisher: The University of Chicago Press, URL https://www.journals.uchicago.edu/doi/abs/10.1086/261725.

Romer, D., 2012. Advanced Macroeconomics, 4e. McGraw-Hill, New York.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vision (IJCV) 115 (3), 211–252. http://dx.doi.org/10.1007/s11263-015-0816-y.

Russell, S., Norvig, P., 2020. Artificial intelligence: a modern approach, forth ed..

San Francisco Fed, 2022. Total factor productivity. URL https://www.frbsf.org/economic-research/indicators-data/total-factor-productivity-tfp/.

Schumpeter, J.A., 1942. Capitalism, socialism and democracy. routledge.

Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbhahn, M., Villalobos, P., 2022a. Compute trends across three eras of machine learning. arXiv preprint arXiv:2202.05924.

Sevilla, J., Heim, L., Ho, A., Hobbhahn, M., Besiroglu, T., Villalobos, P., 2022b. Estimating training compute of Deep Learning models. Tech. rep..

Sharma, U., Kaplan, J., 2020. A neural scaling law from the dimension of the data manifold. arXiv preprint arXiv:2004.10802.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., et al., 2017. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. arXiv preprint arXiv:1712.01815.

Skaperdas, S., Vaidya, S., 2012. Persuasion as a contest. Econ. Theory 51 (2), 465–486.

Sourati, J., Evans, J.A., 2023. Accelerating science with human-aware artificial intelligence. Nat. Hum. Behav. 7 (10), 1682–1696.

Sun, C., Shrivastava, A., Singh, S., Gupta, A., 2017. Revisiting unreasonable effectiveness of data in deep learning era. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 843–852.

Sutton, R., 2019. The bitter lesson. URL http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

Teplitskiy, M., Ranu, H., Gray, G., Menietti, M., Guinan, E., Lakhani, K.R., 2019. Do experts listen to other experts?: Field experimental evidence from scientific peer review.

Thompson, N., Ge, S., Manso, G.F., 2022. The importance of (exponentially more) computing power. arXiv preprint arXiv:2206.14007.

Thompson, N., Greenewald, K., Lee, K., Manso, G.F., 2020. The computational limits of deep learning. arXiv preprint arXiv:2007.05558.

Trajtenberg, M., 2018. AI as the next GPT: a political-economy perspective. (Working Paper No. 24245) Series: Working Paper Series, National Bureau of Economic Research URL https://www.nber.org/papers/w24245.

Trammell, P., Korinek, A., 2020. Economic growth under transformative AI: A guide to the vast range of possibilities for output growth, wages, and the labor share. URL https://globalprioritiesinstitute.org/wp-content/uploads/Philip-Trammell-and-Anton-Korinek_Economic-Growth-under-Transformative-AI.pdf.

Vojnović, M., 2015. Contest theory: Incentive mechanisms and ranking methods. Cambridge University Press.

Weitzman, M.L., 1998. Recombinant Growth. Q. J. Econ. 113 (2), 331–360. http://dx.doi.org/10.1162/003355398555595.

Zucker, L.G., Darby, M.R., Torero, M., 2002. Labor mobility from academe to commerce. J. Labor Econ. 20 (3), 629–660.