The Economic Advantage of Computer Vision Over Human Labor, and Its Market Implications

by

Maja S. Svanberg

Bachelor of Arts in Computer Science,
Wellesley College (2018)

Submitted to the Institute for Data, Systems, and Society
in partial fulfillment of the requirements for the degree of

Master of Science in Technology and Policy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Authored by: Maja S. Svanberg
Institute for Data, Systems, and Society
May 12, 2023

Certified by: Neil Thompson
Research Scientist and Director, MIT FutureTech
Thesis Supervisor

Accepted by: Noelle Selin
Professor, Institute for Data, Systems, and Society and
Department of Earth, Atmospheric and Planetary Sciences
Director, Technology and Policy Program

The Economic Advantage of Computer Vision Over Human Labor, and Its Market Implications

by

Maja S. Svanberg

Submitted to the Institute for Data, Systems, and Society
on May 12, 2023, in partial fulfillment of the
requirements for the degree of
Master of Science in Technology and Policy

Abstract

With the emergence of Artificial Intelligence (A.I.), our lives and economy are undergoing a profound transformation. While there are huge benefits to be realized by the technology, we must also prepare for shifting circumstances, including changes in market dynamics and the labor market. Thus, to inform policy, we need to understand and forecast the implementation of A.I.

Previous forecasts of A.I. proliferation have focused on the technical feasibility of replacing human labor in existing tasks. However, since the decision to deploy a technology is ultimately an economic one, I develop a framework that compares the cost of A.I. to the cost of worker compensation. As such, this approach considers not only technical feasibility, but also the economic advantage of A.I. over human labor.

Using the framework, I examine the case of Computer Vision in the U.S. non-farm economy, drawing on previous work on the cost of Computer Vision, as well as government data on wages, tasks, and the size of firms. The results suggest that while Computer Vision can replace human labor across sectors and industries, it will only have an economic advantage over human labor in the very largest enterprises. In smaller companies, the sum of task-specific employee compensation does not exceed system development costs. Data is identified as the main driver of total Computer Vision development costs, placing incumbent firms at an advantage in the race to realize the economies of scale that Computer Vision, and A.I. in general, enable.

Based on my findings and related work on labor markets, I argue that automation is not the only way in which the introduction of A.I. could harm workers. Increased market concentration, stemming from access to data being restricted to firms with existing operations as well as enhanced production efficiency, might cause a systemic power shift from workers to firms. I point to the facilitation of industry data-sharing as a tool for policy-makers to mitigate these effects by lowering the barriers to entry into A.I.-centric markets.

Thesis Supervisor: Neil Thompson
Title: Research Scientist and Director, MIT FutureTech

Acknowledgments

I would like to express my gratitude to the many people who have supported me throughout my time at MIT.

First and foremost, I am extremely grateful for my advisor, Dr. Neil Thompson, whose consistent guidance and support have been essential to my success. Without his dedication and commitment, I could not have undertaken this journey. I am also grateful for the generous funding of MIT-IBM Watson AI Lab, which made this research possible.

I would not be where I am without Dr. Dean Eckles' mentorship, thank you. The same goes for Dr. Henric Johnson, thank you for inspiring me and for giving me the candid advice that I needed to stay on track.

Thank you Ben Jun-Hong Tang, Leon Yao, and Dr. Wensu Li for their collaboration and co-creation of ideas, which were essential to my research.

I am also grateful for the many professors who taught classes that changed my view of the world, including Dr. Nicholas Ashford, Dr. Anna Stansbury, Dr. Richard Zeckhauser, and Danny Weitzner, as well as to past mentors at Wellesley, Spotify, and Palantir. A special shout-out goes to Oscar Söderlund for hosting me at Einride last summer. Also, a big thank you to Minna Sandberg opening doors.

I would be remiss not to mention my peers in the Technology and Policy Program, who have been a tremendous source of joy over these two years. They have taught me how to find a balance between imagining how the world could be, and seeing it for what it is. Additional thanks to Kali, Joy, and Allie, for helping me through surgery, and for their friendship.

I would like to thank Barb DeLaBarre, Dr. Frank Field, and Elena Byrne, for being the backbone of the arguably most important graduate program in America and for their excellent work. Thank you, also, to Dr. Noelle Selin for leading TPP.

Finally, to my family and friends in Sweden. I can't wait to get back to doing life with you all, and I will see you soon!

Contributions

Although first-person singular pronouns are used throughout the document, this thesis is based on joint work with Dr. Neil Thompson and Dr. Wensu Li. Neil provided the idea and overarching direction for the project. In addition to providing insights and feedback, Wensu contributed to the thinking on the trade-offs of the costs and benefits of marginal performance improvements.

I wrote the thesis with Neil and Wensu reviewing. Neil, Wensu, and I plan to submit a co-authored version of this paper for publication.

I am grateful to Brian Goehring, Dr. Martin Fleming, Dr. Anna Pastwa, Kate Soule, Dr. Christophe Combemale, Ben Jun-Hong Tang, and Dr. Subhro Das for helpful comments and suggestions.

Funding

This research was supported in part by MIT-IBM Watson AI Lab.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Machines will steal our jobs" is a sentiment frequently echoed during times of rapid technological change. It even dates back to Keynes [39], who noted that technological change can outpace job creation. Modern economists, including Autor [8], Bishop [16], and Bessen et al. [14] [13], mostly agree that the aggregate amount of labor demanded relative to the population will not change over time, i.e., we are unlikely to have a jobless future. However, new developments can still be disruptive to individuals and worsen economic inequality. The Luddite movement of the early 18th and 19th centuries in the UK organized violent riots in response to the effect of automation on the quality of life of workers, as recounted by Mueller [46]. Michaels [44] found that digital technology aggravated inequality by increasing demand for high-income labor at the expense of middle-income labor. Autor et al. [9] and Krueger [41] report that computerization, in particular, increases the returns to education, favoring those who already have the resources to invest in their careers over those who do not. Therefore, to inform social and industrial policy to combat these inequality-increasing effects, we need to forecast the scale and nature of future technological advances.

The technology at the root of our current collective labor market-anxiety is Artificial Intelligence (A.I.). Fleming [32] sees the technology as an enabler of a fourth industrial revolution, and Brynjolfsson and McAfee [21] believe digitalization will underpin most, if not all, of the economy. Fleming et al. [33] and Acemoglu et al. [1] present evidence that A.I. is already changing the job market. Recent advances, including tools such as ChatGPT [59], are driven by increased computing power, which allows us to build large neural networks trained on vast amounts of data, so-called *Foundation Models*, a term coined by Bommasani et al. [18]. These generalist models can be fine-tuned for specialized downstream use cases using *Transfer Learning*, a technique that leverages relatively small amounts of additional data and computing power. In the case of ChatGPT, it was created by fine-tuning the foundation model GPT-3 [59]. The launch of ChatGPT reignited worries about technology changing the way we live and work, bringing the issue of studying the impact of A.I. on the economy to the forefront of public discourse.

The timeline and scope of A.I. proliferation are incredibly hard to predict, as discussed by Armstrong and Sotala [5], yet many researchers have created methods to predict the suitability of A.I. to replace or augment work currently done by humans. Frey et al. [34] were the first to make headlines when they stated that 47% of occupations are at high risk of automation. Subsequent studies have contested their findings, including work by Brynjolfsson et al. [22], Felten et al. [30], and Webb [72]. However, a limitation these studies have in common is that they focus solely on technical feasibility as a proxy for whether human labor is exposed to A.I. While technical feasibility is a necessary prerequisite for deployment, we also need to consider whether A.I. is actually more cost effective than human labor. With the computing- and data-intensive nature of neural networks, it is not obvious that it always is.

The economic advantage of A.I. over human labor matters. Agrawal et al. [4] and Acemoglu and Restrepo [2] emphasize financial viability as a determinant of technological improvements. Similarly, Habakukk [36] discusses the relative cost of machines and people when deploying labor-saving technologies. When automating existing human tasks, it is therefore clear that not all tasks in the economy are done at a scale that is sufficient to offset the development costs of an A.I. system. Borge [19] refers to this as the *minimum viable scale*. A task that could easily be done by a sufficiently sophisticated A.I. system might only be performed by a handful of people in a company. Hence, that company is unlikely to invest in its development and deployment. Therefore, in Chapter 2, my first contribution is to develop a framework for comparing the relative costs of A.I. and human labor at different scales of deployment. I use cost data for fine-tuning Computer Vision systems, developed by Thompson et al. [69, 68], as well as government data on wages and tasks, to implement this framework. I aggregate these costs across firms, industry groups, subsectors, and sectors, to determine where we find this minimum viable scale and, hence, the economic advantage of machines over human labor.

In Chapter 3, I present my findings. While Computer Vision has an economic advantage over human labor across the economy, that advantage is rare on the firm level because of the large development costs. When the advantage exists, it is only found in the largest enterprises. The largest cost item is data, suggesting that reducing the cost of data could help Computer Vision proliferate across more firms and thereby accelerate the replacement of human labor. Alternatively, data-sharing within industries can enable platform businesses to form, selling inferences across industry groups, subsectors, and sectors. I close the chapter by discussing to what extent we can generalize my results to other A.I. domains, including language models. Chapter 4 summarizes my method, results, and claims, and concludes the thesis.

Chapter 2

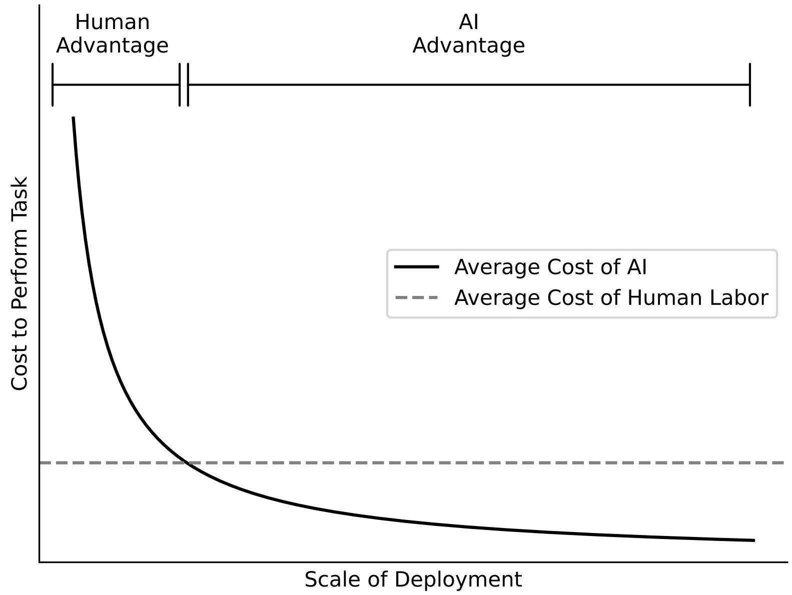
Comparing Humans to Machines

This chapter describes my framework for determining when A.I. has an economic advantage relative to human labor. I take a task-based approach to my comparison, drawing on a predictive model for the cost of developing a Computer Vision system, as well as official data on wages, industries, firms, and employment in the U.S. non-farm economy.

In the spirit of Autor et al. [10], Autor [7], and Acemoglu and Restrepo [3], my model assumes that the output of an economic process can be broken down into discrete tasks. Some of these tasks have the potential to be performed by a sufficiently sophisticated Computer Vision system, i.e., they are *technically feasible*, as explored by Brynjolfsson et al. [22], Felten et al. [30], and Webb [72] among others. However, these approaches do not take the economically motivated choice of deploying a technology into account.

Technology does not only progress from impossible to possible but also from expensive and rare to cheap and abundant (Agrawal et al. [4]). To understand to what extent human labor is exposed to A.I. replacement, we must therefore consider the current state of the markets and business decisions that control the deployment of the technology. If a technology offsets fewer costs than it takes to acquire it, no rational actor would choose the technology. There are, of course, many potential dimensions for comparative advantages between humans and machines (according to a McKinsey Report [24], the advantages of machines include speed and consistency), but consider-

Figure 2-1: Relative Returns to Scale of Artificial Intelligence



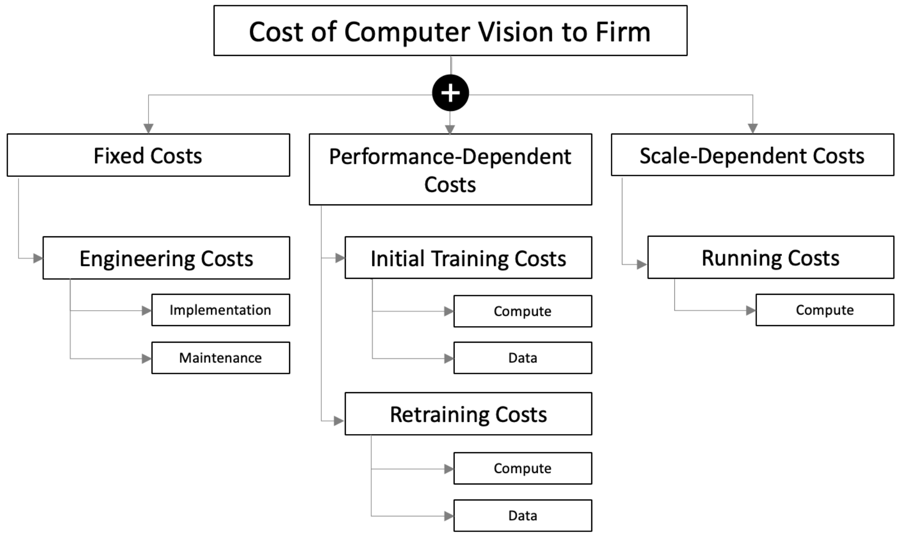
ing the economic advantage will advance our understanding one step beyond technical feasibility.

Where I make a contribution is, therefore, that I consider for which tasks Computer Vision systems have an economic advantage over human labor, i.e., when the cost to develop and run those systems is smaller than the cost of employing humans to perform the same tasks. Due to the economies of scale of A.I., there exists a *minimum viable scale*, a concept identified by Borge [19], for A.I. to be profitable, a trend shown in Figure 2-1. I explore where we can find this minimum viable scale for Computer Vision within the economy of existing tasks, i.e., do the economics of replacing humans for a specific task make sense on the firm, industry group, subsector, or sector levels?

2.1 Defining Economic Advantage

I say that Computer Vision has an economic advantage over human labor for a given task j and section s of the economy (i.e., a firm, industry group, subsector, or sector) when it not only has technical feasibility T_j, but also when the cost of Computer

Figure 2-2: Breakdown of Cost of Computer Vision



Vision , C_j^M, where M is for *Machine*, is less than the cost of labor, $C_{j,s}^H$, where H is for *Human*, as expressed by:

$$T_j \wedge (C_j^M < C_{j,s}^H)$$

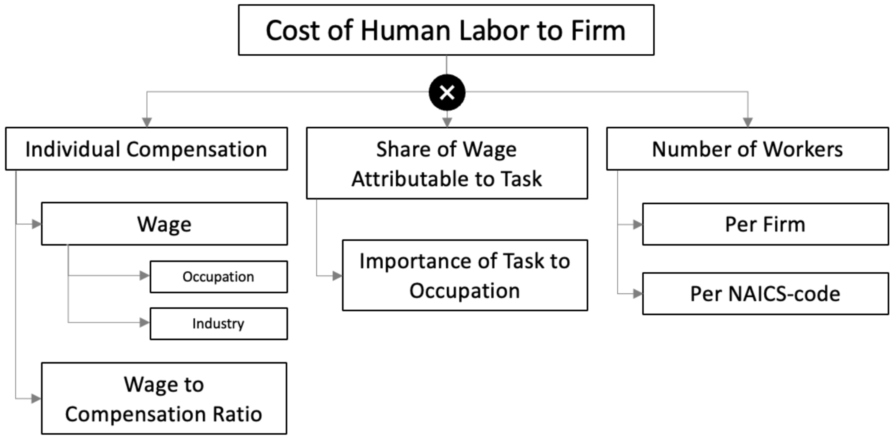
Figures 2-2 and 2-3 break down C_j^M and $C_{j,s}^H$ and into their main components. The rest of this chapter will go into detail about each of the components of T_j, C_j^M, and $C_{j,s}^H$, as well as explain my approach to data gathering.

2.2 Technical Feasibility

Many tasks currently performed by humans could be carried out by a sufficiently sophisticated Computer Vision system. For example, the system could check products for quality at the end of a factory assembly line or scan medical imagery for anomalies. However, other tasks have little use for vision technologies, e.g., negotiating the salary of subordinates. To find the technical suitability, T_j, we need to identify which tasks in the economy are suitable vision tasks and which are not.

While prior work on technical feasibility inspired my approach to this paper, they do not provide the data needed for my use case. Felten et al. [30] base their

Figure 2-3: Breakdown of Cost of Human Labor



method on linking A.I. progress to abilities, which does not translate to my task-based model of replacement. Webb [72] uses a task-based model but does not allow us to easily distinguish between Computer Vision and other A.I. domains. Eloundou et al. [29] only cover language tasks. Brynjolfsson et al. [22] did produce task-level data with an indicator for vision tasks, but when filtering their results for highly scoring image-based tasks, the output contained many tasks for which I, upon manual inspection, could not see an obvious Computer Vision use case for, e.g., "Analyze market conditions or trends" or "Dispose of biomedical waste in accordance with standards." Therefore, I create my own data on technical feasibility.

I take a manual approach to identifying T_j. Like Webb [72], Eloundou et al. [29], and Brynjolfsson et al. [22] I choose to rely on the O*NET Database 27.1 [50]. O*NET contains standardized characteristics of work and workers in the United States. By doing so, I assume that those tasks are an appropriate unit of replacement and that technological change does not otherwise change the overall processes and task breakdowns. The U.S. Department of Labor collects the data through survey instruments. It contains descriptions of the nature of 1016 occupations with 19,265 unique associated tasks, which are in turn mapped to 2087 different Direct Work Activities (DWAs) through a many-to-many relationship. Although the word "task" is a category in the O*NET schema, I find that the descriptions of the O*NET-

Table 2.1: Examples of *Tasks* and Technical Feasibility

j	O*NET-task	DWA	Suitable Vision Task?
1	Operate diagnostic equipment, such as radiographic or ultrasound equipment, and interpret the resulting images.	Analyze test data or images to inform diagnosis or treatment.	Yes
2	Operate diagnostic equipment, such as radiographic or ultrasound equipment, and interpret the resulting images.	Operate diagnostic imaging equipment.	No
3	Examine trays to ensure that they contain required items.	–	Yes

Each row is a distinct *task*. Note that rows 1 and 2 are associated with the same O*NET-Task. However, as the DWAs refer to different aspects of the O*NET-Task, only one of the rows is deemed suitable for Computer Vision.

Tasks are too broad and include too many different capabilities. Therefore, I define a *task* for the purposes of a unit of A.I. system capabilities as the combination of an O*NET-Task and DWA, as shown in Table 2.1. When O*NET-Tasks do not have any associated DWAs, I treat them as one DWA.

There are slight differences in occupational taxonomies that need to be rectified. Because O*NET includes more granularity than the Standard Occupational Classification (SOC) [51], which is used for data on wages, I truncate the O*NET-SOC Codes by removing the decimal points to match the SOC (see Appendix A).

The large number of O*NET-Tasks makes manual identification of vision tasks challenging, but because of the lack of prior art, it was still my preferred approach. To classify the combinations of almost 20,000 tasks and 2,000 DWAs, I first identify 190 DWAs that suggest the possibility of being replaceable by Computer Vision in any way. These include DWAs like ”Assess skin or hair conditions”, ”Examine patients to assess general physical condition”, ”Inspect items for damage or defects”, and ”Monitor facilities or operational systems”. Filtering on these 190 DWAs yields 1922 possible O*NET-Task-DWA combinations.

Naturally, there are instances where tasks are ambiguous, or where I lack context or knowledge to determine their suitability. To aid my decision-making process, I employ the following heuristics:

- When an O*NET-Task has multiple DWA components but only one of the DWAs refers to something visual, I only label one of the DWAs as a vision task. For example, for the O*NET-Task "Diagnose fractures using X-rays" consisting of the DWAs "Diagnose conditions" and "Analyze medical data", I only consider the latter DWA suitable. I assume the healthcare professional will still make the official diagnosis based on the output of the Computer Vision system.
- Even if I do not take into account the frequency at which the task is performed, I require repeatability. If a task needs new criteria each time it is performed, e.g., if a costume attendant "check[ing] the appearance of costumes on stage or under lights to determine whether desired effects are being achieved" only does so once per production, I do not consider it a suitable vision task.
- If something can be done with Computer Vision but there is simpler way to do it, it is not suitable. For example, I do not consider a Computer Vision system suitable to read gauges if they instead could be directly digitally encoded.
- If a task refers to a different technology, such as GIS, I do not consider it suitable even if it could be done by Computer Vision.
- If the vision part of the task comes for free when a human carries out all other components, it is not suitable. As an example, the DWA "Locate suspicious objects or vehicles" in the context of "Search prisoners and vehicles and conduct shakedowns of cells for valuables and contraband, such as weapons or drugs." is not suitable, although in other contexts, it is, e.g., "Locate suspicious bags pictured in printouts sent from remote monitoring areas, and set these bags aside for inspection."
- If a task requires prohibitively complex supplementary systems, e.g., "Piloting aircraft" or "Driving ground vehicles", I do not consider it suitable even if it can be done with Computer Vision.

- When evaluating a task's suitability, I do not consider the ethics of replacement, nor the ethics of the camera surveillance that is implicit in many applications. I ask whether something *could* be replaced, not whether it *should* be.

By applying these heuristics and my subjective judgement, I identified a total of 420 suitable vision tasks.

2.3 Cost of Computer Vision

To estimate the cost of a Computer Vision system, I rely on the prior work of Thompson et al. [68, 69], which breaks down and calculates individual cost components. In general, the cost of fine-tuning and deploying a Computer Vision system to perform a task j can be divided into three categories:

- Fixed Costs, or Engineering Costs[1], which include implementation costs, $C^{eng/imp}$, and maintenance cost, $C^{eng/main}$.
- Performance Dependent Costs, which include the cost of data, C_j^{Δ}, and the compute cost per training round, C_j^{τ}.
- Scale Dependent Costs, i.e., Running Costs, C_j^R.

To estimate the total cost of replacing human labor for a given task, I calculate the Net Present Value (NPV) of the cost of a system of a given lifespan. In addition to the initial round of fine-tuning, changes in the real world can lead to a decline in accuracy due to data drift, as explained by Moreno-Torres et al. [45]. To address this accuracy drop, the network must be retrained at regular intervals of K times per year. Denoting the capital discount rate d, the yearly rate of decrease in computing costs as m, and the system lifespan as L, the total cost of building a Computer Vision system to perform task j, C_j^M, is:

[1] Although Thompson et al. [68] also include *infrastructure cost* as a fixed cost, I ignore this since I assume the use of cloud computing.

$$C_j^M = C^{eng/imp} + C_j^{\Delta} + C_j^{\tau} + \sum_{i=0}^{L-1}(\frac{C^{eng/main} + C_j^{\Delta} \times K}{(1+d)^i} + \frac{C_j^R + C_j^{\tau} \times K}{((1+d) \times (1+m))^i})$$

Throughout the experiments, I make the following assumptions:

- A flat discount rate of 8% is applied across the economy, i.e., $d = 0.08$, corresponding to a conservative expected stock market return based on historical values [65].
- The yearly rate of decrease in computing costs is 22%, i.e., $m = 0.22$ based on findings by Hobbhahn and Besiroglu [38].
- The system lifespan is set to $L = 5$ years based on the software depreciation rate published by the Bureau of Economic Analysis [48].

2.3.1 Fixed Costs

The engineering project for a Computer Vision system involves two phases: implementation and maintenance. These phases correspond to the variables $C^{eng/imp}$ and $C^{eng/main}$ in the equation discussed in Section 2.3. For this analysis, I assume that the implementation and maintenance costs are the same for all tasks, reflecting the complexity and repeatability of the engineering process rather than the complexity of individual tasks. To estimate these costs, I refer to the IBM case study presented by Thompson et al. [67], which describes a Deep Learning time series prediction project. The study reports an upfront implementation cost of $C^{eng/imp} = \$1,765,000$ for a 6-month project and a yearly maintenance cost of $C^{eng/main} = \$242,840$. Refer to Tables 2.2 and 2.3 for a detailed breakdown of these costs.

2.3.2 Performance Dependent Costs

Training a neural network requires significant computing power, making it a resource-intensive process. For example, Brown et al. [20] report that training the foundation

Table 2.2: Implementation Team Costs ($C^{eng/imp}$) [67]

Worker Type	#	Utilization	Monthly Cost Per Worker	Monthly Cost
IBM Engineers	6	100%	$40,000.00	$240,000.00
Client Engineers	4	80%	$16,666.00	$53,333.33
Subject Matter Experts	1	5%	$16,666.00	$833.33
			Monthly Cost Sum:	$294,166.66
			Total 6 Months Cost:	$1,765,000.00

Table 2.3: Maintenance Team Costs ($C^{eng/main}$) [67]

Worker Type	#	Utilization	Monthly Cost Per Worker	Yearly Cost
IBM Engineers	0	0%	$40,000	$0
Client Engineers	4	30%	$16,666.66	$240,000.00
Subject Matter Experts	1	1.42%	$16,666.66	$2,840.00
			Total Yearly Cost:	$242,840.00

model GPT-3 took "several thousand petaflop/s-days of compute", which would require approximately $10,000,000[2]. My costs will be smaller, but likely not negligible, since I am assuming that it is possible to leverage existing foundation models and fine-tuning[3].

The cost of the Computer Vision system will depend on the required quality. As established in Section 2.2, a *sufficiently sophisticated* Computer Vision system can perform a suitable vision task. Therefore, the technical feasibility depends on the system meeting minimum standards. While fairness and other intangible measures can impact system quality, I focus on two tangible measures, namely *accuracy* and *entropy*. Accuracy is the percentage of correct predictions made by a system, and entropy is a measure of the complexity of information in the possible outcomes of the system, as defined by Shannon [61], e.g., how many different categories does the model need to be able to recognize? Hence, I assume that, for each task j, there exists

[2]Assuming a rate of $0.34 per hour, 4k petaflop/s-days corresponds to 96k petaflop/s-hours or 24M hours, which yields a compute cost of approximately $8,160,000.

[3]Fine-tuning the process of adapting a pre-trained deep learning model to a specific use case to enhance its performance, also known as transfer learning.

minimum levels of accuracy a_j and entropy e_j that a human worker must achieve to be deemed fit for the job. I further assume that a Computer Vision system that meets those requirements would be technically suitable to replace the human at task j. In that sense, a_j and e_j are implicit in my definition of technical feasibility T_j.

The cost of training is going to depend heavily on the required accuracy and entropy. The higher the number of categories a system needs to distinguish between, the more data it needs to be trained on, in order not to confuse it. Similarly, the less error-tolerant the system can be, the more examples it needs to have seen. Therefore, the cost of a Computer Vision system is partially driven by its required performance and quality.

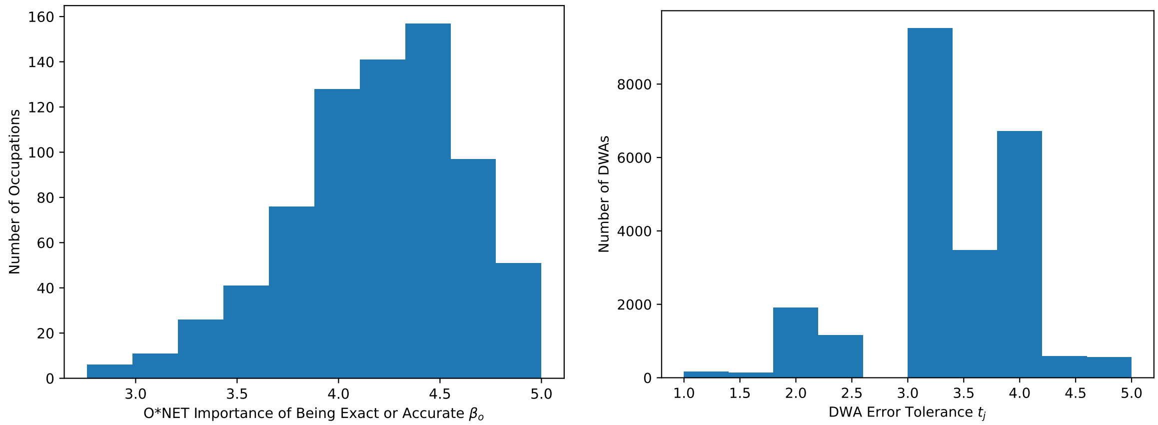
To determine the minimum required accuracy a_j for a given task j, I use the "Importance of Being Exact or Accurate" variable from the O*NET database, β_o shown in Figure 2-4a. It rates an occupation's accuracy on a scale from 1 to 5, where 5 represents the highest level of accuracy. Additionally, I incorporate the "Error Tolerance" score, t_j, shown in Figure 2-4b, from Brynjolfsson et al.'s [22] study. It rates the level of error tolerance required for a task on a scale from 1 to 5, where 5 represents the least tolerance for error. Using these scores, I calculate a_j as follows:

$$a_j = min(\frac{max(6 - t_j, \beta_o)}{5}, 0.995)$$

The value 0.995 accounts for a level of human error. Figure 2-5 shows a histogram of the required accuracies for any given task.

To find the entropy level, e_j, I assume that the foundation model was trained on ImageNet [27], and I set a target entropy level that represents the mid-range of entropy for the Computer Vision deployments included in the tools developed by Thompson et al. [69], a value roughly corresponding to 20 equally sized classes.

To determine the number of required datapoints and its costs, C_j^{Δ}, and thereby also the computational requirements and costs, C_j^{τ}, I refer to the work by Thompson et al. [69], which assumes an exponential relationship between the requirements for the fine-tuned model and the number of required datapoints. Using the number of



(a) O*NET Importance of Being Exact or Accuracy per Occupation (β_o)

(b) Error Tolerance per DWA (t_j), according to Brynjolfsson et al. [22]

Figure 2-4: Inputs for Minimum Required Accuracy per Task

datapoints and the size of the foundation model, we can calculate the number of GPU hours required. I model the total cost of data using $f(a_j, e_j)$, the number of datapoints required for a given accuracy a_j and entropy e_j, as follows:

$$C_j^{\Delta} = f(a_j, e_j) \times p^{\Delta}$$

Here, p^{Δ} is the cost per datapoint. We assume that $p^{\Delta} = \$0.05$, five times higher than the smallest possible assignment award on the crowd-sourcing platform Amazon mTurk[4].

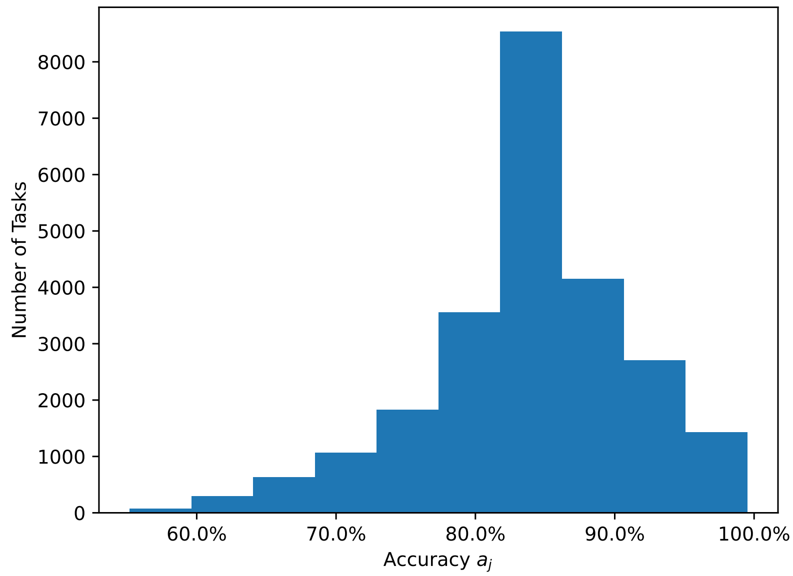
To calculate the cost of compute, C_j^{τ}, we use the same number of datapoints, $f(a_j, e_j)$, and the following equation:

$$C_j^{\tau} = \frac{f(a_j, e_j) \times 2 \times \# \text{ Model Connections} \times 3 \times \# \text{ Epochs}}{\text{GPU FLOPs/h}} \times \frac{p^{GPUh}}{U}$$

Here, the numerator of the left factor is the number of floating point operations (FLOPs) required to train the model, based on research by Sevilla et al. [60]. The denominator is the number of FLOPs a given GPU can perform in one hour at peak

[4] $0.01 according to `https://www.mturk.com/pricing`, Accessed: 2023-04-09. We use a higher than minimum rate to account for the requirement for specialized skills for many tasks, as well as the cost to collect the images.

Figure 2-5: Histogram of Minimum Required Accuracy Per Task



utilization. The right factor is the price per GPU hour p^{GPUh} over the utilization U of that GPU.

I assume the use of a 4 FP-32 TFLOPS GPU, with an hourly rate of $p^{GPUh} = \$0.340$[5]. I also assume a GPU utilization of $U = 85\%$, which is consistent with the utilization when training large Computer Vision models like ResNet50, as reported by Yeung et al. [74]. Finally, I assume that 50 epochs are used, and that the foundation model used has the size of VGG-19 [62], i.e., 1.44×10^8 parameters, the largest configuration studied by Thompson et al. [69].

2.3.3 Scale Dependent Costs

In addition to the fixed and performance dependent costs, there is a marginal cost of running the model, i.e., making inferences. According to anecdotal data from IBM practitioners, the cost of making inferences using a Computer Vision system can be between three to eighteen times as expensive as the training cost. However, these numbers are based on the total scale of deployment and not the minimum viable scale. For instance, in a case where a model was trained for \$1,000 and then required

[5] `eia2.xlarge` pricing in U.S. East region on AWS `https://aws.amazon.com/machine-learning/elastic-inference/pricing/`, Accessed: 2023-04-09

\$18,000 in inference costs, there were 2,500 datapoints for training and 1,300,000 inferences were made. The minimum viable scale might only have required a fraction of that number of inferences to break even, so we cannot take these numbers at face value when calculating the running costs for finding the economic advantage.

Instead, to determine the economic advantage between human and machine labor, we will consider running costs that are proportional to the amount of human labor being displaced. According to a McKinsey report [24], machines have an advantage over human labor in terms of speed. It is very conservative to assume that a machine can perform a task in less time than a human. For instance, a 4 FP-32 TFLOPS GPU hour is enough to make approximately 50,000,000[6] inferences using VGG-19 [62], one of the larger publicly available Computer Vision models. No human could match the pace of almost 14,000 inferences per second achieved by that machine. Other A.I. hardware setups can be more powerful and therefore more expensive, like the NVIDIA DRIVE Orin system of 254 TOPS[7], which the company intends to be the brain of autonomous vehicles. However, even if high-end setups give a cost per inference that is 1000 times greater than our assumed model, it would still be well below our upper bound. Therefore, I believe that the upper bound holds for a wide range of possible setups.

To estimate the yearly running costs, C_j^R, for a Computer Vision system to perform task j instead of n employees, I assume that task j takes up w_j of the employees' duties. I model the yearly running costs as follows:

$$C_j^R = \frac{p^{GPUh}}{U} \times 40 \times 50 \times w_j \times n$$

40 is the number of hours worked per week, and 50 is the number of weeks worked per year. Like for training costs, I use a GPU hourly rate of $p^{GPUh} = \$0.34$ and assume a GPU utilization of $U = 85\%$. My method for finding w_j and n is outlined in Subsections 2.4.1 and 2.4.2, respectively.

[6] $(4 \times 10^{12} \times 3600)/(2 \times 1.44 \times 10^{8})$, where the nominator is FLOPs per hour and the denominator is FLOPs per inference.

[7] `https://www.nvidia.com/en-us/self-driving-cars/drive-platform/hardware/`, Accessed: 2023-05-08

Table 2.4: Bare-Bones Implementation Team Costs ($C^{BB/eng/imp}$)

Worker Type	#	Utilization	Monthly Cost Per Worker	Monthly Cost
IBM Engineers	0	0%	$0.00	$0.00
Client Engineers	2	80%	$16,666.00	$26,666.66
Subject Matter Experts	1	5%	$16,666.00	$833.33
			Monthly Cost Sum:	$27,500.00
			Total 6 Months Cost:	$165,000.00

Table 2.5: Bare-Bones Maintenance Team Costs ($C^{BB/eng/main}$)

Worker Type	#	Utilization	Monthly Cost Per Worker	Yearly Cost
IBM Engineers	0	0%	$40,000	$0
Client Engineers	2	30%	$16,666.66	$120,000.00
Subject Matter Experts	1	1.42%	$16,666.66	$2,840.00
			Total Yearly Cost:	$122,840.00

2.3.4 Alternative: Bare-Bones Setup

There are instances where costs can be reduced or eliminated completely. For example, a foundation model might already be fit for the task, or sufficiently close to it that fine-tuning can be done with available data and hardware. Therefore, in addition to the setup above, I explore the possibility that the only cost of Computer Vision is that of a small engineering team, with the breakdown of costs shown in Tables 2.4 and 2.5.

Using the same assumption of $d = 0.08$ and $L = 5$ as elsewhere in this thesis, the total cost C_j^M of implementing this bare-bones Computer Vision system for task j can, hence, be written as

$$C_j^M = C^{BB/eng/imp} + \sum_{i=0}^{L-1} \frac{C^{BB/eng/main}}{(1+d)^i}$$

2.4 Cost of Human Labor

Unlike the cost of Computer Vision, human labor does not exhibit obvious economies of scale. In fact, it is not unreasonable to assume that the only cost of human labor is the marginal cost of compensation per worker. Because of these differences, human labor has an advantage at smaller scales, as shown in Figure 2-1. It is therefore relevant to consider the different business models that could drive investments into A.I. deployment; either as a third-party platform (A.I.-as-a-Service) or in-house within one firm.

To estimate the cost of human labor for a given task, I consider the economy at different levels of detail, including sectors, subsectors, and industry groups, i.e., 2-, 3-, and 4-digit North American Industry Classification System codes (NAICS) [47], as well as individual firms of different sizes. I refer to these different units of the economy as s. For a given s and task j, where j is technically replaceable by a Computer Vision system with a lifespan of L, I define the NPV cost of labor as follows:

$$C^H_{j,s} = \sum_{i=0}^{L-1} \frac{w_{o,s} \times r \times w_j \times n_{o,s}}{(1+d)^i}$$

Here, $w_{o,s}$ is the mean wage for occupation o in s, r is the wage to total compensation ratio, w_j is the fraction of an occupation's duties that is task j, $n_{o,s}$ is the number of workers of a given occupation in s, and d is the capital discount rate. I assume $L = 5$ and $d = 0.08$.

I explain my approach for calculating $w_{o,s}$, r, and w_j in Subsection 2.4.1. Subsection 2.4.2 outlines my method for finding $n_{o,s}$ for different scales of companies and parts of the economy.

2.4.1 Task Costs for Individual Workers

To estimate wage costs $w_{o,s}$, I use the 2021 Occupational Employment and Wage Statistics (OEWS) data tables for the U.S. non-farm economy created by the U.S.

Department of Labor's Bureau of Labor Statistics. These tables provide the average wage and number of employees per NAICS code for each occupation o and section s of the economy.

There is missing data for smaller occupations and sections in more detailed NAICS codes (3- and 4-digit), likely for privacy reasons. To impute missing average wage and employment numbers, I use a bottom-up summation followed by a top-down distribution of employment numbers, as well as a top-down propagation of average wage.

To convert employee wages to employer costs, I use the Bureau of Labor Statistics September 2022 wage to compensation ratio of $r = 1.449$[8] [52].

To assign a fraction, w_j, of an employee's wage to a given task and calculate labor cost, I weight each task by its score on the O*NET-Task-Importance scale, following the example of Brynjolfsson et al. [22] and Webb [72]. For O*NET-Tasks that have multiple DWAs associated with them, I distribute the score of the O*NET-Task equally among the DWAs, giving me a weight for my definition of tasks described in Section 2.2. In the made-up example of a Liberal Arts Professor who only has the two O*NET-Tasks *teach* and *research*, with an Importance of 5 and 3 respectively, *teach* would account for 5/8 of compensation and *research* 3/8. If *research* had two DWAs, each of them would account for 3/16 each. In general, for an O*NET-Task Y with an O*NET-Task-Importance score Φ_Y and q associated DWAs, we find q different task-weight scores w_j using the following formula:

$$w_j = \frac{\Phi_Y}{\sum_{\text{T}\in\text{O*NET-Tasks in Occupation}} \Phi_{\text{T}} \times q}$$

2.4.2 Scale of Deployment

Whether a Computer Vision model is deployed to replace the labor of one employee or one million employees will affect its profitability. The cost of development and implementation, which may be prohibitively expensive for a single firm, can be offset by widespread adoption across an industry or even the entire economy. If a system

[8] 1/0.69

costs \$5,000,000 to develop and run, it will not make sense for it to replace the tasks of two nurses in an Urgent Care office, but it would be a great idea to replace that task for all nurses in America. As such, I analyze two deployment scales for computer vision: platform deployment and in-house deployment.

Platform Deployment

Platform deployment involves deploying the Computer Vision system as a service across a segment s of the economy. To assess the potential size of these markets, I first obtain the number of employees $n_{o,s}$ for a given occupation o within each sector, subsector, and industry group, using imputed OEWS data as described in Subsection 2.4.1. We want to know $n_{o,s}$ across different levels of granularity, since we do not know how similar a task is across segments. For instance, a startup selling access to a model that checks the quality of food may find that the model is valuable to replace human labor across all restaurants, but not in adjacent industries like catering. Performing the calculations for the different levels allows us to do a sensitivity check on that assumption.

Internal Deployment

To estimate the number of employees $n_{g,o,s}$ for an individual firm g in occupation o and NAICS code s, I combine the average occupation distribution of NAICS codes with publicly available data on firm sizes across those sections, using the following formula:

$$n_{g,o,s} = \frac{n_{o,s}}{\text{Total employment in } s} \times \text{Size of } g$$

For example, with 1,028,940 employed *35-2014 Cooks, Restaurant* in *7225 Restaurants and Other Eating Places* across the United States, an industry group with 9,431,910 employees total, a company with 50 employees would have 5 and a half cooks[9]. I obtain the employment in an occupation in a given NAICS code using the

[9] $\frac{1{,}028{,}940}{9{,}431{,}910} \times 50 \approx 5.45$. This number does not include other food preparation workers, supervisors, or head chefs.

OEWS data. To approximate the firm employment size distributions in each NAICS code, I use the 2019 SUSB Annual Data Tables by Establishment Industry [23], as detailed in Appendix B.

It is reasonable to assume that many firms will have a different distribution of occupations than their respective sectors, subsectors, or industry groups, potentially leading to a high enough concentration to make a Computer Vision deployment viable for them. However, unless there is a strong correlation between occupation concentration and firm size close to the minimum deployment size, this should have a minimal effect on the overall results.

2.5 Limitations

A criticism of my approach is that A.I. would not simply replace human labor, but that there are other comparative advantages to the technology that makes it both strategic and attractive to deploy, not to mention A.I. products that do things beyond human abilities. Bessen et al. [15] find that only 50% of A.I. startups help customers reduce labor costs, whereas 75% build products to improve product and process quality. Plotz and Fink [54] and Dranove and Garthwaite [28] discuss the potential value of A.I. as increased quality of output. Hence, we might see companies deploy A.I. for applications where it would not make economic sense when comparing only to the size of their existing payroll. Instead, the technology might increase the company's market share. This could, of course, reduce demand for labor in competing firms in the short-term.

While not strictly a limitation of this method, I caution readers not to interpret the results of applying my method as a prediction of a net-job loss in the aggregate. Technological change can be devastating for individuals who cannot adapt, as evidenced by the higher likelihood of early retirement in jobs exposed to automation shown by Yashiro et al. [73]. However, economists, including Autor [8], Bishop [16], and Bessen et al. [14, 13], largely agree that the aggregate amount of labor demanded is unlikely to change over time, since freed labor can be absorbed by the freed capital, creating new tasks and products.

Chapter 3

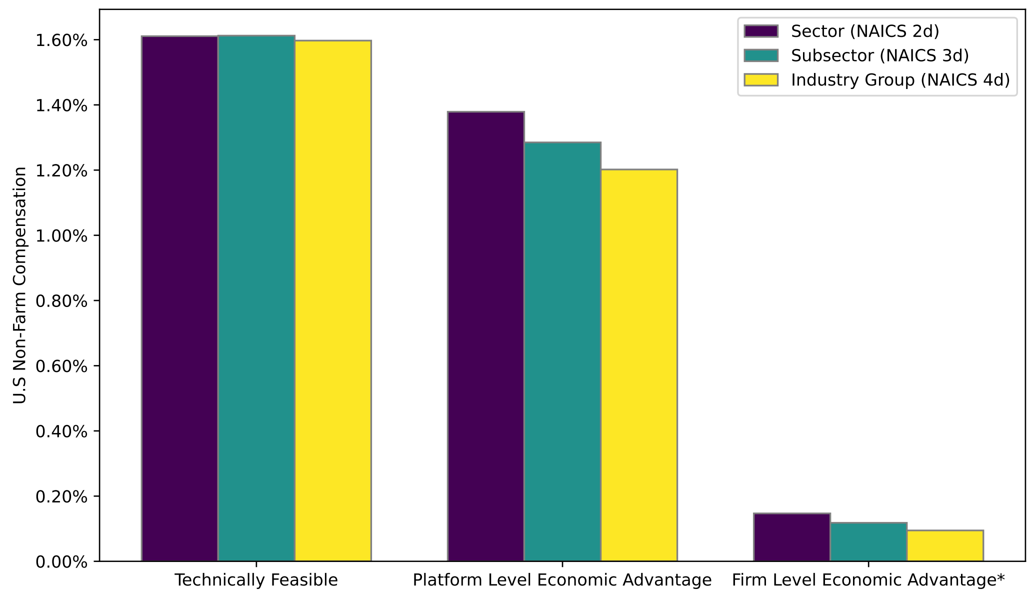
The Economics of Computer Vision

In this chapter, I present the results of applying the method from Chapter 2. The findings suggest that Computer Vision is generally more cost-effective than labor when tasks are aggregated across sectors, subsectors, and industry groups, but the minimum viable scales are too large for most firms. I identify data as the largest cost item in development, accounting for over 70-90% of total platform-system costs, and discuss its role in Computer Vision proliferation as well as market competition. Further, I find that Computer Vision tasks are concentrated in low- to mid-income occupations as well as occupations requiring a high school degree with some training, a trend that remains even after economic advantage has been considered. I close the chapter with a discussion about the extent to which these conclusions apply to other types of A.I., including language models.

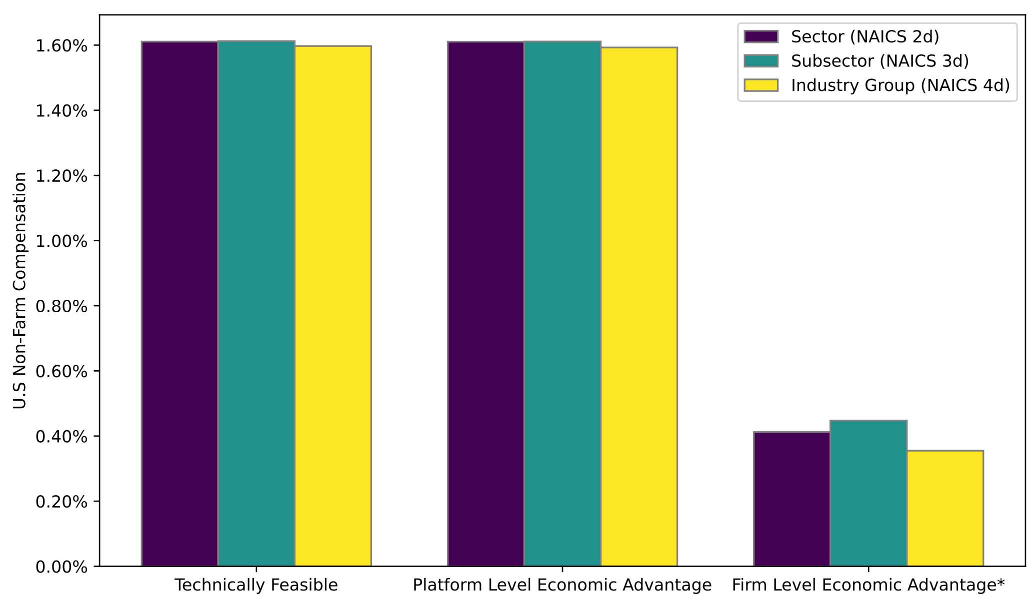
3.1 Market Structure Implications

Although Computer Vision can be more cost-effective than human labor when tasks are aggregated across sectors, subsectors, and industry groups, the minimum viable scale is too large for most firms. Figure 3-1a shows that vision tasks comprise 1.6% of U.S. non-farm compensation. Computer Vision would have an economic advantage over human labor for 80% of this work (1.3% of compensation) when aggregating the labor across sectors, subsectors, and industry groups. But only 6% (0.1% of compen-

Figure 3-1: Economic Advantage Across Sectors, Subsectors, Industry Groups, and Firms



(a) Feasibility across the economy using my main method.



(b) Feasibility across the economy using the minimal setup described in Subsection 2.3.4.

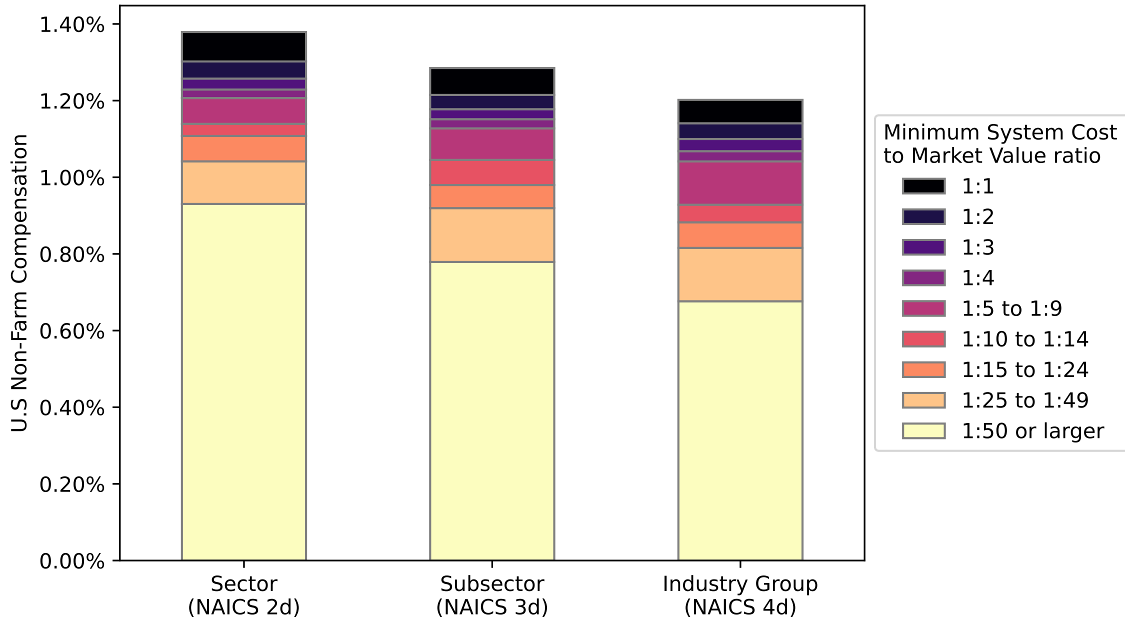
*because the SUSB Annual Data Tables lack firm size data for several NAICS codes, the denominator only includes the codes for which I do have firm size data.

sation) has an economic advantage when deployed at the company level and then, only in larger companies. When considering only the bare-bones costs of Computer Vision described in Section 2.3.4, with free data, free compute, and only minimal engineering effort, the amount of compensation that is exposed on the firm level increases to 25% (0.4% of compensation), as shown in Figure 3-1b, but the implications for small companies remain.

The impact of the findings is consequential. For widespread adoption of Computer Vision at its current cost, there needs to be restructuring of the economy. One possibility is that larger companies with A.I. capabilities will simply outcompete smaller ones, leading to a higher proportion of the economy being controlled by a smaller number of firms. Alternatively, third-party platforms could emerge to provide these services to companies of all sizes, or large companies could sell their systems to competitors. Most likely, we will observe a combination of these business models across various markets and use-cases. However, depending on the level of competition we desire in our markets, and what we want companies to be competing on, these observations have profound implications for policy. To the extent that access to A.I. is a barrier to entry into a market, we might want to enable and incentivize the creation of third-party platforms, and as such democratize access to A.I. to include smaller players. Enabling these platforms to gain access to data to train the models with will be crucial.

In order for a platform business model to actually be viable in a market based only on economic advantage, there needs to be a reasonable margin between the size of the addressable market, i.e., the total compensation for the task within the defined scope of the system, and the cost of development. In fact, the market needs to be many times bigger than the cost of development. Why? Because making enterprise sales is a complex task. Hannan and Freeman [37] describe how inertia, i.e., resistance to change, is a powerful force within companies, and Mueller [46] illustrates that this resistance is especially powerful when it comes to avoiding automation of existing tasks. Closing rates are as low as 20% in Enterprise IT sales, according to HubSpot Sales blogger Jay Fuchs [35]. Combining these insights, I say that it is likely that

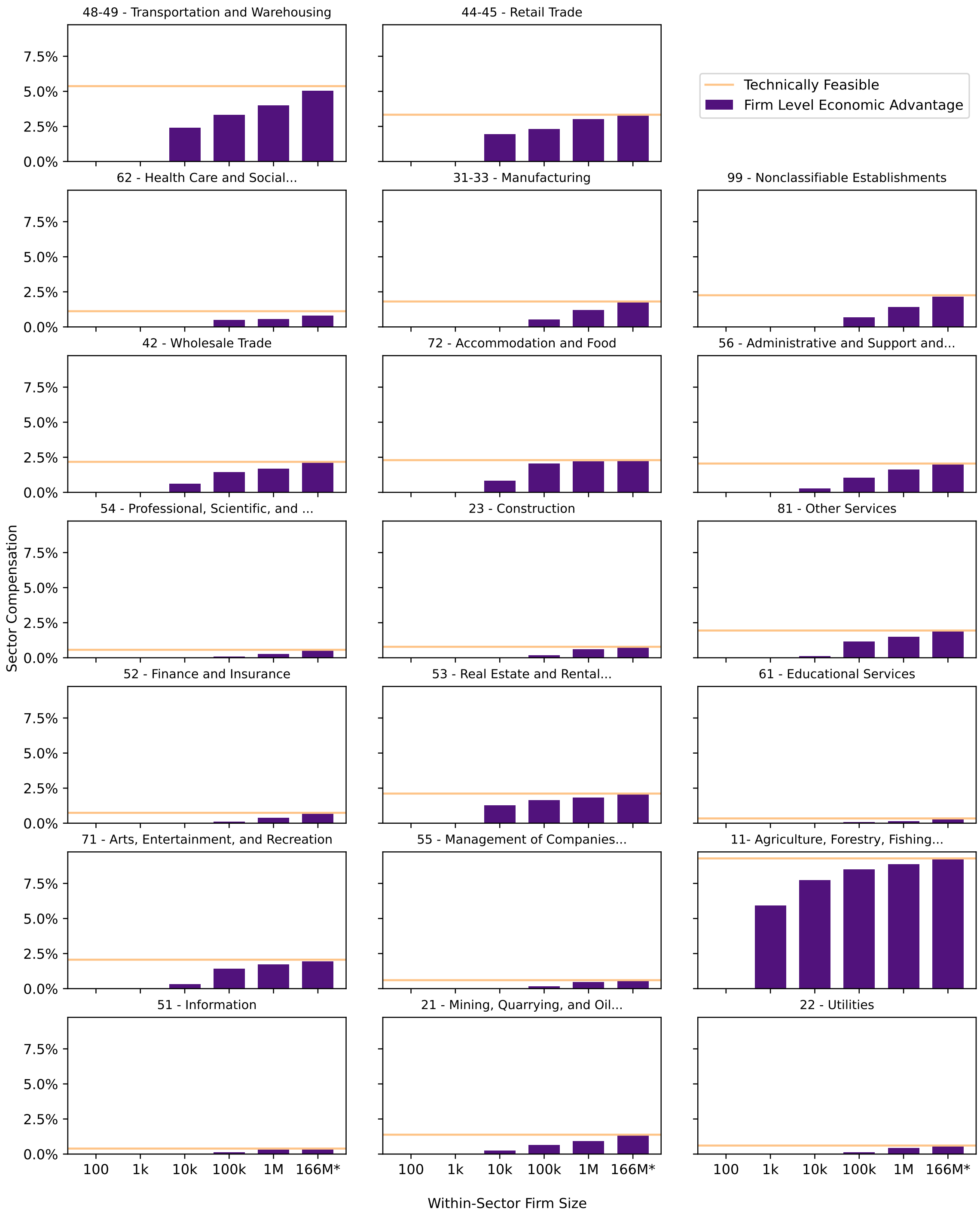
Figure 3-2: Ratio of System Cost to Addressable Market Value



any third-party vendor could only capture a fraction of the total market. In this context, Figure 3-2 breaks down the addressable market sizes for Computer Vision platform systems by their ratio to the cost of the minimum viable system for the given task. It is clear that the majority of the task market value is in markets where the market size is many times larger than the cost of developing a Computer Vision system to automate that task. This gives promise to platform business models even where Computer Vision has no advantage over human labor other than its price tag.

The replacement of tasks is mostly restricted to very large firms, as shown in Figure 3-3. This figure shows each sector and the percentage of compensation made up by tasks for which company of a given size (ranging from 166 million employees, the size of the U.S. civilian workforce, to 100 employees) would be able to profitably replace its human labor using Computer Vision. No tasks can be profitably replaced by a company with 100 employees, no matter the sector. In fact, the only sector where a company with 1000 employees should deploy Computer Vision internally is *Agriculture, Forestry, Fishing and Hunting*. This finding is consistent with empirical evidence on A.I. adoption by U.S. companies by Zolas et al. [76], which reveals that

Figure 3-3: Economic Advantage by Firm Sizes and Sector (NAICS 2d)



*166M is the size of the U.S. civilian workforce. Sectors are sorted by total vision task compensation (high to low).

only 1 in 16 firms uses any form of A.I.

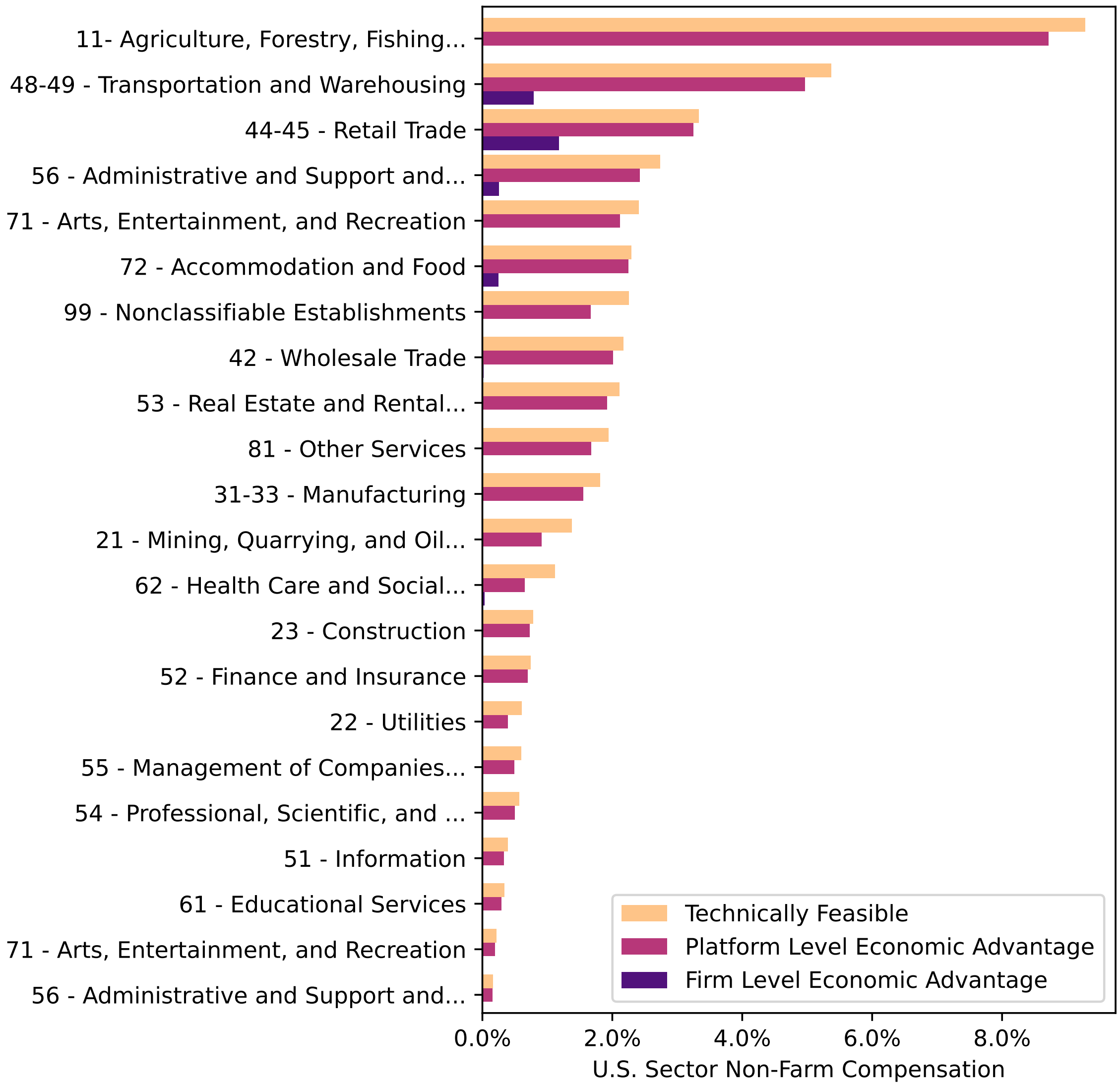
The business model required to achieve the scale that makes Computer Vision cheaper than human labor, whether in-house or platform, will have a huge impact on the scale and pace of Computer Vision proliferation. Although in-house development is a risky venture where the costs and benefits are hard to estimate upfront, according to Stamelos et al. [63], and organizations suffer from inertia, the lowered marginal cost of inference gives increased returns to scale, making it an attractive investment to individual decision-makers. Therefore, we can consider the firm-level economic advantage as a proxy for technological change that could happen on a shorter time-scale, and industry group, subsector, and sector level advantage as a proxy for what could happen on a longer time-scale.

The platformized Computer Vision business models might indeed take a long time to proliferate, but the trade-offs are arguably better because of the larger potential market. However, access to data is a big hurdle for these business models where the platforms are not already incumbents. There are examples where industry actors have come together to establish data-sharing agreements where a third party could not otherwise collect the required data, such as the NVIDIA Drive collaboration described by Thompson [66], or where regulatory bodies have introduced bills to facilitate these exchanges, such as the Data Act put forward by the EU in February 2022 [25, 26], but largely, this still remains an obstacle to the platformization of A.I. Effective government policy could enable more industries to overcome that obstacle.

3.1.1 Business Exposure

Opportunities for platformized Computer Vision replacement of vision tasks exist in all sectors, but firm-level opportunities are concentrated in only a handful. Figure 3-4 shows the sector-specific (NAICS 2d) aggregates of technically feasible tasks, as well as the economic advantage of platform and firm deployments. In fact, the fraction of technically feasible tasks that have an economic advantage as a platform is rather similar across all sectors, with one of the few noticeable outliers being the *Health Care and Social Assistance* sector. The sectors where we do see firm-level tasks exceeding

Figure 3-4: Economic Advantage by Sector (NAICS 2d)



the minimum viable scale include *Retail* and *Transportation and Warehousing.* An explanation for this could be that these sectors contain some of our largest enterprises, including Walmart and Amazon. In contrast, we can imagine that the *Agriculture, Forestry, Fishing and Hunting* sector, which has a high percentage of vision tasks yet no tasks with firm level economic advantage, does not benefit from the same economies of scale as *Retail* or *Transportation and Warehousing* do. Hence, companies in that sector are not large enough to deploy Computer Vision technologies, at least not at the current state of technology.

3.2 Labor Market Exposure

When considering the economic benefits of using Computer Vision technology over human labor, it is important to examine which workers will maintain their economic advantage and which will not. While predicting the labor market effects of a productivity-enhancing technology is challenging (e.g., introducing ATMs counter-intuitively did not change the number of bank tellers, as noted by Bessen [12]), it will inevitably have an effect on individual workers. Previous studies by Felten et al. [30] and Webb [72] suggest that A.I. primarily affects high-skilled tasks and high-income occupations. However, my findings indicate that Computer Vision does not fit this pattern.

Instead, Computer Vision tasks are more prevalent in occupations that have lower incomes and require less training. Figures 3-5 and 3-6 show the aggregate concentrations of Computer Vision tasks across income deciles and job preparation levels[1] respectively. While these trends are evident when examining the percentage of compensation, the R^2 values for individual occupations are very low: the most extreme R^2 is -0.04[2]. Therefore, while the overall trends are strong, a well-paying job is not significantly less likely to be exposed to Computer Vision simply because it is

[1]For Job Preparation, I use the O*NET Job Zones. Job Zones are groupings of occupations according to preparation, which roughly corresponds to a level of education.

[2]Comparing Job Zone to Percentage of Platform Economic Advantage per Occupation on the Sector Level.

Figure 3-5: Economic Advantage by Income Decile

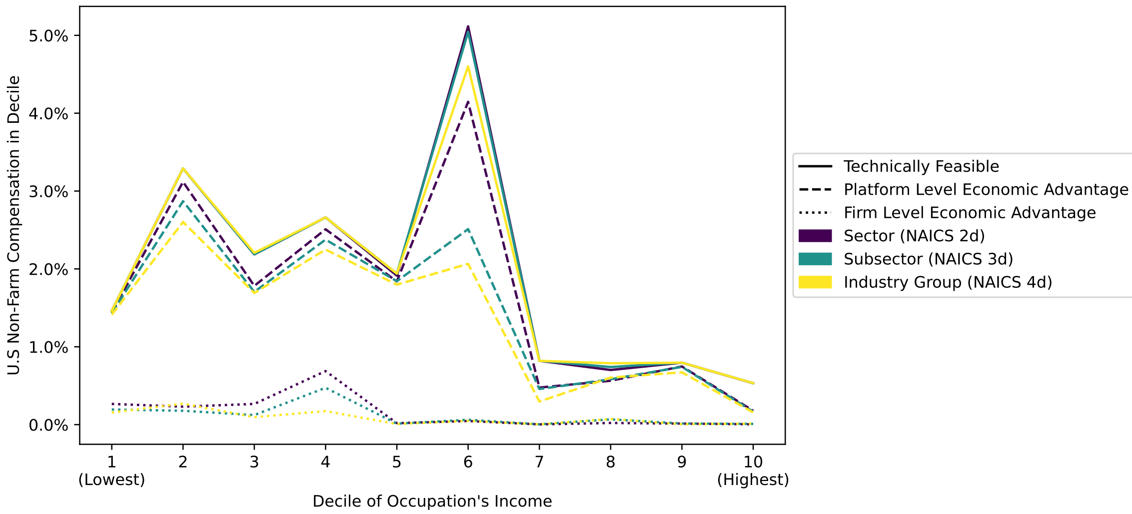
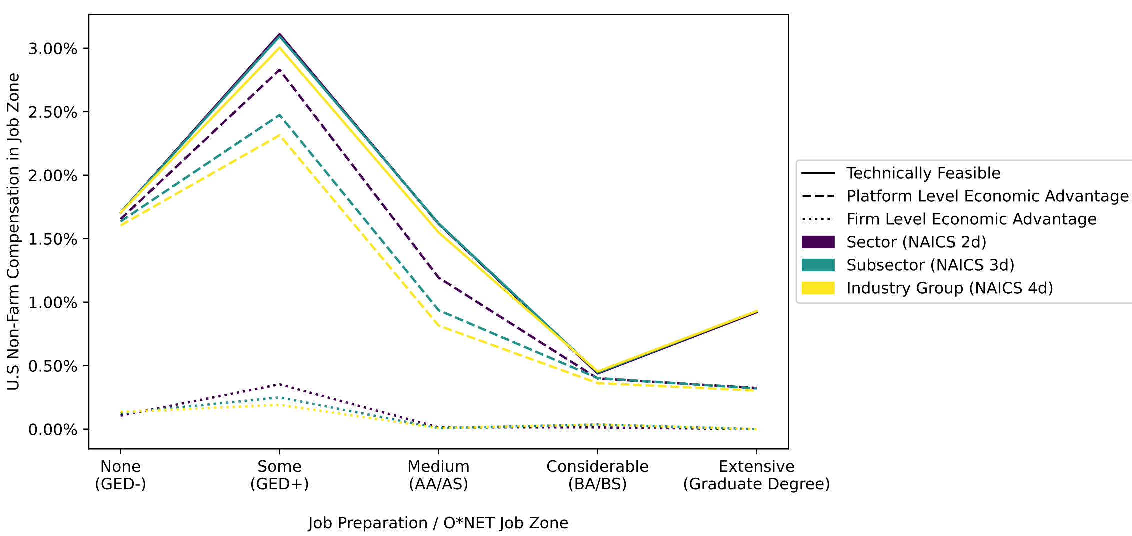


Figure 3-6: Economic Advantage by Job Preparation



well-paying.

The trends observed in the previous paragraph remain consistent when accounting for economic advantage, with the exception of occupations requiring extensive preparation, i.e., a graduate degree or similar. As shown in Figure 3-6, platform-level economic advantage is found for most of the vision task compensation, except among occupations that require a graduate degree, where less than half of the vision task compensation is economically feasible for replacement on the platform level. This economic disadvantage mostly consists of healthcare occupations such as Dentists, Physicians, and Nurses, who are likely to maintain their economic advantage because of the exponential cost of training Computer Vision to levels of accuracy approaching 100%. It is worth noting that occupations with "Some preparation" have a higher exposure to Computer Vision compared to both "No preparation" and higher-requirement occupations, suggesting that Computer Vision systems replace tasks that require some level of knowledge and training but not those that are increasingly complex.

The macroeconomic implications of technological change, and automation in particular, are difficult to predict. Even if machines replace human labor, there may be offsetting effects from increased demand or complementarity, according to Autor [8]. While some theories suggest that technological progress exacerbates inequality by favoring capital over labor, such as those presented by Piketty [53] and Ashford and Hall [6], Stansbury and Summers [64] argue that, historically, workers have shared some of the benefits of productivity increases. Thus, we should not worry about a completely jobless future, but we should stay vigilant about how A.I. could worsen inequality.

The belief that technological change changes the tasks we perform is less controversial. As Bessen [12] points out, while the number of bank tellers did not change when ATMs were introduced, the tasks they carried out did, which required a new set of skills. Therefore, in addition to staying cautious about increasing inequality, the scale, nature, and speed of A.I. proliferation is highly relevant to what re-skilling policies we need to make sure that humans are able to rise to their new tasks.

3.3 Breakdown of Costs

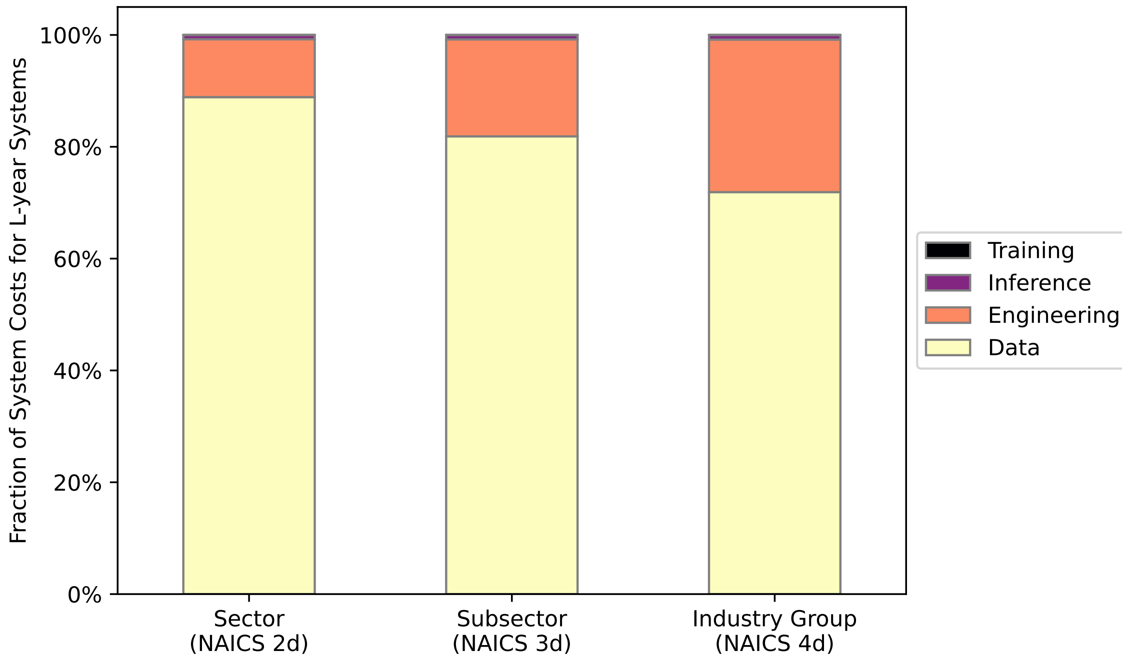
The cost of data is a driver of the overall cost of Computer Vision systems. Figure 3-7 shows that between 70% to 90% of the cost of building and running platform systems consists of data costs. We can infer that the marginally most expensive system has an even higher data and training cost to engineering ratio, since the latter is fixed for any accuracy and entropy of the models, but the data costs are not. Given that the cost per datapoint is not fixed and varies across use cases and acquisition methods, it will be a significant determinant in the proliferation of Computer Vision models. The cost of data can range from $0 where the data is already collected and labeled, e.g., pairing X-ray images with existing medical records, to very high when highly skilled humans are required for labeling, e.g., when the X-ray images need to be both taken and labeled by medical professionals. By taking advantage of the former situation, where it has already been produced as a byproduct of some other business process, costs can be almost eliminated. Furthermore, by enabling data sharing across industries and economies, more actors could derive value from that data, pointing to data sharing as a tool that can drive the possibility, not to mention profitability, of Computer Vision proliferation.

The running costs compared to paying a human being are vanishingly small. In my cost breakdown in Figure 3-7, it is almost invisible to the human eye. This is in contrast to news reports about the exorbitant costs of running large Deep Learning models, including an article by Koetsier in Forbes [40] stating that ChatGPT costs OpenAI millions of dollars every day. However, strictly in terms of replacing existing human tasks, this is reasonable. If we pay a GPU 1.7%[3] of what we pay a human per hour, then the running costs at the minimum viable scale should indeed be close to 1.7% of total system costs. Hence, we have to assume that running cost of ChatGPT is driven by the fact that it is run at a scale where it could potentially replace hundreds of millions of human hours, if correctly aligned.

The overall cost structure of Computer Vision is a cause for concerns about in-

[3] $\frac{0.34}{\$20}$. The cost per GPU hour in our method $0.34 and for this example, we assume an hourly rate for the human of $20.

Figure 3-7: Breakdown of Average Cost of Computer Vision System with Economic Advantage on Platform Level



creased market concentration, and thereby decreased competition. The low running costs point to significant productivity gains when Computer Vision is deployed at scale. Viscusi et al. [71] point out that while efficiency is desirable, it comes with a cost of increased market concentration, which has negative effects for both consumers and workers. Together with the fact that access to data is tied to market incumbent status, these effects can be great. For consumers, there is a risk of anti-competitive behavior and raised prices. For workers, theory and empirical studies by Robinson [55], Boal and Ransom [17], Manning [43], and Schubert et al. [57], find that monopsony power can depress wages. To the extent that A.I. is a general-purpose technology, this could affect all markets.

3.4 Sensitivity Analysis

My conclusions are robust to changes in assumptions. Although the results are slightly different when inputs are changed, the fluctuations do not alter my reasoning in a

Table 3.1: Key Assumptions and Inputs for Sensitivity Analysis

		Low Cost	Base	High Cost
C^{eng}	Engineering costs	0.2x Base	1x Base	2x Base
p^{Δ}	Data costs	\$0	\$0.05	\$5
p^{GPUh}	Cloud pricing	\$0.1	\$0.34	\$1
L	System lifespan	10 years	5 years	2 years
K	Retraining cadence	Never	1 year	2 months
a	Accuracy	a^2	a	$\sqrt{a}$
e	Entropy	5 classes	20 classes	100 classes
	Bare-bones	See Section 2.3.4.		

major way. Table 3.1 provides an overview of the variables I controlled for. I choose the values based on the assumptions I believe are on the extreme ranges of possible costs, including the bare-bones setup from Section 2.3.4 where the only cost is a minimal engineering team. I transform the accuracy using power functions, to ensure that the range of possible values remains between 0 and 1.

The percentage of compensation made up by vision tasks with an economic advantage over human labor on the sector and firm levels for our changing assumptions can be found in Figures 3-8 and 3-9. While different engineering costs, cloud compute costs, and system lifespans barely change the results we obtain compared to our base assumptions, changes to retraining cadence, accuracy, entropy, and data costs are more consequential. This makes sense, given that these factors affect the total spending on data, which, as established in Section 3.3, is already the largest cost item for Computer Vision. They are also the assumptions that I am least certain of, which calls for further research on what a reasonable range is. However, despite this variability, I note that the economic advantage on the firm level for our most generous assumption, i.e., the bare-bones setup at 0.4%, is only about half as big as the most conservative assumption on the platform level, i.e., high accuracy at 0.8%. Further, the gap between sector and firm advantage is large within each assumption. In other words, for values within the range of assumptions in Table 3.1, we will need to see platform business models to realize the full potential of Computer Vision.

Figure 3-8: Economic Advantage with High Cost Assumptions (NAICS 2d)

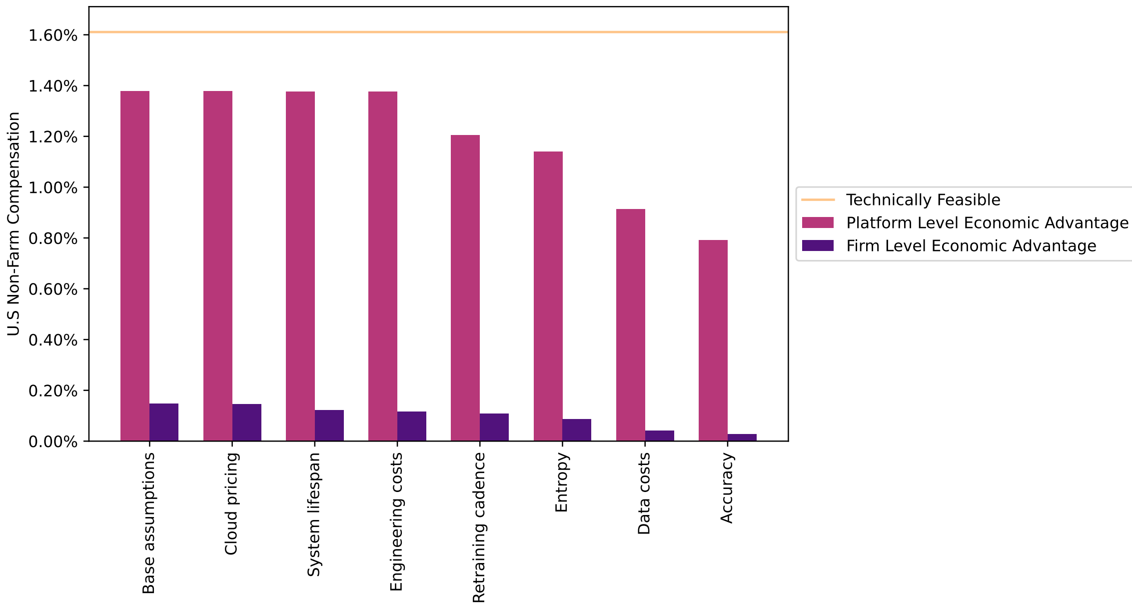


Figure 3-9: Economic Advantage with Low Cost Assumptions (NAICS 2d)

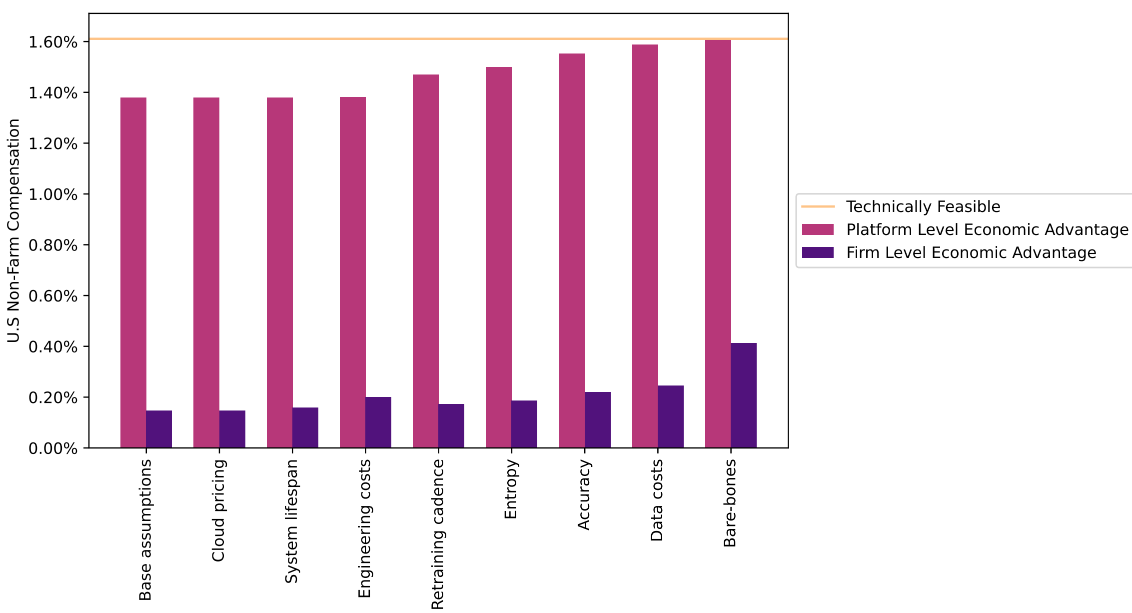
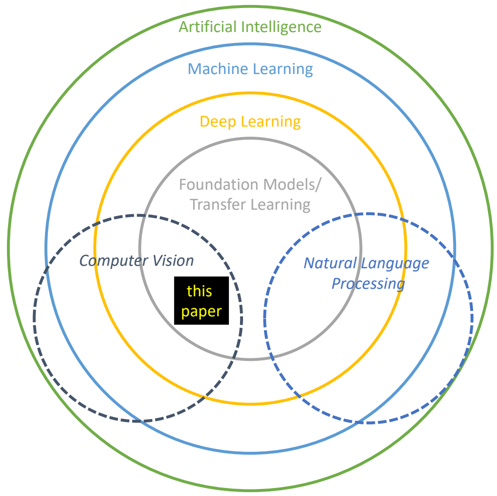


Figure 3-10: Venn Diagram of Artificial Intelligence



3.5 Beyond Computer Vision

The impact of A.I. extends beyond the scope of Computer Vision, and my findings raise questions about whether similar trends can be observed in other areas of Machine Learning. My results are limited to a specific domain of A.I., as illustrated in Figure 3-10, at the intersection of Computer Vision and Foundation Models/Transfer Learning. The release of ChatGPT has brought Natural Language Processing (NLP) into the limelight, and while the impact on the labor market may differ depending on the A.I. domain and the specific tasks involved, some of the insights gained in this study will also be applicable to other A.I. systems that rely on foundation models and transfer learning, such as Large Language Models (LLMs) and NLP.

NLP development, like Computer Vision, often utilizes generalist foundation models that are then fine-tuned to specific applications. For example, ChatGPT is a fine-tuned version of GPT-3 [59]. Although both models are proprietary and owned by OpenAI, other LLMs with similar performance have been released for public use, including Meta's OPT [75] and Google's BLOOM [56] foundation LLMs, and Meta's LLaMa [70], a specialized chat model. Therefore, the landscape of open-source foundation models is similar to that found in Computer Vision. However, there are three key differences between the two that impact economic trade-offs in NLP: (1) the amount of labor offset by language versus vision tasks, (2) the cost of acquiring training data for specific tasks, and (3) the computationally intensive nature of prediction.

NLP tasks make up a bigger portion of worker compensation than those exposed to Computer Vision. Eloundou et al. [29] confirm this, citing that 15% of tasks are exposed to NLP, compared to this study's assessment that only 1.6% of compensation is exposed to Computer Vision. Although including generative Computer Vision in this study would have expanded the set of vision tasks, few tasks in the O*NET database are generative image tasks. On the other hand, generative language tasks are more common. Due to their generative nature, NLP tasks tend to be more time-intensive than Computer Vision tasks, contributing to more significant wage savings for each NLP system compared to Computer Vision systems.

The cost of data acquisition and labeling was the major driver of cost in large Computer Vision systems, but it might contribute less to the cost of generative NLP. LeCun and Misra's observation [42] that NLP can take advantage of self-supervised learning to a larger extent than Computer Vision has implications for the cost of data acquisition. Self-supervised learning utilizes unlabeled data that can be obtained inexpensively, for example, through web-crawling. It remains to be seen whether these models can be fine-tuned using unlabeled data or reinforcement learning. In the case of ChatGPT, the model was trained on data created by humans for the specific purpose of serving as an example of the Q&A format of the chatbot before reinforcement learning was applied, as explained in early press releases on OpenAI's website [58]. However, if an enterprise already has examples of data for the task in the form of customer support chats, email exchanges, or internal knowledge hubs, this data may be cheaper than Computer Vision data. I note that even if data for fine-tuning an NLP system is free, the overall trends compared to Computer Vision would not be drastically different, as seen in Figure 3-9.

There will be differences in how much computing power is necessary to power systems in each of these domains. My conclusions about Computer Vision running costs differ greatly from those one could draw from the Koetsier Forbes article [40], which reports that ChatGPT costs millions of dollars every day. As mentioned above, part of the explanation pertains to scale of deployment. But there are also technical reasons: generative language models make one prediction per token, or word, whereas Computer Vision, specifically image recognition, only does one inference per example. However, given the speed and effort it takes for a human being to write one word, it is still fair to assume that the marginal cost compared to human labor is almost nothing.

My demographic analysis of the impact of Computer Vision is based on the distribution of vision tasks among occupations, which we cannot expect to correlate with language tasks. Studies inspired by ChatGPT, including Felten et al. [31] and Eloundou et al. [29], have found that NLP tasks have a positive correlation with wages, while my results with Computer Vision show a very weak negative correlation. It is possible that the trend of having technically feasible but economically unviable tasks among occupations with graduate degrees could hold true across A.I. domains, but further research is necessary.

While my results are not limited to Computer Vision and NLP, they are limited to A.I. that uses foundation models and transfer learning. Bommasani et al. [18] demonstrate that transfer learning and foundation models are effective not only in NLP and Computer Vision, but also in robotics and strategic games. Other types of deep learning, such as developing models from scratch, would face even more extreme upfront-to-marginal cost ratios. On the other hand, shallow learning, any machine learning type that is not deep learning, is recognized for being less expensive in terms of engineering, data, and computation. Therefore, in cases where shallow learning can achieve the desired inference quality, it may be more economically advantageous.

Chapter 4

Conclusion

I have highlighted the economic cost of Computer Vision compared to the cost of employing human labor. Generally, Computer Vision is more cost-effective than labor if you look across the economy. However, only the largest firms have the minimum viable scale to profitably replace their existing workforce with it. Under my assumptions, data is the main driver of the cost of Computer Vision systems. While Computer Vision is only a small part of the field of A.I., I believe that this cost structure generalizes to other domains that use foundation models and fine-tuning, including the language domain.

The combination of data being the driver behind Computer Vision costs, data being most cheaply collected as a byproduct of existing processes, and only the largest firms having the minimum viable scale for deployment, could make it much harder for small and new firms to compete than it is today. It could aggravate the effects of the increased market concentration stemming from increased efficiency. This would not only have effects on consumers, but it could also lead to depressed wages because of increased monopsony power in the labor market. To the extent that A.I. is a general-purpose technology, this would apply to competition in all markets and industries.

Task replacement could have devastating effects for the individual worker, but a systemic increased market power of large firms could pose a larger threat to the returns of human labor in the aggregate. It is hard to argue against a world where humans do less for more, but when the bounty from productivity is not shared, that vision becomes dystopian. Therefore, A.I. policy needs to, in addition to ensuring algorithmic safety and access to reskilling, also favor market competition. One approach to doing so could include data-sharing frameworks that break down data as a barrier to entry into A.I.-enabled markets.

With policy, we can shape the course of technology and its effects on the world. While we probably should accept an increasingly technological future, we should not accept an increasingly unequal one. The good news is, we do not have to.

Appendix A

O*NET-SOC to SOC

O*NET offers a crosswalk from occupations listed in the O*NET-SOC 2019 taxonomy to the 2018 SOC [49]. However, due to the vast difference in tasks between O*NET occupations that are mapped to the same SOC code, e.g., 11-1011.00 Chief Executive Officer (CEO) and 11-1011.03 Chief Sustainability Officer (CSO), I exclude occupations with a non ".00" decimal notation for which a corresponding ".00" exists. This has the effect that the compensation for all 11-1011 Chief Executive Officers in OEWS is instead allocated to CEO tasks instead of the average of the tasks of the CEOs and CSOs. Because of this logic, 149 out of 1016 occupations are discarded. Although this is a large number, we assume that their absence in the 2018 SOC occupations speaks to their relative small size in the economy. For other cases where multiple O*NET-SOC occupations map to the same SOC occupation, we aggregate all tasks into the same occupation.

In addition, when comparing the datasets, I found multiple occupations without a corresponding SOC-code in the target OEWS dataset, where I created a manual mapping to an occupation with a similar occupation title (Table A.1).

O*NET	O*NET Title	SOC	SOC Title
13-1021	Buyers and Purchasing Agents, Farm Products	13-1020	Buyers and Purchasing Agents

13-1022	Wholesale and Retail Buyers, Except Farm Products	13-1020	Buyers and Purchasing Agents
13-1023	Purchasing Agents, Except Wholesale, Retail, and Farm Products	13-1020	Buyers and Purchasing Agents
13-2022	Appraisers of Personal and Business Property	13-2020	Property Appraisers and Assessors
13-2023	Appraisers and Assessors of Real Estate	13-2020	Property Appraisers and Assessors
21-1011	Substance Abuse and Behavioral Disorder Counselors	21-1018	Substance Abuse, Behavioral Disorder, and Mental Health Counselors
21-1014	Mental Health Counselors	21-1018	Substance Abuse, Behavioral Disorder, and Mental Health Counselors
25-2055	Special Education Teachers, Kindergarten	25-2052	Special Education Teachers, Kindergarten and Elementary School
25-2056	Special Education Teachers, Elementary School	25-2052	Special Education Teachers, Kindergarten and Elementary School
25-9042	Teaching Assistants, Preschool, Elementary, Middle, and Secondary School, Except Special Education	25-9045	Teaching Assistants, Except Postsecondary
25-9043	Teaching Assistants, Special Education	25-9045	Teaching Assistants, Except Postsecondary
25-9049	Teaching Assistants, All Other	25-9045	Teaching Assistants, Except Postsecondary
29-2011	Medical and Clinical Laboratory Technologists	29-2010	Clinical Laboratory Technologists and Technicians
29-2012	Medical and Clinical Laboratory Technicians	29-2010	Clinical Laboratory Technologists and Technicians
31-1121	Home Health Aides	31-1120	Home Health and Personal Care Aides
31-1122	Personal Care Aides	31-1120	Home Health and Personal Care Aides

39-7011	Tour Guides and Escorts	39-7010	Tour and Travel Guides
39-7012	Travel Guides	39-7010	Tour and Travel Guides
45-3031	Fishing and Hunting Workers	–	*Not in OEWS*
47-4091	Segmental Pavers	47-4090	Miscellaneous Construction and Related Workers
47-4099	Construction and Related Workers, All Other	47-4090	Miscellaneous Construction and Related Workers
51-2022	Electrical and Electronic Equipment Assemblers	51-2028	Electrical, Electronic, and Electromechanical Assemblers, Except Coil Winders, Tapers, and Finishers
51-2023	Electromechanical Equipment Assemblers	51-2028	Electrical, Electronic, and Electromechanical Assemblers, Except Coil Winders, Tapers, and Finishers
51-2092	Team Assemblers	51-2090	Miscellaneous Assemblers and Fabricators
51-2099	Assemblers and Fabricators, All Other	51-2090	Miscellaneous Assemblers and Fabricators
51-1042	First-Line Supervisors of Helpers, Laborers, and Material Movers, Hand	53-1047	First-Line Supervisors of Transportation and Material Moving Workers, Except Aircraft Cargo...
51-2043	First-Line Supervisors of Material-Moving Machine and Vehicle Operators	53-1047	First-Line Supervisors of Transportation and Material Moving Workers, Except Aircraft Cargo...
51-1044	First-Line Supervisors of Passenger Attendants	53-1047	First-Line Supervisors of Transportation and Material Moving Workers, Except Aircraft Cargo...
51-1049	First-Line Supervisors of Transportation Workers, All Other	53-1047	First-Line Supervisors of Transportation and Material Moving Workers, Except Aircraft Cargo...

Table A.1: O*NET-SOC to SOC Mapping

Appendix B

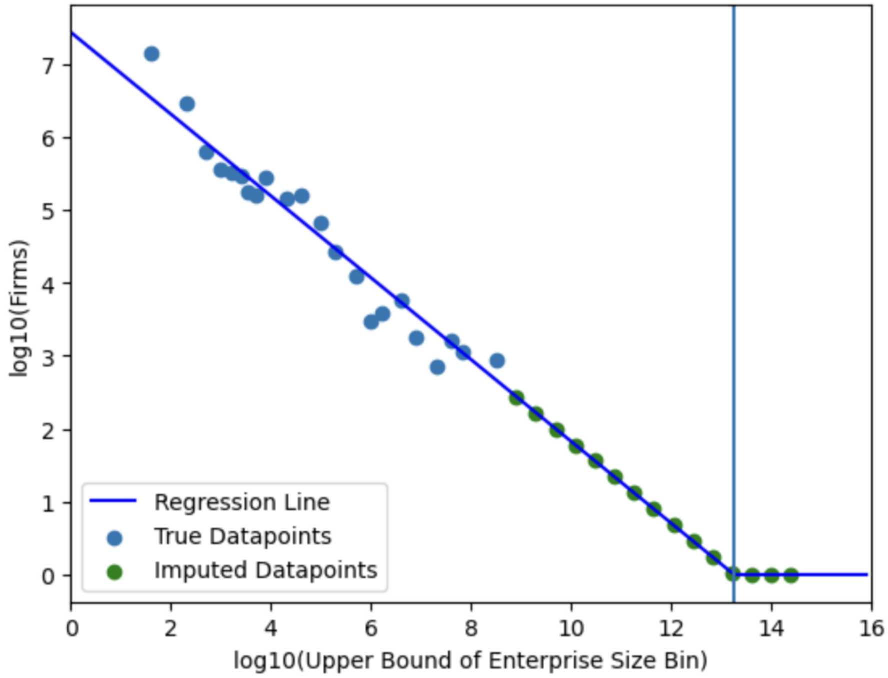
Imputing Firm Size Data

I impute the approximate firm size distributions in each NAICS section of the U.S. economy using the 2019 SUSB Annual Data Tables by Establishment Industry [23]. From the published histogram binned by total cross-section enterprise size, with a catch-all bin for any enterprise larger than 5,000 employees, I first use a power-law assumption, based on Axtell's claim that Zipf's law approximates firm size distributions [11], data about the largest enterprises in the U.S. economy by employment, and a linear regression with custom penalties to estimate the shape of the upper tail of the number of firms per total enterprise size in each section (see Figure B-1a). I then use a parabolic estimator to infer the mean section-specific employment by enterprise size for the inferred bins (see Figure B-1b). The result is a histogram containing section-specific employment per cross-section employment numbers up to and including the size of the largest enterprise by U.S. employment, Walmart, with 1,600,000 employees[1]. I then consider the mean section-specific employment per enterprise size bin when calculating $n_{g,o,s}$.

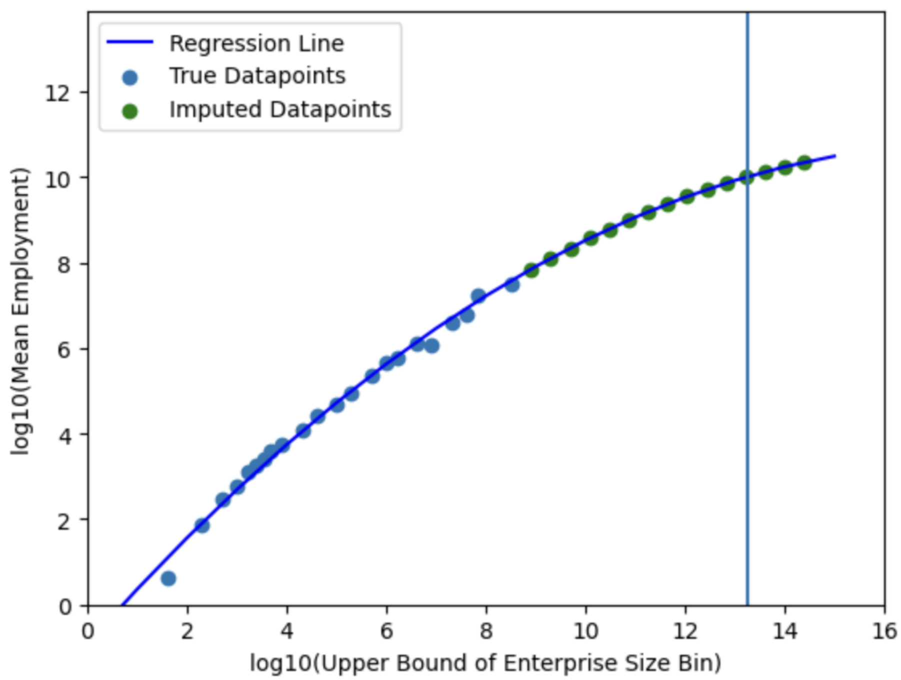
The variable of interest to us is the number of people employed by a given firm within a given NAICS-section. The latter is important because we find higher concentrations of individual occupations for smaller sections of the economy, which we use as an assumption for calculating the minimum firm size to compare to the minimum deployment size. The 2019 SUSB Annual Data Tables [23] contain data relevant for

[1] https://corporate.walmart.com/about

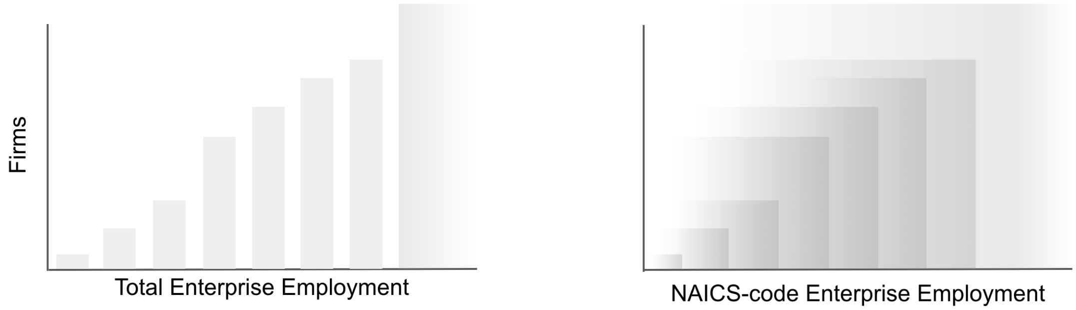
Figure B-1: Imputed Data for NAICS 22 - Utilities



(a) Imputed data for the frequency of firms per total Enterprise Size. The blue vertical line represents the largest imputed firm size.



(b) Imputed data for the mean number of employees per firm in the given NAICS code per total Enterprise Size. The blue vertical line represents the largest imputed firm size, beyond which there is no employment (mean employment times no firms is always 0).



(a) The largest of the histogram bins groups together all enterprises with larger than 5,000 Total Enterprise Employment.

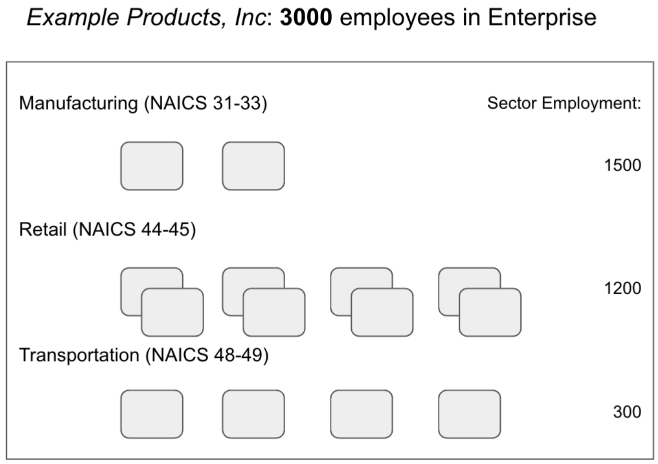
(b) The variable used for the histogram bins, Total Enterprise Employment, is correlated, but not the same, as our variable of interest, the NAICS-section specific Enterprise Employment. However, the Total Enterprise Employment is an upper bound for that variable.

Figure B-2: Limitations of the SUSB Annual Data Tables

this purpose, i.e., the number of Firms and amount of Employment for each NAICS-section of the U.S. economy, in the shape of a histogram on Total Enterprise Size. Although this is the best publicly available option we could find, the problem is twofold: (1) the largest bin has a lower bound of 5,000, and since we expect that the minimum viable deployment size will often be larger than that, we need to split this into smaller bins (see Figure B-2a), and (2) the histogram bins is based on total enterprise size, which is correlated with but not the same as our variable of interest, which is the NAICS-section specific employment (see Figure B-2b). Therefore, I impute the distribution of the number of Enterprises bigger than 5,000 U.S. employees, as well as their mean employment within each given NAICS code.

The first issue to bridge is that while the Employment and Firms data is specific to each of the NAICS codes, the histogram bins represent the total U.S. Enterprise Size. This means that if a company like "Example Products, Inc" has 4000 employees in the country, it will appear in the "2500-4999" bin of the Manufacturing sector, Retail sector, and Transportation sector tables even it will only contribute 500, 1200, and 300 to the Employment column of each of those sectors, respectively (see Figure B-3 for an illustration). Consequently, we need to estimate not only the number of Firms within each histogram bin, but also infer the Employment count to avoid

Figure B-3: Total Enterprise Employment versus NAICS-specific Employment



overestimating the size of each NAICS section.

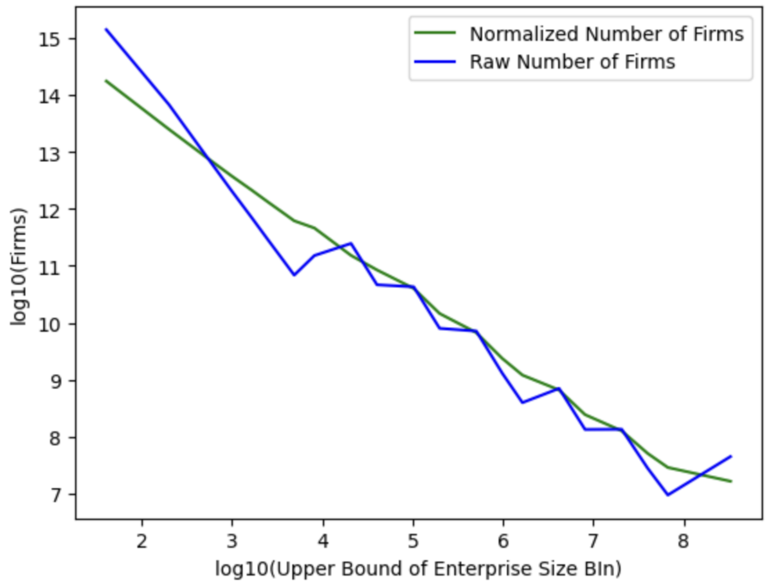
I first define the range of the histogram and the width of each bin in the catchall 5000+ category. The lower bound of the range is 5,000, which is the upper bound of the true data, and the upper bound of the range should be large enough to include the largest private enterprise in America by U.S. employment, Walmart, with 1,600,000 employees[2]. The SUSB data tables use a quasi-exponential bin size, i.e., approximately exponential in size but aligned with base-10 numbers. Therefore, for my imputed bin sizes, I calculate the mean of the log ratio of the upper bound of each bin to the lower bound of the same bin from the SUSB tables and use that as the constant bin-size log ratio for the imputed bins. Because of this quasi-exponential scale of the smaller bins, I also normalize the number of Firms in each of the bins to account for this difference in bin-size (see Figure B-4).

In order to impute the number of Firms in each of the larger bins, I use a linear regression with custom penalties. It has been well-established that firm sizes are distributed in a way that resembles a power law distribution, including by Axtell [11]. Figure B-4 shows that this holds true even for our data, and that a linear regression on the log-log scale presented is not an unreasonable approximation of the

[2]Largest global U.S. employer: `https://en.wikipedia.org/wiki/List_of_largest_United_States%E2%80%93based_employers_globally`, Accessed: 2023-05-08, 1,600,000 U.S. employees according to `https://corporate.walmart.com/about`, Accessed: 2023-05-08

Figure B-4: Normalized Firm Numbers

Normalizing Number of Firms using Mean Log Ratio of Upper to Lower Bound of Bin.

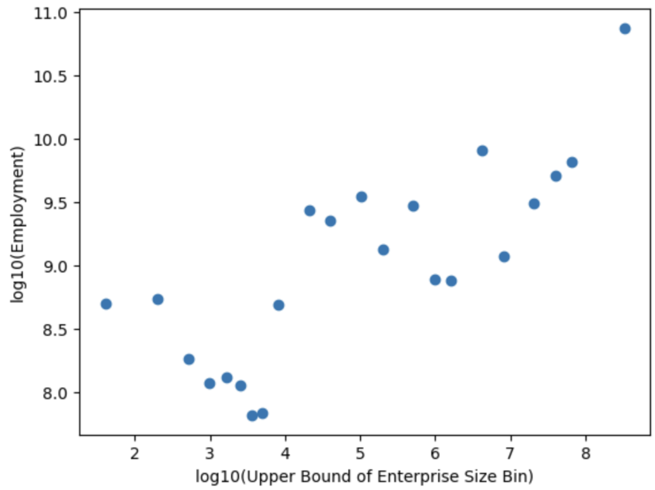


distribution (see Figure B-1a). However, given that we have information about both the exact number of firms in the 5,000+ bin as well as information about the size of the biggest employers in the country[3], in addition to a sum-of-squared residual loss from failing to fit the existing small bins, I include additional squared loss for lines that either poorly estimate the total number of large firms or that claim that the large bins include more firms than actually exist based on the true data about the 15 largest firms. I then multiply the additional loss by the number of small bins available for that NAICS-section as well as a factor of 25 for importance. I use the `optimize.minimize` function from the python package `scipy` to find the optimal line.

The next step is to impute the Employment per bin. When plotting the raw Employment numbers, I find that there is little pattern in its distribution (see Figure B-5) but when plotting the mean employment, it is clear (see Figure B-1b). Using the same approach as for the firms, I devise a loss function as a weighted sum of both the fit to the true data points in the smaller bins as well as an overall sum equating

[3]`https://en.wikipedia.org/wiki/List_of_largest_United_States%E2%80%93based_employers_globally`, Accessed: 2023-05-08, with `https://corporate.walmart.com/about`, Accessed: 2023-05-08, and `https://www.businessinsider.com/amazon-employees-number-1-of-153-us-workers-head-count-2021-7`, Accessed: 2023-05-08, giving us the number of U.S. employees for the two largest employers.

Figure B-5: Employment by Enterprise Size



to the total number for the large bin and minimize it using `scipy`. However, instead of using a linear function, I use a concave parabolic function. An example of the resulting predictions can be seen in Figure B-1b.

Bibliography

[1] Daron Acemoglu, David H. Autor, Jonathon Hazell, and Pascual Restrepo. Artificial intelligence and jobs: evidence from online vacancies. *Journal of Labor Economics*, 40(S1):S293–S340, 2022.

[2] Daron Acemoglu and Pascual Restrepo. The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6):1488–1542, 2018.

[3] Daron Acemoglu and Pascual Restrepo. Tasks, automation, and the rise in U.S. wage inequality. *Econometrica*, 90(5):1973–2016, 2022.

[4] Ajay Agrawal, Joshua Gans, and Avi Goldfarb. *Prediction machines: the simple economics of artificial intelligence*. Harvard Business Press, 2018.

[5] Stuart Armstrong and Kaj Sotala. How we're predicting AI–or failing to. In *Beyond artificial intelligence*, pages 11–29. Springer, 2015.

[6] Nicholas A. Ashford and Ralph P. Hall. *Technology, Globalization, and Sustainable Development*. Routledge, 2018.

[7] David H. Autor. The "task approach" to labor markets: an overview. *Journal for Labour Market Research*, 46(3):185–199, 2013.

[8] David H. Autor. Why are there still so many jobs? the history and future of workplace automation. *Journal of economic perspectives*, 29(3):3–30, 2015.

[9] David H. Autor, Lawrence F. Katz, and Alan B. Krueger. Computing inequality: have computers changed the labor market? *The Quarterly journal of economics*, 113(4):1169–1213, 1998.

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

[11] Robert L. Axtell. Zipf distribution of us firm sizes. *Science*, 293(5536):1818–1820, 2001.

[12] James E. Bessen. Toil and technology: Innovative technology is displacing workers to new jobs rather than replacing them entirely. *Finance & Development*, 52(001):16, 2015.

[13] James E. Bessen, Maarten Goos, Anna Salomons, and Wiljan van den Berge. Automation: A guide for policymakers. *Economic Studies at Brookings Institution: Washington, DC, USA*, 2020.

[14] James E. Bessen, Martin Goos, Anna Salomons, and Wiljan Van den Berge. Automatic reaction-what happens to workers at firms that automate? 2019.

[15] James E. Bessen, Stephen Michael Impink, Lydia Reichensperger, and Robert Seamans. The business of AI startups. *Boston Univ. School of Law, Law and Economics Research Paper*, (18-28), 2018.

[16] Matthew Bishop. *Essential economics: an A to Z guide*, volume 22. John Wiley & Sons, 2009.

[17] William M. Boal and Michael R. Ransom. Monopsony in the Labor Market. *Journal of Economic Literature*, 35(1):86–112, 1997.

[18] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

[19] Nicholas J. Borge. Deep pockets: The economics of deep learning and the emergence of new AI platforms. Master's thesis, Massachusetts Institute of Technology, 2022.

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

[21] Erik Brynjolfsson and Andrew McAfee. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company, 2014.

[22] Erik Brynjolfsson, Tom Mitchell, and Daniel Rock. What can machines learn, and what does it mean for occupations and the economy? In *AEA papers and proceedings*, volume 108, pages 43–47, 2018.

[23] U.S. Census Bureau. 2019 SUSB Annual Data Tables by Establishment Industry, 2021.

[24] Michael Chui, James Manyika, and Mehdi Miremadi. Where machines could replace humans-and where they can't (yet). *McKinsey Quarterly*, 2016.

[25] European Commission. Communication. Towards a Common European Data Space, 2018.

[26] European Commission. Data Act: Commission Proposes Measures for a Fair and Innovative Data Economy, 2023. `https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113` Accessed: 2023-03-12.

[27] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE, 2009.

[28] David Dranove and Craig Garthwaite. Artificial Intelligence, the Evolution of the Healthcare Value Chain, and the Future of the Physician. Technical report, National Bureau of Economic Research, 2022.

[29] Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models, 2023.

[30] Edward W. Felten, Manav Raj, and Robert Seamans. A Method to Link Advances in Artificial Intelligence to Occupational Abilities. In *AEA Papers and Proceedings*, volume 108, pages 54–57, 2018.

[31] Edward W. Felten, Manav Raj, and Robert Seamans. How will Language Modelers like ChatGPT Affect Occupations and Industries? *arXiv preprint arXiv:2303.01157*, 2023.

[32] Martin Fleming. *Breakthrough: A Growth Revolution*. Business Expert Press, 2022.

[33] Martin Fleming, Wyatt Clarke, Subhro Das, Phai Phongthiengtham, and Prabhat Reddy. The Future of Work: How New Technologies Are Transforming Tasks. *MIT-IBM Watson AI Lab*, 2019.

[34] Carl Benedikt Frey and Michael A. Osborne. The Future of Employment: How Susceptible are Jobs to Computerisation? *Technological forecasting and social change*, 114:254–280, 2017.

[35] Jay Fuchs. How Close Rates are Shifting in 2023 [New Data], Sep 2022. `https://blog.hubspot.com/sales/new-sales-close-rate-industry-benchmarks-how-does-your-close-rate-compare`, Accessed: 2023-04-04.

[36] Hrothgar John Habakkuk. *American and British Technology in the Nineteenth Century: the Search for Labour Saving Inventions*. Cambridge University Press, 1962.

[37] Michael T. Hannan and John Freeman. Structural Inertia and Organizational Change. *American sociological review*, pages 149–164, 1984.

[38] Marius Hobbhahn and Tamay Besiroglu. Trends in GPU price-performance, 2022. `https://epochai.org/blog/trends-in-gpu-price-performance`, Accessed: 2023-03-04.

[39] John Maynard Keynes. Economic Possibilities for our Grandchildren. In *Essays in persuasion*, pages 321–332. Springer, 1933.

[40] John Koetsier. ChatGPT Burns Millions Every Day. Can Computer Scientists Make AI One Million Times More Efficient? *Forbes*, Feb 2023. https://www.forbes.com/sites/johnkoetsier/2023/02/10/chatgpt-burns-millions-every-day-can-computer-scientists-make-ai-one-million-times-more-efficient/?sh=53c980646944, Accessed: 2023-04-04.

[41] Alan B. Krueger. How Computers Have Changed the Wage Structure: Evidence From Microdata, 1984–1989. *The Quarterly Journal of Economics*, 108(1):33–60, 1993.

[42] Yann LeCunn and Ishan Misra. Self-supervised Learning: The Dark Matter of Intelligence. *Meta AI Research*, 2021. https://ai.facebook.com/blog/self-supervised-learning-the-dark-matter-of-intelligence/, Accessed: 2023-04-04.

[43] Alan Manning. Monopsony in Motion. In *Monopsony in Motion*. Princeton University Press, 2013.

[44] Guy Michaels, Ashwini Natraj, and John Van Reenen. Has ICT Polarized Skill Demand? Evidence from Eleven Countries Over Twenty-Five Years. *Review of Economics and Statistics*, 96(1):60–77, 2014.

[45] Jose G. Moreno-Torres, Troy Raeder, Rocío Alaiz-Rodríguez, Nitesh V Chawla, and Francisco Herrera. A Unifying View on Dataset Shift in Classification. *Pattern recognition*, 45(1):521–530, 2012.

[46] Gavin Mueller. *Breaking Things at Work: The Luddites are Right About Why You Hate Your Job*. Verso Books, 2021.

[47] John B. Murphy. Introducing the North American Industry Classification System. *Monthly Lab. Rev.*, 121:43, 1998.

[48] U.S. Bureau of Economic Analysis. Fixed Assets and Consumer Durable Goods in the United States, 1925–97, 2003.

[49] U.S. Department of Labor. Crosswalk O*NET-SOC 2019 to 2018 SOC, Feb 2023. https://www.onetcenter.org/taxonomy/2019/soc.html, Accessed: 2023-04-04.

[50] U.S. Department of Labor. O*NET, Feb 2023. https://www.onetcenter.org/overview.html, Accessed: 2023-04-04.

[51] U.S. Bureau of Labor Statistics. Standard Occupational Classification (SOC), 2018. https://www.bls.gov/soc/, Accessed: 2023-04-04.

[52] U.S. Bureau of Labor Statistics. Employer Costs for Employee Compensation - SEPTEMBER 2022, 2022. https://www.bls.gov/news.release/pdf/ecec.pdf, Accessed: 2023-04-04.

[53] Thomas Piketty. *Capital in the 21st Century*. Harvard University Press Cambridge, MA, 2014.

[54] Thomas Plötz and Gernot A Fink. Markov Models for Offline Handwriting Recognition: a Survey. *International Journal on Document Analysis and Recognition (IJDAR)*, 12(4):269–298, 2009.

[55] Joan Robinson. *The Economics of Imperfect Competition*. Springer, 1933.

[56] Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, et al. Bloom: A 176b-parameter Open-Access Multilingual Language Model. *arXiv preprint arXiv:2211.05100*, 2022.

[57] Gregor Schubert, Anna Stansbury, and Bledi Taska. Employer Concentration and Outside Options. *Available at SSRN 3599454*, 2021.

[58] John Schulman. ChatGPT: Optimizing Language Models for Dialogue. *OpenAI*, 2022. `https://web.archive.org/web/20221130180912/https://openai.com/blog/chatgpt/`, Accessed: 2023-04-04.

[59] John Schulman, Barret Zoph, Christina Kim, Jacob Hilton, Jacob Menich, Jiayi Weng, Juan Felipe, Ceron Uribe, Liam Fedus, Luke Metz, Michael Pokorny, Rapha Gontijo Lopes, Shengjia Zhao, Arun Vijayvergiya, Eric Sigler, Adam Perelman, Chelsea Voss, Mike Heaton, Joel Parish, Dave Cummings, Rajeev Nayak, Valerie Balcom, David Schnurr, Tomer Kaftan, Chris Hallacy, Nicholas Turley, Noah Deutsch, Vik Goel, Jonathan Ward, Aris Konstantinidis, Wojciech Zaremba, Long Ouyan, Leonard Mogdanoff, Joshua Gross, David Medina, Sarah Yoo, Teddy Lee, Ryan Lowee, Dan Mossing, Joost Huizinga, Roger Jiang, Carroll Wainwright, Diogo Almeida, Steph Lin, Marvin Zhang, Kai Xiao, Katarina Slama, Steven Bills, Alex Gray, Jan Leike, Jakub Pachoki, Phil Tillet, Shantanu Jain, Greg Brockman, Nick Ryder, Alex Paino, Qiming Yuan, Clemens Winter, Ben Wang, Mo Bavarian, Igor Babuschkin, Szymon Sidor, Ingmar Kanitscheider, Mikhail Pavlov, Matthias Plappert, Nik Tezak, Heewoo Jun, William Zhuk, Vitchyr Pong, Lukasz Kaiser, Jerry Tworek, Andrew Carr, Lilian Weng, Sandhini Agarwal, Karl Cobbe, Vineet Kosaraju, Alethea Power, Stanislas Polu, Jesse Han, Raul Puri, Shawn Jain, Benjamin Chess, Christian Gibson, Oleg Boiko, Emy Parparita, Amin Tootoonchian, Kyle Kosic, and Christopher Hesse. Introducing ChatGPT. *OpenAI*, 2022. `https://openai.com/blog/chatgpt`, Accessed: 2023-04-04.

[60] Jaime Sevilla, Lennart Heim, Anson Ho, Marius Hobbhahn, Tamay Besiroglu, and Pablo Villalobos. Estimating training compute of deep learning models. Technical report, Tech. rep, 2022.

[61] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[62] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[63] Ioannis Stamelos, Lefteris Angelis, Maurizio Morisio, Evaggelos Sakellaris, and George L Bleris. Estimating the Development Cost of Custom Software. *Information & Management*, 40(8):729–741, 2003.

[64] Anna M. Stansbury and Lawrence H. Summers. Productivity and Pay: Is the Link Broken? *National Bureau of Economic Research*, 2018.

[65] Bob Sullivan. Average stock market return, Feb 2023. `https://www.forbes.com/advisor/investing/average-stock-market-return/`, Accessed: 2023-05-08.

[66] Neil C. Thompson. NVIDIA: Building a Compute and Data Platform for Self-Driving Cars. Dec 2021. `http://www.neil-t.com/wp-content/uploads/2022/01/NVIDIA-case-study-2021-12-14.pdf`, Accessed: 2023-04-04.

[67] Neil C. Thompson, Nicholas J. Borge, Aparna Pande, and Martin Fleming. Demand Forecasting with A.I.: Building the Business Case, 2021.

[68] Neil C. Thompson, Martin Fleming, Subhro Das, Brian Goehring, and Nicholas J. Borge. Where is it Cost Effective To Deploy AI: MIT-IBM Industry Showcase, 2022.

[69] Neil C. Thompson, Martin Fleming, Ben Tang, Anna Pastwa, Brian Goehring, and Subhro Das. An Economic Model for AI: Evaluating Cost-Performance Tradeoffs in Deep Learning Technique Selection. 2022. Forthcoming.

[70] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and Efficient Foundation Language Models. *arXiv preprint arXiv:2302.13971*, 2023.

[71] W. Kip Viscusi, Joseph E. Harrington Jr, and David E.M. Sappington. *Economics of Regulation and Antitrust.* MIT press, 2018.

[72] Michael Webb. The Impact of Artificial Intelligence on the Labor Market. *Available at SSRN 3482150*, 2019.

[73] Naomitsu Yashiro, Tomi Kyyrä, Hyunjeong Hwang, and Juha Tuomala. Technology, labour market institutions and early retirement: evidence from Finland. *VATT Institute for Economic Research Working Papers*, 136, 2020.

[74] Gingfung Yeung, Damian Borowiec, Adrian Friday, Richard Harper, and Peter Garraghan. Towards GPU Utilization Prediction for Cloud Deep Learning. In *Proceedings of the 12th USENIX Conference on Hot Topics in Cloud Computing*, pages 6–6, 2020.

[75] Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, et al. OPT: Open Pre-Trained Transformer Language Models. *arXiv preprint arXiv:2205.01068*, 2022.

[76] Nikolas Zolas, Zachary Kroff, Erik Brynjolfsson, Kristina McElheran, David N Beede, Cathy Buffington, Nathan Goldschlag, Lucia Foster, and Emin Dinlersoz. Advanced Technologies Adoption and Use by U.S. Firms: Evidence From the Annual Business Survey. *National Bureau of Economic Research*, 2021.