

The Economic Impacts of AI: A Multidisciplinary, Multibook Review

Kevin A. Bryan*

August 29, 2025

Abstract

This essay reviews seven books from the past dozen years by social scientists examining AI's economic impact. These works offer valuable insights - AI as cheap prediction, architectural barriers to adoption, data as an economic asset, implementation challenges. However, they offer little guidance when it comes to the transformative scenarios considered plausible by many AI researchers. Economists have made great progress in explaining how to use AI within existing production functions, who benefits, and why; what remains needed is rigorous advice to policymakers concerned about rapid increases in labor churn, scientific development, labor-capital shifts, or existential risk.

1 Introduction

Consider the following. “We are on course for AGI by 2027. These AI systems will basically be able to automate basically all cognitive jobs...To be clear - the error bars are large”

*Thanks to David Romer, Anton Korinek and an anonymous referee for very valuable comments. This essay was written by the author, with AI assistance only on the grammar and style editing process, as well as during brainstorming of structure.

(Aschenbrenner, 2024). “Anyone in 2035 should be able to marshal the intellectual capacity equivalent to *everyone* in 2025; everyone should have access to unlimited genius to direct however they can imagine” (Altman, 2025). “On average, it takes...10 years and billions of dollars to design just one drug. We can maybe reduce that from years to maybe months or maybe even weeks” (60 Minutes, 2025). “AI could wipe out half of all entry-level white-collar jobs - and spike unemployment to 10-20% in the next one to five years” (VandeHei and Allen, 2025).

These four claims do not originate from science fiction or uninformed pundits. They are from ex-OpenAI researcher Leo Aschenbrenner’s “Situational Awareness” treatise, an essay by OpenAI cofounder Sam Altman, an interview with Nobel Prize-winning DeepMind founder Demis Hassabis, and an interview with Anthropic founder Dario Amodei. One can certainly argue with these claims - certainly, many even in the computer science community do! - and they won’t go unchallenged in this review. But let us begin by agreeing that *if* we take these claims seriously, the core economic questions related to AI would be very large indeed: will we see hyper-growth, rapid shocks to the nature of the labor force, wild swings in interest rates and the value of assets, large spikes in electricity demand, or radical changes in the structure of firms and other organizations? Are we on the leading edge of a Third Industrial Revolution, only run in a half-decade rather than a half-century? Is the pace of scientific progress about to accelerate wildly?

Bob Lucas’s dictum comes to mind, that once one starts thinking about growth, it’s hard to think about anything else (Lucas, 1988). Once industry and technological leaders start throwing around double-digit growth rates, double-digit unemployment rates, order-of-magnitude improvements in the productivity of research, it breaks one’s ability to concentrate on more pedestrian concerns. Thus, I must apologize to the authors whose books on AI are covered in this joint review. These books are, in most cases, wonderfully written, deep, theoretically compelling, and novel. We will absolutely see why. Yet, after reading the entire collection, I am left with the feeling that I am no more able to respond to the questions

above than I was before.

It is not that economists are neglecting the effects of AI; they are studying them in every dimension, and recently with vigor. Until the inaugural NBER Economics of AI conference in 2017, not a single NBER Working Paper had contained the words “artificial intelligence” in its title.¹ There have been 16 such papers just in the first five months of 2025. Focusing on generative models, there were 130 NBER working papers and chapters as of May 2025 using the term “large language model” and 103 using “generative AI”, all but four of which were written after ChatGPT’s release in November 2022. The effect on labor markets, growth, health care, asset pricing, macroeconomic forecasting, innovation: all are being investigated.²

But books play a unique role. Research articles construct a literature. Books summarize it; they situate research articles in a broader context; they draw out implications; they take stands. Not all books need to do all of this, but books are an important vector by which the aggregated knowledge of research journals reaches the public and non-subject-matter experts. This review therefore examines seven books written in the past dozen years by social scientists on AI. “The Second Machine Age” (Brynjolfsson and McAfee, 2014) offers an early argument that changes in computation and digitization were leading to an Industrial Revolution-sized economic shift; “Prediction Machines” and “Power and Prediction” (Agrawal et al. (2018) and Agrawal et al. (2022)) provide a particularly compelling framework for the basic economic feature of AI, its role in reducing the cost of prediction; “The Data Economy: Tools and Applications” (Baley and Veldkamp, 2025) covers the economic theory of data, an important input into that prediction; “The Skill Code” (Beane, 2024) and “Co-Intelligence” (Mollick, 2024) examine practical implementation challenges for AI, via sociology and management research, that are frequently misunderstood by industry practitioners; and “Situational Awareness” (Aschenbrenner, 2024), a book-length treatise self-published online for speed reasons, offers a view from Silicon Valley about the most

¹Though of course there were some fantastic papers on the topic that didn’t use that precise term in their titles, such as Nordhaus (2015) and Acemoglu and Restrepo (2016).

²See, for instance, Section VII of the Oxford Handbook of AI Governance, covering economic issues (Bullock et al., 2024).

transformative possibilities of AI.

Before turning to these texts, we must understand why some in Silicon Valley see AI as so epochal, so economically transformative. We will need some preliminaries - what AI is, how it developed, and what computer scientists are seeing on a technical level. With that background, we can better understand what they will find valuable in books from our field, and where our advice is still needed. The reader should be aware that the boldest claims about the future of AI are subject to a great deal of debate, and uncertainty. One need not *agree* with the claims bandied about in Silicon Valley, either on technical capabilities or their implications. Indeed, even within Silicon Valley and the AI world more broadly, there are disagreements here. All I would like you to keep in mind is that the bold claims are taken seriously by enough people on the AI research and AI policy side that social scientists risk being left out of important conversations if we do not tackle at least the *conditional* economic implications of these technical predictions.

2 Why Silicon Valley Believes What It Does

The conviction among many AI researchers that it will unleash unique economic effects derives from two decades of empirical computer science breakthroughs.³

Throughout the 1980s and 1990s, AI largely meant expert systems. These were elaborate collections of if-then statements that captured human knowledge in rigid logical structures. For instance, language translation might involve trying to break down the syntax of the origin and target language into component pieces, hardcoding in contexts where knowing means *savoir* instead of *connaître*. These systems could excel in their narrow domains, but remained fundamentally brittle. They required armies of engineers to meticulously codify expertise, and they often failed catastrophically when confronted with anything even slightly outside their programmed parameters (Goodfellow et al., 2016). There is a natural analogy

³I will omit the mathematical details in this review, but Ananthaswamy (2024) is a fantastic and very readable book explaining concepts we'll see below like backpropagation without crawling too far into the weeds.

to the experience of economists with large-scale models of the economy in the 1960s and 1970s: hardcoding rules like Okun’s Law became very problematic when the model was applied to the more complex real economy, especially in the presence of strategic actors, as the Lucas Critique made clear (Lucas, 1976).

Deep learning offered a different methodology: instead of trying to codify rules, let an algorithm discover the rules that matter for prediction by fitting a highly nonlinear function on a large set of training data.⁴ These so-called neural networks took features of input data (independent variables in a statistical model), added “layers” of transformations of these variables, then tried to choose weights on the transformed variables at each layer to best predict outcomes in a test set without overfitting. The basic idea of the neural network is not new; they had shown promise in the 1980s and 1990s on small tasks like character recognition. Early networks were shallow, typically limited to one or two “layers” between input and output, which restricted the nonlinear patterns they could match. And when it came to the structure of the data the neural network tried to capture, practical problems arose, chief among them the vanishing gradient problem which made it difficult to train “deep” models. Essentially, if there were hundreds of layers in the network, a positive or negative update on the final result, mapped via the chain rule through the entire network, approaches zero impact on early parts of the network. This means that the earliest layers, which are responsible for extracting the most fundamental features from the data, receive almost no meaningful adjustment during training.

To see how this works, take thousands of sentences either in English or French. How do we know which is which? Let’s tag a subset of them with ground truth of their language. We will try to predict an outcome 1 if English, and 0 if French. The analyst chooses the input representation, such as letter count vectors \mathbf{x} . The first hidden layer computes $z_i = f(\mathbf{w}_i^T \mathbf{x} + b_i)$ for each node i , where f is a function like a sigmoid, and weights are initially randomized. Different nodes learn different linguistic patterns during training; one might

⁴For an up-to-date overview on deep learning for economists, see Dell (2025).

develop high weights for the letter “r” while another might learn low weights for accented letters. The second hidden layer then creates more abstract combinations $h_j = f(\mathbf{u}_j^T \mathbf{z} + d_j)$, potentially learning to detect concepts like “formal English text” by combining outputs from multiple first-layer nodes. The important idea is that hidden layers weighing heavily on the final prediction may involve features that the analyst had not considered. Finally, the output layer produces the classification probability $p = \sigma(\mathbf{v}^T \mathbf{h} + c)$. This hierarchical feature learning, where each layer builds increasingly abstract representations from the layer below, emerges automatically from gradient descent optimization - using the ground truth of a training set and updating the model “backwards” via the chain rule, and not from explicit programming of linguistic concepts. Theoretical developments to help ensure that incorrect predictions led to useful updated weights throughout the model - such as the ‘ReLU’ activation function whose gradient is either 0 or 1 and hence doesn’t multiplicatively shrink - allowed the fitting of models with many hidden layers.

From that basic idea, the resurgence of deep learning came through two further breakthroughs: the explosion of available data, especially text and images digitized online, gave models enough examples to learn from, and advances in hardware, particularly GPUs, enabled the quick training of large networks (LeCun et al., 2015). In 2012, a team including future Nobel Prize winner Geoffrey Hinton and OpenAI cofounder Ilya Sutskever entered a highly-regarded image recognition competition called ImageNet. Their deep convolutional neural network achieved an error rate of 15.3%, smashing the previous best of 26.2% (Krizhevsky et al., 2012). The “ImageNet shock” proved that deep learning from large datasets, without hardcoded rules, could outperform state-of-the-art models in some classification tasks. More importantly, the technique was general, not specifically linked to classifying images. Success stories cascaded across domains: speech recognition error rates plummeted, machine translation quality surged, and deep learning models began mastering complex games like chess and Go.

Reinforcement learning added another dimension to this optimism, particularly through

what AI pioneer Richard Sutton called “The Bitter Lesson” (Sutton, 2019). Consider training a model to play chess. You can hardcode rules for good play, perhaps involving brute force search many moves ahead to spot mistakes, as Deep Blue did for its famous match with Kasparov. Alternatively, you can take a database of millions of games and try to fit a deep learning model that most closely matches actual play. Of course, there are many mistakes in the data, but since the mistakes are not perfectly correlated and the model will learn the most common move on any given board, it will vastly outperform the average player. You can augment that deep learning model by hardcoding in certain “features” like well-known openings or capture patterns.

However, Sutton observed that AI history is littered with examples of clever, domain-specific algorithms that initially outperformed more general methods, only to be surpassed as computational power grew. A reinforcement learning approach taking Sutton’s lesson seriously would take a general model - no hardcoded rules other than “this particular move is legal” and no set of pre-existing human games - then have the model play itself billions of times. The winning moves would have their weights in the deep learning model increased. DeepMind’s AlphaGo exemplified this lesson perfectly. The system initially learned Go by training on vast datasets of human expert games. But its successor, AlphaZero, achieved superhuman performance across Go, chess, and shogi by learning entirely through self-play, starting with nothing but the basic rules (Silver et al., 2018). This demonstration - that an AI system could not only match but significantly surpass human understanding in complex strategic domains by discovering entirely novel strategies - suggested that AI systems might eventually contribute to their own advancement. Reinforcement learning can be applied even in scenarios outside games or mathematics with “ground truth”; for instance, to train a French-English translator, you might use a weaker large language model (LLM) to rate, imperfectly, the output of the reinforcement learning model iterations. As long as the ratings are positively correlated with good translation, the model can improve itself using those ratings as a reward function.

Silicon Valley did not miss the implications. The founding of OpenAI in late 2015, backed by \$1 billion in funding, marked a crucial inflection point (OpenAI, 2015). The 2017 introduction of the Transformer architecture in “*Attention Is All You Need*” led the shift from general deep learning models to generative models like LLMs (Vaswani et al., 2017). The paper, from a team at Google, has been cited nearly 200,000 times in less than eight years, and was cited more in 2024 than any other paper in any single year in the history of science.⁵

The core innovation of the Transformer was the *self-attention* mechanism. Suppose a paragraph discusses Schumpeter and, 100 words later, says “the key to his idea of creative...”. A human reader can see that *Schumpeter*, *idea*, and *creative* are the most important clues, and that the next word is very likely “destruction.” A model that predicts the next word from only the single previous word would pair *creative* with something like *songs*. A 4-gram model (predicting the next word based on the previous four) might guess “his idea of creative writing.” The challenge is that some of the most important cues (*Schumpeter*, *his*) are far away from *creative*. Older approaches such as Recurrent Neural Networks address this by carrying forward a hidden state step-by-step when predicting the following word, weighting each new word in sequence. While this can, in principle, preserve long-range information, the signal often degrades and training is slow because every step depends on the previous one. Self-attention instead computes, for each position, relevance weights to *all* other positions in the sequence in parallel, allowing *creative* to connect directly to *Schumpeter* in a single hop. This removal of the sequential bottleneck dramatically speeds up training and enables much larger models to be trained on unprecedented amounts of text.

Why should we care about predicting the next word⁶ in a given sequence? The core driver of computer scientists’ optimism about AI and the economy comes from the empirical discov-

⁵As measured by Google Scholar via the author’s calculation. The most-cited 21st century paper in the American Economic Review, “The colonial origins of comparative development: An empirical investigation”, has been cited in 25 years as much as Attention is All You Need gets cited every 4 months.

⁶Technically, these models are predicting the next token, often instantiated as a portion of a word but in multimodal models a slightly different object.

ery of so-called “scaling laws”⁷. Researchers from OpenAI and DeepMind found remarkably consistent relationships between model performance for practical tasks and three factors: computational resources for training, dataset size, and model parameters. Increases in resources yield predictable improvements in performance (Kaplan et al., 2020; Hoffmann et al., 2022). For several years, these laws showed little sign of diminishing returns, suggesting that continued aggressive investment in compute and data would yield quantifiable improvements in capabilities. In a sense, the future capability of large language models to solve a variety of problems has been predictable. More recently, so-called “test-time inference”, where the model is trained via reinforcement learning to use tokens to self-reflect before proposing a final response to a query, has shown evidence of its own scaling law, leading in the short run to impressive benchmark performance from models like OpenAI’s O-series, and “thinking” models from Google DeepMind, Anthropic, and DeepSeek (Snell et al., 2024).

These scaling laws explain why much of the breathtaking AI progress in recent years has felt almost predictable to researchers within the field. The trajectory from GPT-2 to GPT-3 to GPT-4 largely followed expected curves of capability improvements commensurate with increases in computational power, dataset size, and parameter counts (Brown et al., 2020; OpenAI, 2023). Researchers could forecast next-generation model performance with reasonable accuracy by extrapolating from these laws. When AI labs make statements about capabilities in 2027 or 2029, they are effectively just “following lines on a (logarithmic) graph” using scaling laws. To be clear, all lines on a graph predict the future until they don’t, and the theoretical basis of each of these scaling laws is quite tenuous. Nonetheless, as an empirical regularity, it has proven quite fruitful over the past half-decade.

Perhaps the most transformative element fueling expectations of first-order economic effects is AI’s potential to contribute directly to its own research and development. Current AI models already assist human researchers in writing code, summarizing literature, generating hypotheses, and designing experiments. As capabilities advance, AI systems will take the

⁷Perhaps “scaling regularity” is better; they are simply an empirical observation like Moore’s Law, not a physical constant.

lead in certain aspects of their own development. This recursive improvement dynamic need not achieve speculative levels of “hard takeoff” - where the pace of AI development becomes very rapid as prior AI models self-improve - to be economically revolutionary. Even modest but sustained acceleration in AI researcher productivity would lead to rapidly compounding progress. It is not hyperbolic to say that many researchers at top AI labs believe that *science itself* is an AI problem. AI will become adept at writing code autonomously, allowing AI to rapidly speed up its own capabilities, in turn allowing that even better AI to help solve open problems in robotics, biology, and so on.

This optimistic outlook isn’t universally shared, of course. Yann LeCun, a Turing Award laureate for his foundational work on neural networks, has argued that LLMs have fundamental limits (LeCun, 2022). Despite their fluent text generation and impressive pattern-matching capabilities, these models do not genuinely reason about causality or possess common sense in human-like ways. LeCun advocates for alternative architectures that aim to learn more comprehensive world models; humans, after all, appear to learn much more efficiently from a limited number of examples of a new task than machine learning models. For an economist, it remains unclear whether this critique is optimistic or pessimistic. That we may have alternative, very productive neural network architectures we have not even started exploring yet offers another way around binding constraints on performance if LLM capabilities do, in fact, stop improving.

Nonetheless, particularly when it comes to predictions more than a few months ahead, there are sound reasons to be skeptical of the boldest claims. Nontrivial technical issues must be overcome if AI - particularly extensions of transformer-based models - is to play a key role in advancing science or recursively self-improving its own performance. On a benchmark intended to test whether frontier models could replicate recent top AI research papers, including data analysis, from scratch, they did so less than 25% of the time. On a more difficult challenge to, with access to a container on which to run code and access internal logs, see whether models could solve internal OpenAI engineering bottlenecks that

took human researchers at least one day to solve, frontier models can still only solve these 2% of the time as of August 2025 (OpenAI, 2025). Reinforcement learning is most amenable to problems with well-defined reward functions - games like chess, coding exercises, and so on - and it is not at all clear how to structure a deep learning model to perform well on more amorphous tasks where “good” performance is harder to define (see Narayanan and Kapoor (2024), by two computer scientists, for a skeptical take on current model ability to overcome these difficulties).

Perhaps these limits should be seen as an open question, where economists should not take the optimistic case as gospel, but still take seriously their possibility. For example, writing clear proofs is not obviously a reinforcement learning problem. AI has certainly made progress; frontier models can score at IMO Gold level and in the FrontierMath benchmark (<https://epoch.ai/frontiermath>), where mathematicians wrote a series of problems they expected AI to not be able to solve, the fraction of problems solved by the best models has increased from 2.07% to 24.83% between December 2024 and August 2025. And yet: where are the completely novel proofs being added to the literature? Where are the novel scientific breakthroughs? The bet being made by the optimists is that the rate of progress on benchmarks is rapid, and that performance on these benchmarks is correlated with general capabilities; “if you’re so sure AI won’t be able to do X next year, give me the problem you think it won’t solve” is a common refrain. Others quite reasonably point out that the mapping between solving a problem on paper and solving more open-ended economically-relevant tasks may not be equivalent.

To push back even more against the optimists, bolder predictions are generally about technical capabilities of a model, not about how those capabilities map into economic activity. A technology being *able* to do something does not mean the technology will be used to do that thing, nor it is obvious what being able to, or actually, doing a given task means for equilibrium aggregate economic outcomes. So for AI researchers who accept the optimistic technical scenario, many questions about economic impact remain. How would, say, a

speedup in robotics research affect wages? Inequality? The return to education? Predicting well the technical capabilities in terms of sensor density on a robot in 2030 does not answer those questions. Should we be concerned about market concentration in frontier AI? Are we perhaps investing too little if foundation model developers cannot appropriate most of the social returns of the models (precisely the argument made to me by a Chinese AI researcher about why China’s government is so positive on open source models)? These questions presuppose the technical development of AI well beyond today’s capabilities. To these AI researchers, asking about the capability of chatbot interface AI interactions in 2025 is like writing a book on automobiles in 1900 filled with details about steering wheel design and engine “knocking” problem: odd details to focus on when the physical structure of cities, the extent of local trade networks, and the location of housing in the suburbs versus the center was about to be changed. All this to say, providing economic advice to the AI optimists requires taking as given technical projections that are not universally accepted; nonetheless, it is advice *conditional* on at least a weak version of the optimist view that drives demand for economic policy advice in much of the AI world.

Now that we understand why many AI researchers are so optimistic, particularly about future projections, let us return to our books. We will note the technical accuracy of discussions of automation, prediction, and data, but we will observe a lack of discussion of the big question many in the AI crowd care most about: can AI cause transformative growth, particularly by speeding up the scientific production function, and what does that mean for wages, inequality, and organizations? The absence of rigorous economic analysis leaves policy discussions proceeding on the basis of intuition and analogy rather than systematic economic thinking, and risks economists being left out of policy discussions altogether.

3 The Second Machine Age

We begin with *The Second Machine Age* (Brynjolfsson and McAfee, 2014). Despite being published over a decade ago, it is the most general of all the books in this review on the society-wide economic impact of AI. Quite amazing for a book written just after the ImageNet breakthrough! Their core economic argument is straightforward: computers see exponential Moore's Law-type improvements in capabilities (the original Scaling relationship for compute?), rapidly growing corpora of digitized content (the original Scaling relationship for data?), and recombinant innovation using those inputs (the original Scaling relationship for methods?). Even at the time the book was written, it was clear that there was scope for artificial intelligence to solve pattern recognition problems; IBM's Watson program was winning Jeopardy contests and progress on automated interpretation of many medical scans was clear. The implications for "bounty and spread" the authors discuss - growing wealth due to increased recombinant invention brought about by digitization and machine learning, and the potential rise in inequality caused by the ability to extend a superstar's resources globally - are still the two key questions policymakers care about.

We, however, have an unfair advantage when evaluating Brynjolfsson and McAfee's book: they wrote in 2014, but we look back in 2025. One anecdote is particularly revealing. The book opens with a shock to the researchers, as they are given a demonstration in 2012 of Google's "Chauffeur" self-driving car project on California's Route 101.

When one of the Googlers hit the button that switched the car into fully automatic driving mode while we were headed down Highway 101, our curiosities - and self-preservation instincts - engaged. The 101 is not always a predictable or calm environment. It's nice and straight, but it's also crowded most of the time, and its traffic flows have little obvious rhyme or reason...The car performed flawlessly. In fact, it actually provided a boring ride. It didn't speed or slalom among the other cars; it drove exactly the way we're all taught to in driver's ed"

(Brynjolfsson and McAfee, 2014).

What makes this vignette interesting thirteen years later? Google’s Waymo division did in fact release self-driving taxis in many U.S. cities, where they are rapidly taking market share from traditional cabs and rideshare companies. But should you want to match this precise drive, from Stanford’s campus to Google’s offices in Mountain View, it is not possible in 2025. Waymo does not operate in this part of Silicon Valley, nor does it operate in 99% of the United States, to say nothing of the rest of the world. The barrier here is not technological; the fundamental AI aspects needed for self-driving have improved rapidly and dropped in cost by orders of magnitude. But these technological breakthroughs do nothing to change laws, get cities to approve trial demonstrations, get customers to try something new, or get corporate partners to develop vehicles that are a natural fit for autonomous vehicles. Indeed, in 2025, one must still watch a physical steering wheel on a Waymo vehicle turn as if by a ghost seated in the empty driver’s side, a very tangible manifestation of the idea that we do not legally and societally trust the car not to need a human to sit down and take over at some point.

And though on the one hand the spread of AI into the economy was slower than a reader might have expected, in general Brynjolfsson and McAfee were, with our perfect hindsight, too *pessimistic* about the *technical* improvements in AI. In their sixth chapter, Artificial and Human Intelligence in the Second Machine Age, they specifically look at the growth of pattern recognition in machine learning models, pointing to early breakthroughs. The chapter concludes with a subsection, “Billions of Innovators, Coming Soon”. Yet these billions of innovators are *not* AI scientists. Rather, “the main impediment to progress has been that, until quite recently, a sizable portion of the world’s people had no effective way to access the world’s stock of knowledge or to add to it.” Cell phones and the internet, by connecting potential innovators in the developing world to the frontier of knowledge, would allow more rapid recombinant growth. That is, the billions of innovators in this chapter are humans with access to digitized data, like the Kerala fishermen of Jensen (2007) whose

access to cell phones let them get their fish to the port with the highest price. This is a very different idea of AI for science from one where AI itself contributes to breakthroughs (and to learning processes to make future breakthroughs quicker).

The Second Machine Age predates the current wave of large language models and generative AI, terms not yet invented when the book was written. Their automation framework treats AI primarily as one of many components of advanced machinery called computers rather than a general-purpose cognitive technology. This is not meant as a critique; an economist who had foreseen the green shoots of deep learning in 2014 becoming a redwood-sized general-purpose technology would have had both incredible foresight and a fight with their editor about reasonableness. Nonetheless, as a popular introduction to how economists think about endogenous growth, about who captures the rents from technological change, about how digitization affects inequality, about how and whether exponential improvements in the productivity of a factor of production affect end markets and prices, there is very little to argue with here, even with the benefit of hindsight.

4 AI as a Prediction Machine

A reasonable question to ask after reading *The Second Machine Age* is this: if there are improvements in the capabilities of machine learning models a la Moore’s Law, why exactly do we care, economically? Both *Prediction Machines* (Agrawal et al., 2018) and *Power and Prediction* (Agrawal et al., 2022) begin with that question: what exactly is the *economic* implication of machine learning, which at its core is just a really good nonparametric estimator, for broad classes of problems, that minimizes loss rather than bias? Naively, you may think the answer is, not a heck of a lot. We are a century into statistics research on how to fit a function on data, and almost seventy-five years into the Savage (1954) world where these statistical models play a decision-theoretic role.

Prediction Machines, as its title suggests, offers a simple model: AI matters because it

drops the cost of prediction, an input into an economic production function. Drop the price of an input, we use more of it. Complements see an increase in their marginal product. The particular other inputs used in production rebalance. In principle, AI can be analyzed as an economic phenomenon like earlier technology shocks, such as the computer (“[c]omputers made arithmetic cheap”) or the internet (“a drop in the cost of...search”) (Agrawal et al., 2018).

Agrawal, Gans and Goldfarb, based at the University of Toronto, were in the right place at the right time, down the street from Hinton and coauthors when they were winning ImageNet (Krizhevsky et al., 2012). But the computer science concerns of the era were quite far from broad economic explanations. ImageNet was, literally, about classifying a picture as a rock or a bird or a dog. *Attention is All You Need* introduced large language models not as general AI, but as a tool to help with language translation (Vaswani et al., 2017). These are not *uninteresting*, but they are also not economically transformative.

Prediction is broader: the general attempt to understand something unknown based on available data. Ada Lovelace famously saw use cases for music generation from the Babbage “computer” - finding notes that sound good together is an arithmetic problem at heart⁸. Driving a taxicab is also a prediction problem. London black cab drivers spent years studying “The Knowledge” to know the quickest route from any location in the city to any other. Now, of course, that prediction can be done by mapping software on anyone’s phone. Who is harmed? Those whose competitive advantage was pure human prediction. Who benefits? Those whose complementary skills - flexibility in time, safe driving - are a complement to this prediction. Indeed, research including Kanazawa et al. (Forthcoming) has shown that AI predictions have flattened the distribution of performance on skills like “guess where customers will be” that previously were pure human prediction problems.

Economists who understand causal inference know that prediction alone may not be that

⁸“Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the [Babbage computer] might compose elaborate and scientific pieces of music of any degree of complexity or extent” (Lovelace, 1843)

useful in a decision problem. The authors give the lucid example of Kasparov’s team training an algorithm in the 1980s to mimic Grandmaster play in chess. The algorithm on a move-by-move basis performed brilliantly. When put to the test in a full game, it often sacrificed its queen right away! Why? Queen sacrifices by grandmasters often occur as part of brilliant strategic strokes (Kasparov and Greengard, 2017). Correlation is, of course, not causality.

This leaves an important role for humans: judgment. What is judgment? The authors compellingly define it decision-theoretically as well: knowing the set of possible decisions and their payoffs. AI can predict whether driving faster will increase the probability of a crash, and by how much; judgment tells me how to weigh those two factors. AI will predict whether a job candidate will perform well, how long until they attrit, or how they might perform in management roles if later promoted. Judgment tells you how those factors matter for your business. When causality is not correlation, judgment tells you the unstated assumptions required for the prediction to be useful. The separation of judgment, our maximand and our potential actions, from prediction, the likelihood of outcomes given those actions, is an incredibly valuable insight that has rightly been widely adopted as a framing by practitioners since *Prediction Machines* was released.

So AI predicts. We use our judgment to decide what we care about doing, and how we value various outcomes, then use its prediction to decide. Because prediction and judgment are complements, a fall in the cost of prediction increases the marginal return to good judgment, at least until AI predictions of optimal action, given strategic interaction and other limits on the value of pure correlation, trump human error. To prevent strategic gaming, search engines add human-coded rules on how websites are ranked, using machine prediction as an input. Human drivers use traffic predictions to decide when to leave or whether to take mass transit. When training chess engines, we understand that predicting what a grandmaster would do across a whole game versus on a one-move problem are different problems with different payoffs. And since human judgment is still “in the loop”, all of our old mechanism design problems about incentives, delegation, and task allocation remain

relevant (Athey et al., 2020).⁹ A reasonable reader will quibble, and the authors will grant the objection, that prediction and judgment do not seem as clean-cut a distinction as one may like. Consider an AI helping you decide to bring an umbrella or not. You know how annoying it is to carry, and how much you dislike being wet. The AI is better than you at predicting whether it will actually rain. But one can imagine an AI predicting how much you dislike being wet as well, based on prior behavior, shrinking the remaining penumbra of judgment. The reader must, perhaps, use their own judgment when thinking about tasks of options and value which AI will be unable to predict using the data it possesses.

Prediction Machines closes with a section, undercooked relative to the decision-theoretic model, but an attempt to address the big questions with rigor. The conclusions are perhaps not as optimistic as you might imagine.

1. Will there still be jobs? Yes.
2. Will this generate more inequality? Perhaps.
3. Will a few large companies control everything? It depends.
4. Will countries engage in race-to-the-bottom policy-making and forfeit our privacy and security to give their domestic companies a competitive advantage? Some will.
5. Will the world end? You still have plenty of time to derive value from this book (Agrawal et al., 2018).

Why these predictions? They are a simple application of the prediction/judgment dichotomy. On the first, there will still be jobs, but those solely reliant on human prediction can quickly disappear. Luckily, from radiologist to school bus driver, many tasks go beyond pure prediction or rely on non-digitizable data. Lawyers do not simply return a prediction of “yes, this contract breaks the law” or “no, you are fine”; they consult, they advise, they brainstorm, they help the client decide what needs to be predicted in the first place. Again, one can argue that these tasks are all prediction in the end. Likewise, on the other four questions, the line between prediction and judgment plays a critical role.

⁹Of course, humans in the loop may screw up predictions. Goh et al. (2024) shows a case of doctors making diagnoses alone, doctors doing this with an LLM, and the LLM making the diagnosis alone. The latter was the most accurate.

The other major issue with *Prediction Machines* is itself solved in the authors' follow-up, *Power and Prediction* (Agrawal et al., 2022). Recall the taxi driver example. Assume AI reduces the cost of prediction. You may think, aha, we can reduce the amount of training time for black cab drivers. But you are stuck in the short run, where the cost of a factor of production has changed but the production technology has not. In the longer run, we can use an entirely different technology that is now optimal given the shock to prediction cost. And indeed, this is literally what Uber does.

The company begins with Uber Black: take existing private car service drivers, predict when they have downtime, give them a smartphone to connect with customers, and let them earn money on the side. Interesting, but not transformative. What was transformative was understanding that if predicting the optimal route to drive required no special knowledge, then *anyone could be a driver*. A schoolteacher after class, wanting to earn a few bucks, could shuttle customers without fear of getting lost. The authors refer to the distinction between these two as “point solutions” versus “system solutions”.

The conceptual idea here is related to what Rebecca Henderson and Kim Clark call “architectural disruption” (Henderson and Clark, 1990). Some technology adoption requires easy changes: a pharmaceutical company marketing a different drug requires no change in its supplier relationships, workforce, information environment, or internal incentives. But a major shift in demand from pharmaceuticals to, say, health club memberships would be very difficult for the incumbent drug firm to handle. Sticky parts of firms and industries would need to change. Non-architectural innovations, therefore, tend to be adopted quickly; architectural ones can take decades to spread. We have often seen this pattern, such as the quick replacement of gas lamps with electric lamps after Edison's Pearl Street Station opened, but the slow reconfiguration of industry around a location-independent power source. As David (1990) noted, it took four decades for half of U.S. industry to electrify, with the biggest benefits coming at firms that completely shifted their production processes. Ford moved his factories to the suburbs, turned them sideways, and implemented an assembly

line, realizing that without the need for belts and turbines to generate power, the whole structure of factory production could be remade.

Solow famously saw the computer everywhere but in the productivity statistics in the late 1980s (Solow, 1987). The distinction between point and system solutions was an important reason why. With AI, a tool for reducing the cost of prediction, this same issue arises. How does the modal company first implement AI? By looking at its finite tasks and trying to reduce costs. A point solution! Indeed, *economics* research on AI adoption at firms tends to investigate precisely the same type of “improve productivity at existing tasks” problems (Brynjolfsson et al., 2025; Eloundou et al., 2023). If AI drops the cost of prediction, we benefit not just from changing our cost structure to do what we did, but also from changing the tasks we do. The long run! This is difficult. Anyone can look at the finite tasks one does, where implementing AI requires no major changes. But thinking through the infinite tasks one doesn’t currently do but should do now? If AI reduces the cost of weather prediction, a farmer can better choose when to plant and harvest. But should the industrial organization of farming look different? Should we massively rebalance labor versus capital? Does optimal farm size change? Can we grow wholly different crops?

Because a drop in prediction cost is architecturally disruptive, AI’s largest benefits will take time to spread. Incumbent institutions with the resources to implement AI are architecturally a bad fit for many of these implementations; consider the speed at which a university would, for instance, redirect even 10% of its spending toward compute if it turned out the marginal product of compute for learning and research was much higher than for, say, lab equipment or postdoc staff. If a drop in the cost of prediction meant university degrees should be one year longer, or one shorter, or the tenure system did a poor job maximizing research under AI, would changes on these margins be implemented even a decade on? Two decades? A well-known problem with reacting to architectural changes is that since they require up-front costs in response to an uncertain future - remember, the variance in the potential of AI ten years out is very high - we may see particularly poorly fit incumbents if

AI capabilities grow in line with the optimistic projections.

5 Data and the Macroeconomy

The Data Economy: Tools and Applications (Baley and Veldkamp, 2025) considers how data affects economic outcomes. Note the term: data, not AI. For Baley and Veldkamp, data is a *nonrivalrous input to prediction*, potentially with increasing returns to scale, and often created as a byproduct of other economic activity. It is therefore an asset controlled by a firm. Many chapters implicitly assume that predictive accuracy for a given dataset improves over time, but little of the book requires that.

This book is by far the most mathematically rigorous of those covered in our joint review. If the reader wants to learn how to apply the Kalman-Bucy Filter to update beliefs about a state, with practice problems at chapter's end, she can. And likewise for understanding how to acquire data via experimentation (a bandit problem), the value of that information (in terms of entropy), or how to extract signal about, say, inflation from noisy measurements. The formalism is well appreciated. A perceptive audience may note that nearly all books covered are written by business school researchers for practitioners. Economists of course have many models of how to use data in decision-theoretic and game-theoretic problems, including dynamic ones. *The Data Economy* is the most comprehensive and rigorous collection of insights in these models, providing enough detail for PhD students to build on them. It is also remarkably up to date - the book is filled with discussions of papers released only months before the book was published.

The mass of formalism proves its worth in later chapters. You are convinced that data is an asset, useful to firms, collected deliberately and as a byproduct of economic activity, and used to drive production efficiencies and set prices. All this has general equilibrium effects. So does a more data-intensive economy have different business cycles? Yes, particularly when data is generated as a byproduct of economic activity, hence its decline during

recessions creates a negative feedback loop. Does a data-heavy economy change the nature of competition? Sure, such as when heterogeneity in data access creates winners' curses in asset purchases, sharpening misallocation via an auction-theoretic strategic effect. Should we, and can we, measure the value of data as an asset held by firms? The more important proprietary data becomes, the more important this measurement is; Baley and Veldkamp offer a dozen microfounded ways a financial analyst might perform this measurement.

Shifting focus from AI algorithms back to data's fundamental theoretical properties is a welcome rejoinder. Consider Google's self-driving car, which we met in *The Second Machine Age*. Google has spent billions of dollars collecting data about city streets. What data should it collect? When is it valuable? Can it be sold? How should an investor assess Google's data hoard relative to, say, Tesla or Uber? When has it collected "enough" data? Without data, there is nothing to train an AI algorithm on, and no ground truth on which to judge its predictions. Without understanding the role of data in market competition, markets will not allocate the right amount of resources to data collection in the first place.

Again, the question arises: what does this have to do with AI? And one can't blame the authors: their book is called *The Data Economy*, not *The AI Economy*. But a fundamental idea in AI is that so-called foundation models can outpredict more specialized ones - recall Sutton's Bitter Lesson. Is your unique data, analyzed reasonably well, or my public data, analyzed with frontier prediction tools, more valuable? Regardless of where this balance falls, general predictive accuracy improves rapidly.

Taking both the theoretical models in Baley and Veldkamp (2025) and the perspective of the optimistic portion of AI researchers seriously, let us try to make progress here on their combination. Assume foundation models improve rapidly, including their ability to self-improve. Assume AI can, in the near future, gather any public data, including creating data sources from unstructured content. Assume AI can run models that optimally balance bias and variance for a given task, and that it can apply causal methods correctly. Assume that these models are available at low cost to any user who wants them. Besides judgment

being valuable, private data that can be kept private rises in marginal value as long as there is no highly correlated public analogue. What does that mean? Consider trying to hire. A large company with data on how each applicant panned out may have a competitive advantage, but unless the company is idiosyncratic, predictive accuracy from an AI using public data may look similar. If you take AI's ability to *gather* data seriously, the datasets that provide a competitive advantage look smaller, and problems like algorithmic pricing or asset pricing with fewer noise traders become more important. In precisely those settings, economic models clarify.

6 Practical Implementation

Two popular books - *The Skill Code* by a sociologist (Beane, 2024), and *Co-Intelligence* by a management scholar (Mollick, 2024) - help shift the focus from prediction to data to even more microeconomic phenomena. Both authors begin with a simple observation: AI is used by organizations made up of people. It is not just an autonomous tool, and humans in the loop may be necessary for the creation of economic value, or neutral, or even degrade the performance compared to AI alone on some problems. How then should we use AI, and where are the biggest pitfalls?

Beane's *The Skill Code* reveals a critical blind spot in AI economics through ethnographic research on how AI - and indeed, any new technology - disrupts skill formation processes. Workers often gain skills dynamically through "osmosis" as they observe and assist senior workers. A surgeon does not become a surgeon simply by reading books, but by assisting experienced surgeons in the operating room. Though the formal apprentice model is less common, many professions still operate on it implicitly, with new workers paid less as they learn the ropes while working closely with senior colleagues.

The problem with this informality is that the experienced workers and their firms do not capture the full rents of this training. A market is missing. When learning is an unavoidable

byproduct of day-to-day production, no problem. But consider the introduction of a surgical robot with four mechanical arms. The senior surgeon can control all four arms, obviating the need for flesh-and-bone models attached to the surgical resident. This robot prevented mistakes and reduced operation time. Who could object? In a 2019 paper, Beane finds these surgical robots improve efficiency by 40% (Beane, 2019). But the byproduct is that they block skill acquisition for medical trainees, whose participation now resembles watching a movie rather than participating in the procedure.

AI introduces a similar problem. AI's effect on wages depends on whether human skills are complements or substitutes. But human skills are dynamically developed. Wages grow rapidly early in the career life-cycle. Understanding whether AI might interrupt skill development even as it improves short-run productivity or wages is critical.

There is of course an analogue in economic history with the famous de-skilling debate (for example, Atack et al. (2024)). In both the 18th-century Industrial Revolution and the late 19th-century mechanization of industry, commentators worried that machines limited skill acquisition. Even if productivity enhancing in the short run, lost skills might harm long-run productivity. Both in this historical case and with AI, the fundamental problem is one of externalities: the trainer is not paid for the learning of the trainee and hence may “over-mechanize” (see Acemoglu and Restrepo (2020)).

The impact of this skill development change depends, of course, on what the human is to do later, and in how institutions adopt. If AI pioneer Geoff Hinton's mid-2010s claim that we would no longer need to train radiologists had held true, where AI could replicate their full job, then the lost skill acquisition would not be worrying, at least societally (Thompson, 2017).¹⁰ And perhaps we'll see a reinvigoration of formal on-the-job training and mentorship programs, where knowledge is transferred not as a byproduct of work, but via lower wages for the younger workers who get the skill acquisition as a result.

Beane's studies are largely ethnographic field research - useful in what we learn but not

¹⁰Although, as we can see with Hinton's error - radiologists very much still exist - projecting where AI will be autonomously productive is difficult.

easily scalable to society-wide results. How would a policymaker concerned about “AI de-skilling” measure it? We do not directly observe hours of informal training for young workers. While one can advise an HR manager how to ensure younger workers remain *challenged*, with sufficiently *complex* work, and with a personal *connection* to their mentor (Beane’s “three C’s”), public policy has a role if we take the macroeconomic implications seriously.

Inexperienced workers have one saving grace. A firm will hire someone if the marginal product covers the wage. And it will set up structures where young workers spend time with experienced workers if that interaction’s marginal product exceeds alternative arrangements. Surveys of AI adoption like Bick et al. (2025) find younger workers using generative AI at nearly twice the rate of older ones. Perhaps, by helping senior colleagues use AI effectively, younger workers will get the “connection” needed to improve their skills.

Mollick, an innovation researcher at the University of Pennsylvania, worries less about the future of workers than about how they interact with AI today, in *Co-Intelligence* (Mollick, 2024). AI may be adopted slowly because of architectural barriers, but innovations also diffuse slowly because individuals must learn how to use them effectively. Almost 70 years ago, Coleman et al. (1957) showed the slow diffusion of a new antibiotic, tetracycline, in Chicago was driven by the fact that doctors did not know how and when to use it effectively until they had this explained by trusted doctors in their social network.

This slow diffusion is made worse for AI by what Mollick and co-authors call “the jagged frontier” (Dell’Acqua et al., 2023). Identifying what tasks humans, machines, and AI find relatively harder is nonobvious. Seemingly difficult tasks are within AI’s grasp: these models can write proofs for novel IMO-level mathematics questions, but cannot add two ten-digit numbers without computer code. If a worker does not understand where a given task lies on the jagged frontier, they may try to use AI to help with tasks where it degrades performance, or avoid using AI on tasks where it can help; Dell’Acqua et al. (2023) find precisely these mistakes in their study run with Boston Consulting Group.

In the short run, then, how can a worker learn-by-doing most effectively? You are

not normally told to experiment with a new surgical instrument or software; instead, you are trained on its best use. At least for now, this does not work for AI due to jagged frontier effects. So how can workers develop skills with large language models that are most likely to be complements going forward, while not degrading output quality today? Mollick recommends four rules - try using AI as a partner for common tasks, do not fully automate critical tasks, interact with AI as in a conversation with a human that occasionally fabricates or blusters, and do not use it in a way that deskills you.

As practical advice, all well and good, just as Beane’s *Skill Code* is fantastic practical advice for an HR head or manager worried about workforce skills. While both authors speculate about the medium-run implications of how AI is used, they do not formally advise on these grounds. Will AI in fact require a human in the loop for most tasks in ten years, or even benefit from having one? If so, which ones? How will we know? Is de-skilling due to AI relevant at the level of the macroeconomy? If so, for which tasks? How will we know? What are the metrics policymakers should be paying attention to as an early warning?

7 The View from California

We opened with a quote from the economics-trained AI researcher Leo Aschenbrenner. Now that we have seen where social scientists have focused their attention with regard to AI, let us return to Aschenbrenner’s “Situational Awareness” (Aschenbrenner, 2024)¹¹ “You can see the future first in San Francisco,” he writes, which is surely true. Researchers at the AI labs there “are very smart people - the smartest people I have ever met - and they are the ones building this technology. Perhaps they will be an odd footnote in history, or perhaps they will go down in history like Szilard and Oppenheimer and Teller. If they are seeing the future even close to correctly, we are in for a wild ride. Let me tell you what we see.”

¹¹It is worth noting that Aschenbrenner’s view of AI capabilities, as wild as it may seem, is less out-there than texts like AI 2027 (Kokotajlo et al., 2025). In that self-published scenario, a group of very prominent AI safety researchers, including a former high-level OpenAI employee, project recursive self-improvement in the ability to improve AI that becomes meaningful between 2027 and 2029, at which point both capabilities and safety concerns become extreme.

And what they see is a radically more consequential technology than even the most optimistic books described so far have considered. Aschenbrenner does add a caveat that these may be nuts who will be a footnote in history, of course. But if they are correct, consider what is being predicted:

We will see “\$1T/year of total annual AI investment” by 2027, representing approximately 3% of GDP. This investment scale would require fundamental changes to American economic infrastructure: “American electricity production will have grown tens of percent; from the shale fields of Pennsylvania to the solar farms of Nevada, hundreds of millions of GPUs will hum.” Aschenbrenner envisions “every six months another zero added to the boardroom plans,” with AI labs routinely discussing hundred-billion-dollar capital expenditures. This scale implies “our first many-hundred-billion-dollar corporate bond sale” as traditional financing mechanisms strain to accommodate AI development needs.

He predicts “hundreds of millions” of human-level AI researchers investigating at 1,000 times human speed with “superhuman ML intuitions,” effectively “compressing a decade of algorithmic progress into less than one year.” He expects this will be possible before the end of the decade. This “intelligence explosion” would create economic dynamics without historical precedent (see Aghion et al. (2019) and Jones (2024) for formal models of this). If AI systems can recursively improve themselves while operating at superhuman speed, traditional growth models based on human capital accumulation become obsolete. The economic implications extend far beyond productivity improvements to fundamental questions about the nature of economic value creation and distribution.

What of market structure? Aschenbrenner thinks AI intelligence will rapidly become so critical to national security that governments will simply not *allow* private companies to fully determine the growth of this technology. That is, access to AI, its pricing, and the distribution of earnings from its development will all be determined by legislation as well as pure technology.

He understands these claims are wild. But “[a]t this point, you may think that I and

all the other SF-folk are totally crazy. But consider, just for a moment: what if they're right? These are the people who invented and built this technology; they think AGI will be developed this decade; and, though there's a fairly wide spectrum, many of them take very seriously the possibility that the road to superintelligence will play out as I've described..." (Aschenbrenner, 2024).

The reason to take Aschenbrenner seriously is not because one must agree with his predictions. One might not even think them likely. But it is simply a fact that the view of the future expressed in "Situational Awareness" is closer to the modal view of AI researchers and the folks running the most prominent AI labs than less bold analyses of AI, even those that treat AI as a potential general purpose technology with substantial economic importance. If economists are to play a public role in the debate of AI, it is essential to at least understand the economic model in the heads of many of the people we are trying to communicate with.

8 The Big Open Questions

It should be clear to this point that conceptually there is a great deal to like about all of these books. We learn why machine learning is an economic phenomenon; understand that AI reduces the cost of prediction, increasing the value of prediction's complements; see that the highest value use cases of prediction require overcoming architectural barriers in incumbent organizations; derive economic models of data collection and aggregation, the critical input into prediction tasks; and understand how to use AI in practice in the workforce, both for ourselves and for our junior colleagues whose skills might otherwise degrade.

In a sense, the recent wave of books on the economics of AI follows a similar script to the 1990s and early 2000s research on digitization, such as *Information Rules* (Shapiro and Varian, 1998) and *The New Division of Labor* (Levy and Murnane, 2004). It is obvious now that the digitization of industries and their movement to the internet rapidly transformed the nature of many of them (retail, advertising, media, etc.) and their relationship with

labor. Those books correctly laid down the key economic issues of digitization: who will profit, who will lose, what industries will change. But AI may be different, in terms of the magnitude of its effects. What the books in this review don't do is answer Aschenbrenner's question: what if they're right? What if the "totally crazy" folks in San Francisco - the ones developing this technology, to be clear - are correct about the technical milestones AI will pass? Are we as economists providing the right guidance taking that conditional seriously?¹²

Let us conclude with some open questions, none of which are directly answered by the books here, but all of which take seriously the totally crazy folks out West. Again, this is not a critique of the books reviewed here: they have a different audience and a different goal. Nonetheless, here are questions which a reader would still not be able to answer after reading the full collection:

How should monetary policy respond to simultaneous deflationary pressure from productivity gains and potential unemployment from labor demand shifts? What are the implications for interest rate policy when technological change accelerates dramatically? How is public debt affected? If AI adoption reduces employment while requiring large public investments in education, social protection, and infrastructure, how will governments finance these expenditures? What are the optimal tax policies for an economy where capital captures an increasing share of income? It is no surprise that OpenAI's Sam Altman funded the largest economic studies of basic income (Vivalt et al. (2025)).

International economics and AI remain underdeveloped, despite the clear implications for trade, capital flows, and global economic relations. AI development exhibits strong network effects and scale economies that could create winner-take-all dynamics among countries. How should trade policies respond when AI capability determines comparative advantage? What are the implications for international capital flows when AI investment becomes central to national competitiveness? Is "sovereign AI" necessary? How should it be taxed, considering

¹²An anonymous referee for this review essay made the astute observation that if AI actually becomes as capable as the optimists believe, there's really no point in human economists writing the analysis today, since AI in 2029 will do a better job of analyzing its *own* economic impact. Point taken!

the inequality discussion? How can we, as with climate change, coordinate internationally on AI safety concerns? Game theorists have a role to play here.

If AI development requires trillion-dollar investments concentrated among a few companies, how will this affect financial system stability? How should banking regulations adapt as AI systems conduct increasing shares of financial transactions? What are the systemic risks when AI companies become too big to fail?

What should we measure? What early warning signs indicate negative effects or growth takeoff? National accounting systems designed for industrial production will not provide guidance as AI becomes an important share of economic activity.

How should science and innovation be structured? Are patents more or less important? How does information flow across firms? To what extent should we allow trade secrets in an AI-driven intelligence explosion? What should agencies like the NIH or NSF be doing differently? What should universities be teaching differently?

It is not too early to ask these questions. Policy importance does not move at the speed of academic research. In the three to five year gap between when regulations are proposed and when they are implemented, the world may look very different - indeed, the first wave of AI regulation at the start of the decade focused almost entirely on issues that already appear second-order (Bryan and Teodoridis, 2024). Social science needs to be ahead of policy if the latter is to have a theoretically rigorous and empirically sensible basis. We have a role not just in explaining AI ex post, but in shaping its development ex ante. There is a risk here that the optimistic scenario on AI improvement plays out, which means a potential pace of change so quick to require ideas at the nexus of “odd footnote in history” and epoch-defining. This is uncomfortable for the stoic mentality of an academic. But when it comes to AI, it is a scenario within the realm of possibility.

References

- 60 Minutes**, “Google DeepMind CEO Demis Hassabis on Artificial Intelligence,” <https://www.cbsnews.com/news/google-artificial-intelligence-demis-hassabis-60-minutes/> 2025. Interview aired April 20, 2025, CBS News.
- Acemoglu, Daron and Pascual Restrepo**, “The Race Between Machine and Man: Implications of Technology for Growth, Factor Shares and Employment,” NBER Working Paper 22252, National Bureau of Economic Research 2016.
- and —, “Robots and Jobs: Evidence from US Labor Markets,” *Journal of Political Economy*, 2020, 128 (6), 2188–2244.
- Aghion, Philippe, Benjamin F. Jones, and Charles I. Jones**, “Artificial Intelligence and Economic Growth,” in Ajay Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, Chicago: University of Chicago Press, 2019, pp. 237–282.
- Agrawal, Ajay, Joshua Gans, and Avi Goldfarb**, *Prediction Machines: The Simple Economics of Artificial Intelligence*, Boston: Harvard Business Review Press, 2018.
- , —, and —, *Power and Prediction: The Disruptive Economics of Artificial Intelligence*, Boston: Harvard Business Review Press, 2022.
- Altman, Sam**, “Three Observations,” <https://blog.samaltman.com/three-observations> 2025.
- Ananthaswamy, Anil**, *Why Machines Learn: The Elegant Math Behind Modern AI*, New York: Dutton, 2024.
- Aschenbrenner, Leopold**, “Situational Awareness: The Decade Ahead,” <https://situational-awareness.ai/> 2024. Accessed 2025.

- Atack, Jeremy, Robert A. Margo, and Paul W. Rhode**, “De-skilling: Evidence from Late Nineteenth Century American Manufacturing,” *Explorations in Economic History*, 2024, *91*, 101554.
- Athey, Susan C., Kevin A. Bryan, and Joshua S. Gans**, “The Allocation of Decision Authority to Human and Artificial Intelligence,” *AEA Papers and Proceedings*, 2020, *110*, 80–84.
- Baley, Isaac and Laura L. Veldkamp**, *The Data Economy: Tools and Applications*, Princeton: Princeton University Press, 2025.
- Beane, Matthew**, “Shadow Learning: Building Robotic Surgical Skill When Approved Means Fail,” *Administrative Science Quarterly*, 2019, *64* (1), 87–123.
- , *The Skill Code: How to Save Human Ability in an Age of Intelligent Machines*, New York: Harper Business, 2024.
- Bick, Alexander, Adam Blandin, and David J. Deming**, “The Rapid Adoption of Generative AI,” NBER Working Paper 32966, National Bureau of Economic Research 2025.
- Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell et al.**, “Language Models Are Few-Shot Learners,” in “Advances in Neural Information Processing Systems,” Vol. 33 2020, pp. 1877–1901.
- Bryan, Kevin A. and Florenta Teodoridis**, “Balancing Market Innovation Incentives and Regulation in AI: Challenges and Opportunities,” 2024. September 24.
- Brynjolfsson, Erik and Andrew McAfee**, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, New York: W. W. Norton & Company, 2014.

– , **Danielle Li**, and **Lindsey Raymond**, “Generative AI at Work,” *Quarterly Journal of Economics*, 2025, *140* (2), 889–942.

Bullock, Justin B., **Yu-Che Chen**, **Johannes Himmelreich**, **Valerie M. Hudson**, **Anton Korinek**, **Matthew M. Young**, and **Baobao Zhang**, *The Oxford Handbook of AI Governance*, Oxford University Press, 2024.

Coleman, James, **Elihu Katz**, and **Herbert Menzel**, “The Diffusion of an Innovation Among Physicians,” *Sociometry*, 1957, *20* (4), 253–270.

David, Paul A., “The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox,” *American Economic Review*, 1990, *80* (2), 355–361.

Dell, Melissa, “Deep Learning for Economists,” *Journal of Economic Literature*, 2025, *63* (1), 5–58.

Dell’Acqua, Fabrizio, **Edward McFowland**, **Ethan R. Mollick**, **Hila Lifshitz-Assaf**, **Katherine C. Kellogg**, **Saran Rajendran**, **Lisa Kraymer**, **François Cadelon**, and **Karim R. Lakhani**, “Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality,” Technical Report 24-013, Harvard Business School Working Paper 2023.

Eloundou, Tyna, **Sam Manning**, **Pamela Mishkin**, and **Daniel Rock**, “GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models,” Working Paper arXiv:2303.10130, arXiv 2023.

Goh, Ethan, **Robert Gallo**, **Jason Hom**, **Eric Strong**, **Yingjie Weng**, **Hannah Kerman**, **Joséphine A. Cool**, **Zahir Kanjee**, **Andrew S. Parsons**, **Neera Ahuja**, **Eric Horvitz**, **Daniel Yang**, **Arnold Milstein**, **Andrew P. J. Olson**, **Adam Rodman**, and **Jonathan H. Chen**, “Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial,” *JAMA Network Open*, 2024, *7* (10), e2440969.

- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville**, *Deep Learning*, MIT Press, 2016.
- Henderson, Rebecca M. and Kim B. Clark**, “Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms,” *Administrative Science Quarterly*, 1990, 35 (1), 9–30.
- Hoffmann, Jordan, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark et al.**, “Training Compute-Optimal Large Language Models,” Preprint arXiv:2203.15556, arXiv 2022.
- Jensen, Robert**, “The Digital Divide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector,” *Quarterly Journal of Economics*, 2007, 122 (3), 879–924.
- Jones, Charles I.**, “The AI Dilemma: Growth versus Existential Risk,” *AER: Insights*, 2024, 6 (4), 575–590.
- Kanazawa, Kyogo, Daiji Kawaguchi, Hitoshi Shigeoka, and Yasutora Watanabe**, “AI, Skill, and Productivity: The Case of Taxi Drivers,” *Management Science*, Forthcoming.
- Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei**, “Scaling Laws for Neural Language Models,” Preprint arXiv:2001.08361, arXiv 2020.
- Kasparov, Garry and Mig Greengard**, *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*, New York: PublicAffairs, 2017.

- Kokotajlo, Daniel, Scott Alexander, Thomas Larsen, Eli Lifland, and Romeo Dean**, “AI 2027: A Research-Backed Scenario Forecast,” <https://ai-2027.com/> 2025. Accessed August 14, 2025.
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton**, “ImageNet Classification with Deep Convolutional Neural Networks,” in “Advances in Neural Information Processing Systems,” Vol. 25 2012.
- LeCun, Yann**, “A Path Towards Autonomous Machine Intelligence,” *OpenReview*, 2022.
- , **Yoshua Bengio, and Geoffrey Hinton**, “Deep Learning,” *Nature*, 2015, 521 (7553), 436–444.
- Levy, Frank and Richard J. Murnane**, *The New Division of Labor: How Computers Are Creating the Next Job Market*, Princeton: Princeton University Press, 2004.
- Lovelace, Ada**, “Notes by the Translator,” *Scientific Memoirs, Selected from the Transactions of Foreign Academies of Science and Learned Societies, and from Foreign Journals*, 1843, 3, 666–731. Appended to L.F. Menabrea, ”Sketch of the Analytical Engine invented by Charles Babbage Esq.”.
- Lucas, Robert E.**, “Econometric Policy Evaluation: A Critique,” *Carnegie-Rochester Conference Series on Public Policy*, 1976, 1, 19–46.
- , “On the Mechanics of Economic Development,” *Journal of Monetary Economics*, 1988, 22 (1), 3–42.
- Mollick, Ethan**, *Co-Intelligence: Living and Working with AI*, New York: Portfolio, 2024.
- Narayanan, Arvind and Sayash Kapoor**, *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*, Princeton University Press, 2024.

Nordhaus, William D., “Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth,” NBER Working Paper 21547, National Bureau of Economic Research 2015.

OpenAI, “Introducing OpenAI,” 2015. December.

– , “GPT-4 Technical Report,” Technical Report, OpenAI 2023.

– , “GPT-5 System Card,” <https://cdn.openai.com/pdf/8124a3ce-ab78-4f06-96eb-49ea29ffb52f/gpt5-system-card-aug7.pdf> 2025. Accessed August 7, 2025.

Savage, Leonard J., *The Foundations of Statistics*, New York: John Wiley & Sons, 1954.

Shapiro, Carl and Hal R. Varian, *Information Rules: A Strategic Guide to the Network Economy*, Boston: Harvard Business Press, 1998.

Silver, David, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel et al., “A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go through Self-Play,” *Science*, 2018, *362* (6419), 1140–1144.

Snell, Charlie, Jaehoon Lee, Kelvin Xu, and Aviral Kumar, “Scaling LLM Test-Time Compute Optimally Can Be More Effective than Scaling Model Parameters,” Preprint arXiv:2408.03314, arXiv 2024.

Solow, Robert M., “We’d Better Watch Out,” *New York Times Book Review*, 1987, p. 36. July 12.

Sutton, Richard S., “The Bitter Lesson,” 2019. March. Available at: <http://www.incompleteideas.net/IncIdeas/BitterLesson.html>.

Thompson, Clive, “A.I. Versus M.D.,” *The New Yorker*, 2017. March 27. Available at: <https://www.newyorker.com/magazine/2017/04/03/ai-versus-md>, Accessed August 13, 2025.

VandeHei, Jim and Mike Allen, “Behind the Curtain: A White-Collar Bloodbath,” *Axios*, 2025. May 28. Available at: <https://www.axios.com/2025/05/28/ai-jobs-white-collar-unemployment-anthropic>.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, “Attention Is All You Need,” in “Advances in Neural Information Processing Systems,” Vol. 30 2017.

Vivalt, Eva, Elizabeth Rhodes, Alexander W. Bartik, David E. Broockman, Patrick Krause, and Sarah Miller, “The Employment Effects of a Guaranteed Income: Experimental Evidence from Two U.S. States,” NBER Working Paper 32719, National Bureau of Economic Research 2025.