



INTERNATIONAL

# Productivity Monitor



Centre for the Study  
of Living Standards



# International Productivity Monitor

---

The International Productivity Monitor (IPM) is an open-access, peer-reviewed economics journal published by the Centre for the Study of Living Standards in partnership with The Productivity Institute.

The IPM publishes high-quality articles that build the evidence base and serve as a forum for global knowledge exchange on productivity research. Articles are contributed by leading academics and practitioners, combining rigorous analysis with practical insight, and are intended for a broad audience, including researchers, policy analysts, policymakers, the media, and the general public.

# The Potential for Sustained Productivity Impetus from GenAI

Martin Neil Baily, David M. Byrne, Aidan T. Kane and Paul E. Soto\*

*IPM Review Article*

## Abstract

With the advent of generative artificial intelligence (GenAI), the scope of AI has increased dramatically, but its effect on labour productivity remains uncertain. Some innovations raise labour productivity growth as adoption spreads but the effect fades when the market is saturated. In contrast, two types of innovation — general purpose technologies (GPTs) and inventions in the method of invention (IMIs) — have long-lasting effects on productivity growth. GPTs (1) are widely adopted, (2) spur knock-on innovations, and (3) show continual improvement, refreshing the innovation cycle. IMIs increase the efficiency of the research and development process via improvements to observation, analysis, communication, and organization. We conclude there is suggestive evidence that GenAI is both a GPT and an IMI, a sign that its adoption will lead to higher labour productivity growth in the future.

---

\*Martin Neil Baily is Senior Fellow Emeritus and Aidan T Kane is a Research Analyst at The Brookings Institution. David M. Byrne and Paul E. Soto are both Principal Economists at the Federal Reserve Board of Governors. David Byrne is the corresponding author (email: david.m.byrne@frb.gov). Authors are listed in alphabetical order, not in order of relative contribution. The views expressed here are not represented to be the views of the staff or trustees of The Brookings Institution nor of the Federal Reserve. The authors are grateful to Michael Chui, Leland Crane, Avi Goldfarb, Bob Gordon, Shane Greenstein, Anton Korinek, James Manyika, Sid Srinivasan, Scott Stern, and Bill Whyman for helpful conversations.

## 1. Introduction

In late 2022, OpenAI grabbed the world’s attention with ChatGPT, a generative artificial intelligence (GenAI) program that uses a computer model of human discourse to respond to natural-language questions (Figures 1-1 and 1-2). The scope of AI expanded dramatically with the advent of GenAI, including to tasks previously seen as quintessentially human, such as competition-level mathematics (Figure 2). Indeed, more and more challenging benchmark tests have been needed to assess the technology’s progress, as GenAI has matched human performance on one task after another. In another encouraging sign, field test evidence of productivity improvements from GenAI in practical applications has emerged, notably for writing, computer programming, and responding to call center inquiries. We supplement the quantitative evidence with a qualitative assessment of the properties of GenAI, with a view to helping to predict its impact.<sup>1</sup>

We focus on two particularly impactful types of innovation. Unlike discrete innovations that temporarily raise productivity growth as adoption spreads, “general purpose technologies” (GPTs) and “inventions in the method of invention” (IMIs) have longer-lasting growth effects. GPTs are

widely adopted, spur knock-on innovations — new products, process improvements, and business reorganization — and refresh this adoption cycle through ongoing improvement in the core technology (Lipsey *et al.*, 2005). IMIs yield a sustained increase in productivity growth by lowering the cost of research and development (Whitehead, 1925). We conclude there is suggestive evidence that GenAI is both a GPT and an IMI, a sign that its adoption will lead to higher labour productivity growth in the future.<sup>2</sup>

## 2. GenAI as a General Purpose Technology

The future productivity impact of GenAI will depend on whether (1) it is widely adopted, (2) it spurs related innovations, and (3) the technology continues to improve. These are the three distinguishing features of GPTs.

### 2.1 Diffusion

Analysis of job descriptions suggests that AI can be used for a broad range of workplace tasks, indicating high potential for diffusion and rising use of AI (Eloundou *et al.*, 2024).<sup>3</sup> We review evidence that this is taking place. Although few of these indicators are restricted to GenAI, the timing

---

1 For a more extensive discussion of the issues in this article, see Baily and Kane (2025a,b); Kane and Baily (2025a,b).

2 There is a substantial literature on the question of whether machine learning (ML), which preceded GenAI, is a GPT or an IMI, but little such work on GenAI. Eloundou *et al.* (2024), an exception, consider the prospects for GenAI to be a GPT based on the prevalence of tasks that appear likely to benefit from GenAI. On ML as a GPT, see Cockburn, Henderson, and Stern, 2019; Trajtenberg, 2018; Bresnahan, 2019; Goldfarb, Taska, and Teodoridis, 2023; Bresnahan, 2024. Cockburn, Henderson, and Stern (2019) consider if ML is an IMI.

3 Yin, Vu, and Persico (2026) caution that measures of exposure to AI are highly sensitive to the assumptions used.

## Figure 1-1: Web Searches for Artificial Intelligence

Indexed to 100 in month of greatest search



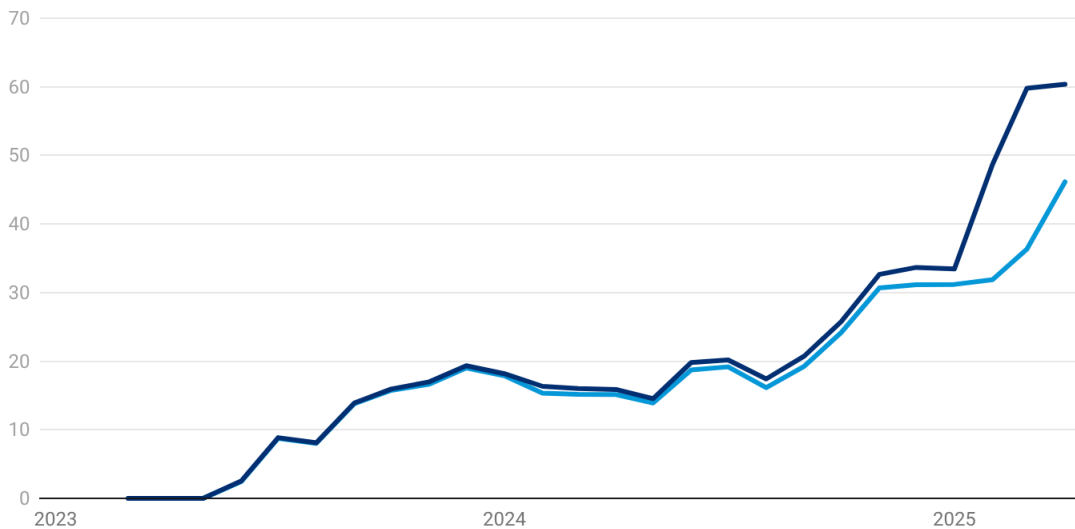
Note: Web searches include related terms in Google's "AI" topic.

Source: Google Trends. • Created with Datawrapper

## Figure 1-2: Generative AI Mobile App Downloads

Millions of downloads

— All Apps — ChatGPT

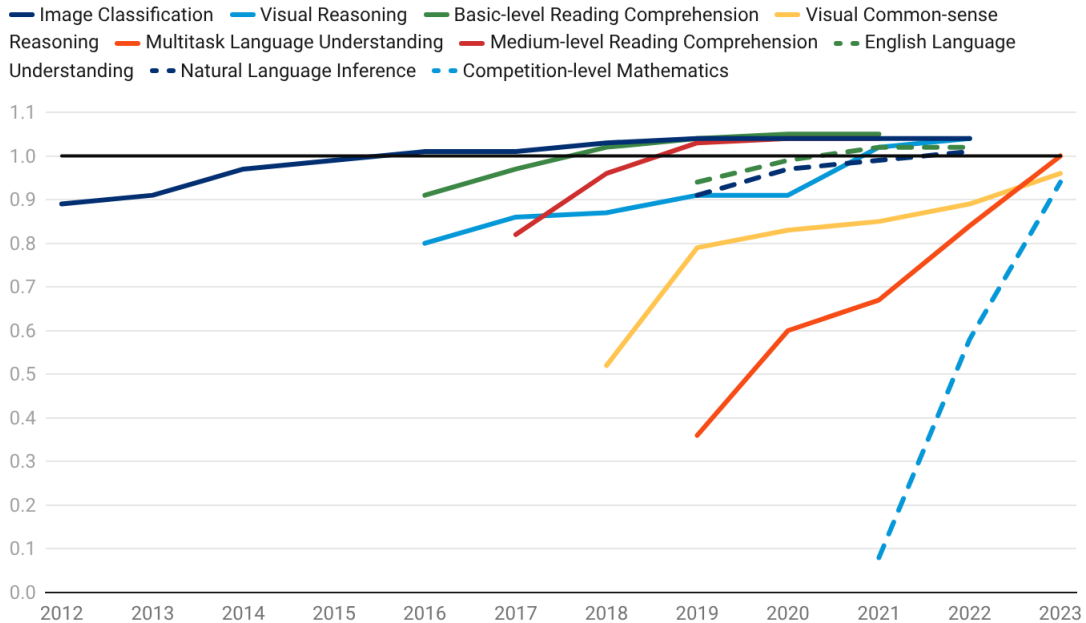


Note: Apps include ChatGPT, Claude, DeepSeek, and Perplexity. Includes Android and iOS versions. Android download information not available for China. Does not account for access via application program interface.

Source: Appfigures. • Created with Datawrapper

## Figure 2: Artificial Intelligence Benchmark Performance

Performance relative to human baseline on various tasks, ratio



Note: The human baseline concept used varies by task. For more challenging tasks, the baseline reflects expert-level performance.

Source: Reproduced with permission from the 2024 AI Index Report, Stanford Institute for Human-centered Artificial Intelligence. • Created with Datawrapper

of their increases is consistent with GenAI driving adoption upward. The level of AI use varies across sources and depends on the meaning one uses for adoption.

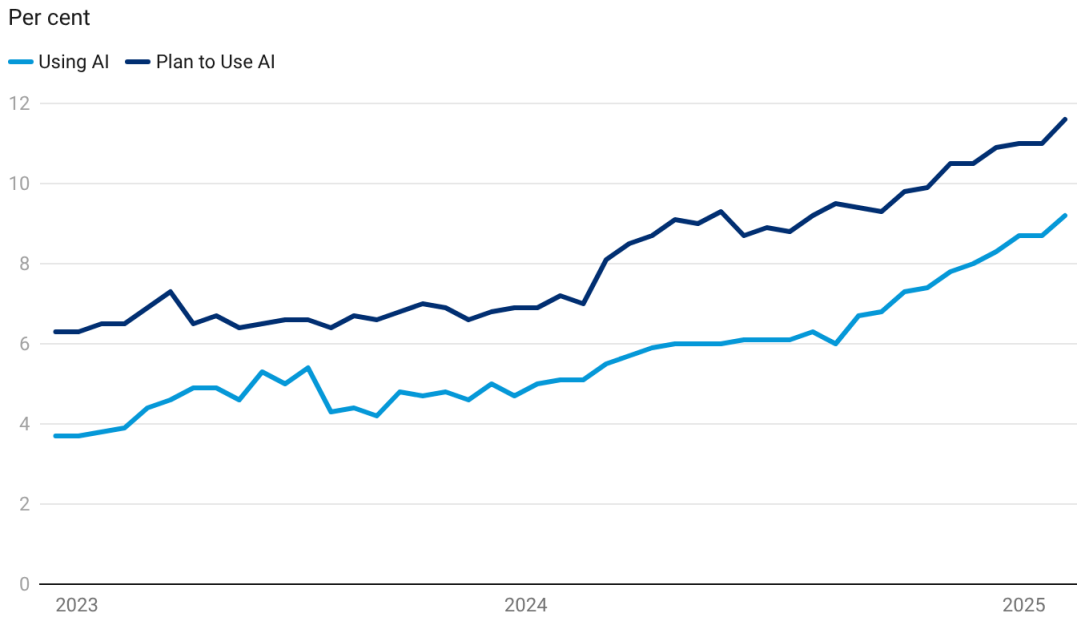
Large-scale firm surveys from the U.S. Census Bureau and McKinsey provide rather different pictures of adoption at first glance (Figures 3-1 and 3-2). The Census Bureau's Business Trends and Outlook Survey (BTOS) found that roughly 17 per cent of firms used AI (GenAI and other types of AI) as of March 2026.<sup>4</sup> In contrast, McKinsey reported that 88 per cent of firms did so in November 2025. Differences in survey coverage likely explain much of the gap. The BTOS is a represen-

tative sample of 200,000 U.S. firms, only a handful of which are large corporations. In contrast, the McKinsey survey is a convenience sample with heavy representation from large corporations (McKinsey, 2024). That is, large firms appear to use AI far more than small firms.

Surveys of workers suggest significant AI adoption with relatively few adopters reporting frequent use. Gallup, Inc. reports 50 per cent of U.S. workers had used AI at work as of the fourth quarter of 2025, though only 13 per cent of respondents used it daily. Pew Research Center reports that 21 per cent of U.S. workers used AI for at least some of their work in

<sup>4</sup> See Bonney *et al.* (2024) for details about the survey and Bonney *et al.* (2026) for a more comprehensive exploration of its implications.

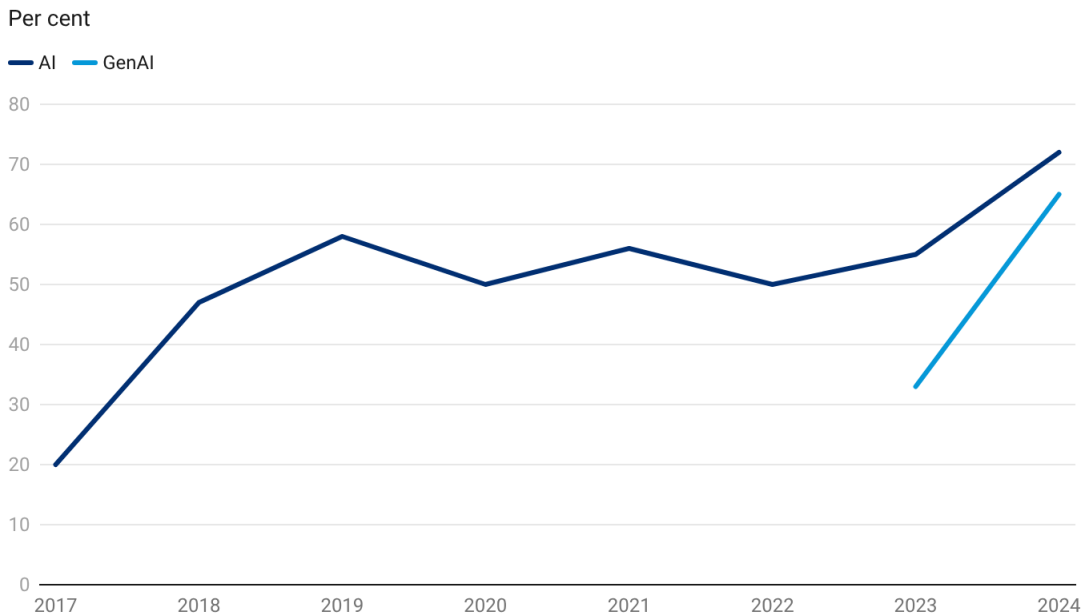
**Figure 3-1: Census Bureau Business Survey (U.S.)**



Note: Respondents were asked about AI use in producing goods or services during the past two weeks and anticipated in the next six months.

Source: Census Bureau, Business Trends and Outlook Survey. • Created with Datawrapper

**Figure 3-2: McKinsey Survey (Global)**



Note: Respondents were asked if they “use AI in at least one business function”.

Source: McKinsey, “The State of AI in Early 2024.” • Created with Datawrapper

September 2025. Bick, Blandin, and Deming (2024) ask specifically about GenAI and find nearly 40 per cent of workers used it in 2024. As new workers enter the labour force, these numbers are expected to rise. The Pew poll finds that 64 per cent of teens have used an AI chatbot and more than half of them have used them to search for information and to get help with school-work.

AI use varies significantly across industries. Both the BTOS and the McKinsey survey show that use is particularly high in the information sector, where computer coding is prevalent. While our focus in this article is not the labour market, we note two indicators of the disruption caused by AI in the information sector. Data from Lightcast show that the share of job postings in the information sector mentioning AI and related skills was 31 per cent in May 2026, up from 10 per cent in 2021, and far higher than the average across all sectors of 7 per cent (Figure 4). Crane and Soto (2026) provide evidence that labour demand growth for computer coders slowed noticeably following the release of ChatGPT.<sup>5</sup>

Kane and Baily (2025a,b) provide industry case studies summarizing the evidence of GenAI adoption in four sectors — information, healthcare, finance and electricity generation and distribution. In some examples, such as computer coding, GenAI has been adopted quickly, and in all these sectors companies were experimenting with ways to cut costs using GenAI. There are

often barriers to adoption, however, including regulation, skill shortages, and high adoption costs. In other cases, efforts to use GenAI have proven disappointing so far.

## 2.2 Knock-on Innovation

Knock-on innovations related to GenAI include novel and improved software, more efficient product and process design systems, and organizational changes to better exploit GenAI capabilities.

A prominent example of knock-on innovation is ChatGPT itself, a user interface (UI) for OpenAI’s GenAI models. UIs provide a channel through which requests and responses can pass between the GenAI model and human users. ChatGPT, released in 2022, is a conversational interface that made GenAI interactions significantly more accessible relative to early approaches that relied on Python programs or websites such as the OpenAI Playground. In 2023, OpenAI introduced custom GPTs, enabling users to create domain-specific large language models (LLMs), such as LegalGPT for legal matters (OpenAI, 2023). In 2024, OpenAI announced integration of their ChatGPT model to Apple’s Siri voice assistant and Google launched NotebookLM, which made it easy to upload documents and transform them into interactive discussions (Johnson, 2024). In addition, there are “copilots” that integrate AI into existing user workstreams, notably GitHub Copilot (computer programming) and Microsoft 365 Copilot (office productivity).

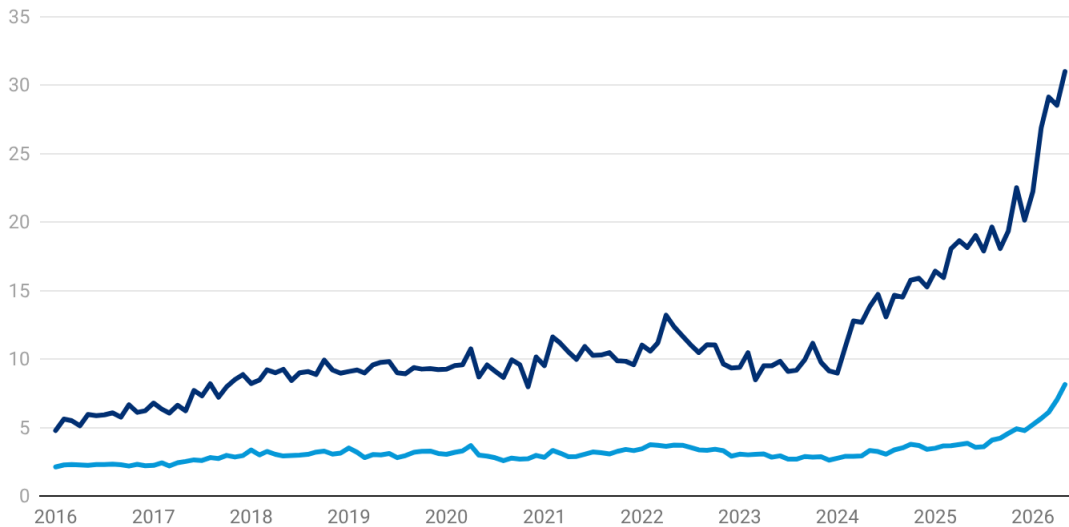
---

<sup>5</sup> On the labour market effects of AI, see Acemoglu *et al.* (2020), Brynjolfsson, Mitchell, and Rock (2018), Felten, Raj, and Seamans (2019), Webb (2019), Eloundou *et al.* (2024).

## Figure 4: AI-Related Job Postings

Per cent of total

Information Postings All Postings



Note: Jobs classified using AI-related terms found in job descriptions, as described in Acemoglu et al. (2020). List of AI-related terms updated by Lightcast.

Source: Lightcast. • Created with Datawrapper

System interfaces are another key locus of knock-on innovation for GenAI. These allow hardware and software systems to access the AI model. For example, Nvidia's Isaac Software Development Kit (SDK) facilitates the integration of AI into robotics. Access to AI through the SDK helps the robot with environmental integration problems, such as simultaneously tracking its location and mapping its environment. Development of multimodal models — which can take in inputs of different kinds (text, images, sensor readings) and output instructions to the robot, such as the rotation and torque for a joint — has pushed robot-AI integration forward (Reed *et al.*, 2022; Brohan *et al.*, 2023). System interfaces also enable the design of agentic AI systems, which collect information during operation, interpret context, make decisions and act autonomously to pursue goals (Park *et al.*,

2023). Innovations in production line operations have followed GenAI advances as well. Serradilla *et al.* (2022) provide an overview of the use of deep learning, including GenAI such as generative adversarial networks, to optimize line configuration, throughput, efficiency, and carbon footprint. Predictive maintenance using synthetic data and scenario simulation is another application for GenAI in industry.

Another area of knock-on innovation has been the creation of AI agents. Agentic AI systems develop strategies to pursue broad goals and recalibrate in response to their environment, in contrast to tool-based AI, which has a stable structure and calibration, and is only equipped to respond to carefully crafted requests. For example, specialized agents can handle health care paperwork including making appointments, following up on treatment protocols

and submitting insurance claims. In software development, agentic tools have extended the copilot approach further. Tools such as Anthropic’s *Claude Code* allow programmers to delegate command-line coding tasks, such as debugging, refactoring, and test execution, with the agent reading and modifying files autonomously rather than suggesting individual lines of code.

### 2.3 Core Innovation

Importantly, labour productivity (economic performance, as opposed to benchmark performance) only rises when more can be accomplished *while holding input costs fixed*. Accordingly, we focus below on (a) how innovations in model architecture raise GenAI model capabilities without raising training costs, (b) how hardware innovations lower the cost of computation, and (c) how richer datasets can be brought to bear on training.

#### 2.3.1 Model Development, Training, and Deployment

Early in the development of modern GenAI, advances in performance were attained primarily by increasing model scale (number of parameters), computational power, and training data (Kaplan *et al.*, 2020). That is, output (benchmark performance) was increased by adding more inputs (Figure 5). For example, GPT models had 117 million parameters in 2018 (GPT-1) and as many as 175 billion in 2020 (GPT-3) (Shree, 2020). More recently, increasing attention has been given to using these inputs more efficiently. These efficiency gains appear in four distinct stages:


leveraging insights into algorithm design to make better models, improving model pre-training, fine-tuning of those models, and strategies to reduce the cost of inference (running the deployed models).

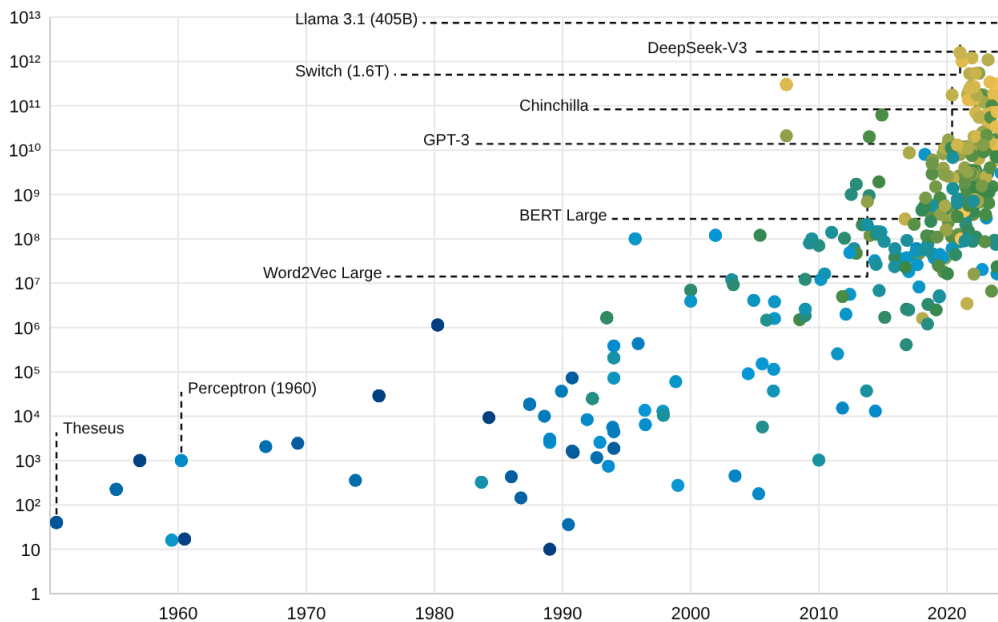
The Transformer, which enables GenAI to efficiently pay attention to context when interpreting text, was the seminal algorithmic innovation for modern GenAI (see Box 1). Further progress on this “attention mechanism” has followed. An example is Mamba (Gu and Dao, 2023). In the original transformer, computational burden is proportional to the square of the number of parameters; Mamba achieves sub-quadratic costs using a state-space model. Other innovations have focused on what can be achieved with smaller scale models; models from Microsoft and Mistral AI have shown strong performance relative to their size (Jiang *et al.*, 2023; Abdin *et al.*, 2024). Open-source model development has played a central role in this process.

Pre-training brings algorithms to a dataset to produce a broadly applicable “foundation model.” Between 2018 and 2022, a key pre-training tactic in the effort to improve GenAI performance was to increase model size. Remarkably, algorithmic innovation has more than offset the resulting increase in computational complexity: the cost of training a model of a given size was halved approximately every eight months through 2024 (Ho *et al.*, 2024). This focus on scale abated to some degree over time: by 2022, diminishing returns to model size for foundation models had appeared and researchers have increasingly explored performance improvements in fine-tuning and inference (Zeff, 2024).

**Figure 5: Number of Parameters and Training Dataset Size**

Number of Parameters (log scale)

Training Dataset Size (log scale, # of obs.) 0  30



Note: Only models with reported parameter size and training dataset size are included in the dataset. For instance, GPT-4 is excluded as its parameter size is not known.

Source: Epoch (2024) with major processing by Our World in Data (Rahman, Owen, and You 2024).

Fine-tuning refines the foundation model for a specific application. Fine-tuning innovations have included transfer learning (adapting an already fine-tuned model to a related task with domain-specific data), instruction tuning (guiding the model to recognize instructions, not just predict the next word; Taori *et al.*, 2023), and reinforcement learning from human feedback (aligning the model outputs with human preferences; Christiano *et al.*, 2017;

Ouyang *et al.*, 2022). Once pre-trained and fine-tuned, the model is used in inference (responding to user requests). The aggregate cost of GenAI inference — in terms of electricity, time, computation, and carbon emissions — has risen with the popularity of GenAI, leading to a focus on techniques to make this step more efficient.<sup>6</sup>

Conversely, some recent models have deliberately extended inference time to improve performance with such techniques

<sup>6</sup>Among these innovations are the Mixture of Experts (MoE) approach — only activating a subset of model parameters (Jacobs *et al.*, 1991; Shazeer *et al.*, 2017); pruning — removing extraneous parameters (Cetin *et al.*, 2024); distillation — compressing large models into smaller ones (Hinton, Vinyals, and Dean, 2015); quantization — reducing numerical precision (Wang *et al.*, 2023); and token caching — storing reusable computations (Pope *et al.*, 2023).

## Box 1: The Transformer

The transformer architecture, introduced by Vaswani *et al.* (2017), was a game-changer in AI, particularly as the engine behind GenAI models. Its key innovation, the “attention mechanism,” steers models to focus selectively on relevant parts of the prompt, enabling more efficient and accurate processing of language. This breakthrough has powered major advancements in natural language understanding, translation, and generation, forming the backbone of today’s most advanced GenAI systems.

Transformers process input data through a series of layers (steps), each consisting of an attention mechanism followed by a multilayer perceptron (MLP, defined below), proceeding as follows.

First, a representation of the prompt (input text) suitable for analysis by the model is created. Specifically, the prompt is broken into tokens (smaller pieces which may be phrases, words, or parts of words). The tokens are converted into embeddings (numerical vector representations) which encode the semantic and syntactic meaning of each token. Loosely speaking, for each token, the closest of the other tokens, as measured by the distance between their embeddings, are the ones most important to understanding its meaning.

Second, the attention mechanism processes the matrix of token embeddings using three large matrices called the “query,” the “key,” and the “value.” For each token in the input, the query is compared to the keys of all tokens to compute attention scores, which are used to form a weighted average of values. This step allows each token’s representation to incorporate information from other tokens in the prompt based on their contextual relevance.

Third, the data passes through an MLP, a type of neural network. While the attention mechanism focuses on pairwise interactions between tokens, the MLP applies nonlinear functions (in contrast to the linear attention mechanism) in refining the token representations.

This sequence—of computing the attention mechanism followed by the MLP—is repeated multiple times depending on how many layers are in the model (for example, the Llama-3 model has 32 layers), enabling the model to capture increasingly abstract features of the input text.

The performance gains from scaling of this system through increasing the size of these matrices—along with larger training datasets and improvements in hardware and processing algorithms—underpins the rising ability to handle complex language tasks.

as chain-of-thought reasoning (Wei *et al.*, 2022). A salient example of the cumulative effect of these innovations is DeepSeek R1, released in January 2025, which blended several of these approaches, including MoE, chain-of-thought reasoning, and reinforcement learning and distillation, to achieve frontier-level performance at a fraction of the cost of comparable models (DeepSeek-AI, 2024).

### 2.3.2 Hardware

In addition to the software and operational choices described above, GenAI model training and inference costs rely critically on the state of electronic hardware. Because the workloads generated by GenAI training are best handled with massive parallel computation, graphics processing units (GPUs), which are designed for

parallel processing, play a central role.<sup>7</sup> Successive GPUs released by NVIDIA have delivered leaps in AI performance through improvements in processing core (CUDA) design, the addition of on-chip tensor cores, which accelerate matrix calculations, and massive amounts of high-speed integrated memory. In addition to these circuit design innovations, hardware cost improvement depends on advances in basic and applied science that permit greater miniaturization and power-efficiency for electronics.<sup>8</sup> While the engineering performance of leading-edge GPUs has rocketed upwards in recent years, prices have increased dramatically as well, but by less than performance: In 2007, a \$349 GPU provided 0.3 teraflops (TFLOPS) of compute and in 2024, a \$299 GPU delivered 15.1 TFLOPS, implying an average annual rate of price decline of 24 per cent that persisted for 17 years (Figure 6). This example is representative of a broader trend.

More recently, NVIDIA’s Blackwell GPU architecture, introduced in 2024, roughly quadrupled the inference throughput of LLMs (tokens per second) at comparable power consumption, relative to prior models like the Hopper. Such hardware gains amplify the algorithmic efficiency improvements described above for training as well. Given that agentic AI systems involve long, multi-turn interactions in which input text

is reused across steps, these types of gains are particularly consequential for the rise of agents discussed in Section 2.2, where inference costs rather than training costs are binding constraints on adoption.

### 2.3.3 Datasets

GenAI models “learn” by adjusting parameters to best represent the content of large amounts of text, allowing them to choose the word that should appear next in response to a user’s prompt. Figure 5 illustrates the increase over time in the size of the datasets used in training and the associated improvement in (benchmark) performance. A crucial nuance to this aspect of model improvement is that access to more information, not more text, is needed to continue to improve GenAI models. Diminishing marginal returns set in as developers move from information-rich content, such as Wikipedia and scientific articles to noisier text like social media posts.

One approach to mitigating the content constraint is transfer learning, where a model pre-trained with public data is improved by further training using proprietary data. Another approach is data augmentation, such as incorporating small, localized modifications of the training data. For example, the performance of an image recognition model may be improved by

---

<sup>7</sup> Specialized chips, such as tensor processing units (TPUs) customized for matrix multiplication, have increased computational efficiency as well.

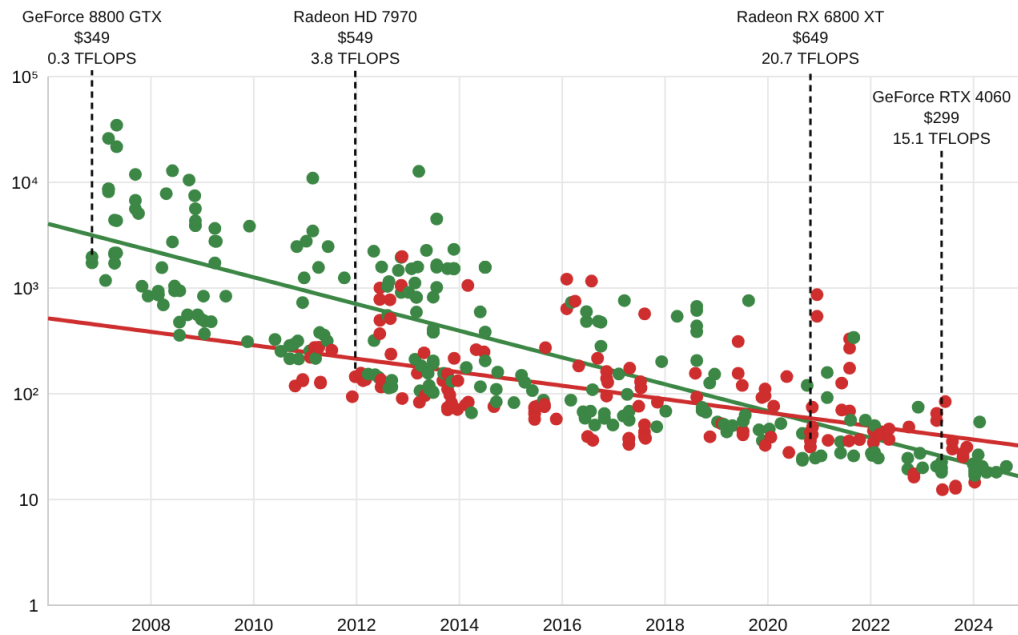
<sup>8</sup> The frequent release of new generations of semiconductors belies the difficult challenges faced in achieving each one. The Institute of Electrical and Electronics Engineers regularly publishes “roadmaps” detailing the problems that must be solved for continued improvement in electronics performance (see International Roadmap for Devices and Systems (IRDS™) 2024 Edition <https://irds.ieee.org/editions/irds2024/>.)

<sup>9</sup> This approach was taken by the developers of AlexNet, a model which revolutionized the field (Krizhevsky, Sutskever, and Hinton, 2012).

## Figure 6: Price Trends for GPUs, Price per TFLOP

Dollars per Trillion Floating-Point Operations per Second (log scale)

Company ● AMD ● NVIDIA



Note: The green line represents the line of best fit for NVIDIA, and the red line represents the line of best fit for AMD.

Source: TechPowerUp.

supplementing the training set of labeled images with their mirror images.<sup>9</sup> “Synthetic data”, where generative models create new data to emulate patterns and characteristics of real data has been explored as well (Liu *et al.*, 2024).<sup>10</sup> Last, datasets can be augmented by harvesting information collected with sensors, particularly in physical environments such as industrial robots and autonomous vehicles (Feng *et al.*, 2019).

### 3. GenAI as Invention in the Method of Invention

Like other sectors, efficiency in the research sector can be increased using appropriate capital, such as inventions in the method of invention (IMI). We consider below whether GenAI is such an invention and whether it can contribute to research productivity beyond what is contributed by predecessor AI technologies. We then review broad indicators of AI’s role in

<sup>10</sup> Some observers have raised concerns that training with synthetic data (and AI-generated text increasingly present on the internet) will yield low-quality or even nonsensical results, a phenomenon known as “model collapse” (Alemohammad *et al.*, 2023; Shumailov *et al.*, 2023). Others have argued that model collapse only occurs when the original training text is replaced by model-generated text (Gerstgrasser *et al.*, 2024).

research: patent filings, the share of AI use by workers in research roles, and new evidence on AI references in company conference calls.

Eloundou *et al.* (2024) note that “scientists and researchers” and “technologists” are the job groups most highly exposed to LLMs, suggesting substantial potential for research and development (R&D) productivity enhancement from GenAI. Prior to GenAI, AI had already diffused widely across scientific disciplines and improved research efficiency (Carobene *et al.*, 2024). Cockburn, Henderson, and Stern (2019; 2023) note that pre-generative AI assists with the “labour-intensive search with high marginal cost of search” involved in many types of R&D. Agrawal *et al.* (2018), emphasize that AI improves prediction, which plays a central role in research; for example, machine learning has also been used extensively for predicting the properties of novel metal alloys, economizing on physical experimentation and computer simulations (Hart *et al.*, 2021). Our focus is on the question of whether GenAI enables additional efficiencies in R&D beyond these and other improvements provided by machine learning. We group IMIs into enhancements to observation, analysis, communication, and organization.

### 3.1 Observational

Observational tools, such as microscopes, telescopes, and cameras are central to scientific advance (Mokyr, 2004). These tools are invariably limited in that they

produce imperfect data due to defects in their components and variation in the environment. GenAI provides a tool to enhance imperfect portions of data. For example, generative techniques for image enhancement perform better than techniques, such as splines, relying solely on smoothness assumptions (i.e., that nature does not make leaps) (Liu *et al.*, 2018; Lugmayr *et al.*, 2022). GenAI can approximate the manifold of the data generating process, implicitly accounting for the actual physics, say, of a remote galaxy seen through an imperfect lens. More generally, imputation of missing observations in datasets of all kinds is possible in a fashion consistent with the apparent properties of the underlying phenomena.

### 3.2 Analytical

GenAI can be used as an analytical tool as well. Caliskan, Bryson, and Narayanan (2017), for example, find that “text corpora contain recoverable and accurate imprints of our historic biases.” This new visibility may promote analysis of previously intractable social science questions. There has been an explosion of sentiment analysis and other forms of natural language processing in recent years fueled by this capability of GenAI.<sup>11</sup> While the identification of underlying sentiment (encoding) is strictly speaking, a function of the LLM, conveying the discovered sentiment to the user is necessarily a generative process. Korinek (2023) documents a variety of potential roles for GenAI in the economic

---

<sup>11</sup> Sentiment analysis is possible with earlier forms of AI but the capabilities of GenAI models are vastly greater (Gentzkow, Kelly, and Taddy, 2019; Dell, 2025).

research process; that GenAI may play a similar role in many other fields is a reasonable conjecture.

### 3.3 Organizational

The organization of science may benefit from the use of GenAI. Institutional organization plays a central role in the effectiveness of R&D (Mowery and Rosenberg, 1999), as do informal associations into professional networks (Wang and Barabasi, 2021) and geographic clusters (Porter and Stern, 2001). Consequently, the method of invention for any given research program properly includes the institutions involved. Emerging applications of AI “digital twins” offer the prospect of R&D with a reduced institutional footprint in many areas of study. Among these are drug discovery (Bordukova *et al.*, 2024), industrial research (Tao, Zhang, and Zhang, 2024), and materials science (Kalidindi *et al.*, 2022). GenAI tools can help with applied science as well, such as designing products that meet technical and aesthetic specifications. Moreover, the design process itself can be transformed to create a range of options, not just one-off designs, together with detailed manufacturing specifications (Saadi and Yang, 2023).

### 3.4 Communication

Perhaps most obviously, GenAI is a communication tool. Although empirical and analytical stages of research projects focus on measurement and calculation, many aspects of the research process involve manipulating language. GenAI may be employed in the writing tasks involved in the conceptual, planning, and dissemination stages of

research projects, such as drafting literature reviews, grant applications, and seminar slides. Whether GenAI improves the efficiency of such tasks on net, once the effort needed for review and editing of the documents drafted by GenAI is accounted for, is an open question. If so, GenAI may play a similar role to the printing press and word processing as a catalyst to the invention process.

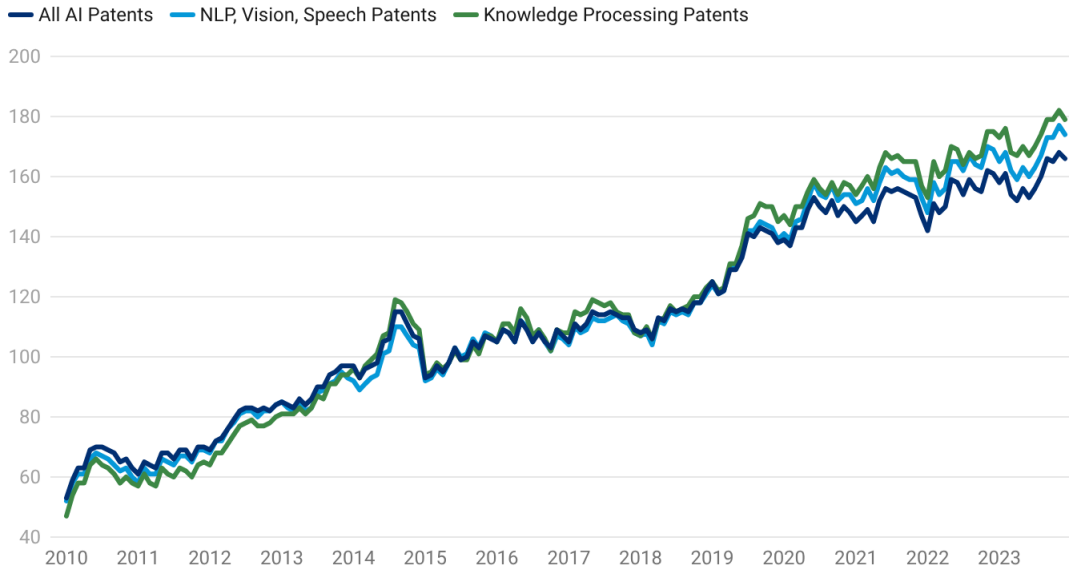
### 3.5 Indicators of GenAI Research and of GenAI Use in Research

Substantial suggestive evidence has emerged that GenAI has enhanced research performance. AI-related patents issued by the United States Patent and Trademark Office (USPTO) increased following the advent of GenAI, suggesting a related surge in GenAI research (Figure 7) (Pairolero, 2025). The USPTO index of AI-related patents began climbing in 2018, shortly after the publication of the seminal paper by Vaswani *et al.* (2017) quickly reaching a level 50 per cent higher, which it has sustained since 2019. We also observe that increases in patent activity for AI modalities particularly related to GenAI — natural language processing, vision, speech, and knowledge processing — have risen even further. This suggests that the recent surge in patenting activity is not merely a reflection of advancements in machine learning.

Handa *et al.* (2025) provide a rich set of information on GenAI use in their Anthropic Economic Index (AEI), assigning millions of conversations from Claude (Anthropic’s premier GenAI system) to roughly 3,500 of the tasks defined by the U.S. Department of Labor’s O\*NET

## Figure 7: Artificial Intelligence Mentions in Scientific Patents

Indexed to 100 in 2015, three month moving average



Source: Artificial Intelligence Patent Dataset (2023), U.S. Patent Office. • Created with Datawrapper

Dataset. Table 1 shows the estimated share of prompts accounted for by occupational groups, their employment share, and the ratio of the two. (If prompts were equally distributed across all workers, these ratios would each be equal to 1.) “Computer and mathematical occupations”, which includes the computer programmers for whom GenAI use is especially intense, have the highest ratio of prevalence of GenAI use to occupational prevalence and use intensity is nearly as high among scientists. Other occupational groups with high relative prevalence of GenAI use include “arts, design, sports, entertainment and media”; “architecture and engineering”; and “educational instruction and library”. The remaining 87.6 per cent of employment is accounted for by occupations which AEI found had a share of Claude prompts roughly equal to or lower than their share of employment, highlighting the concentrated nature of GenAI adoption in

the economy at present.

Figure 8 illustrates significant automation and augmentation of tasks among our groupings of research occupations: programmers exhibit the highest automation rate, with over half of the requests handled by GenAI being automation tasks. Social science researchers show slightly lower automation rates, with economists showing over 23 per cent of their prompts being automation focused. Notably, for hard science researchers (e.g., physicists, biochemists), the share of their GenAI use for automation is nearly 15 per cent higher than their natural science counterparts. This difference likely reflects AI’s strength in data-intensive and simulation-based research such as those found in hard sciences like physics and materials science.

Firm communication also reveals GenAI’s integration into the invention process. Figure 9 plots the number of firms referencing AI in the context of research,

**Table 1: Occupations with High Generative AI Use**

Job Type	Prompt Share	Employment Share	Ratio
Computer and mathematical	37.2	3.4	10.9
Arts, design, sports, entertainment, and media	10.3	1.4	7.4
Life, physical, and social science	6.4	0.9	7.1
Architecture and engineering	4.5	1.7	2.6
Educational instruction and library	9.3	5.8	1.6
Memo: Other occupations	31.8	87.6	0.4

Note: Percent share of prompts submitted to Claude AI and linked to tasks by Anthropic. Task weights are apportioned equally to all occupations which include that task in O\*NET.

Source: Anthropic Economic Index. • Created with Datawrapper

as indicated by the firms mentioning an AI-specific term (“machine learning,” “deep learning,” “artificial intelligence,” “GenAI,” or “generative AI”) within a research-related context (within 10 words of “inventi-”, “research-”, or “discover”). A sudden rise appears in 2023, with approximately 60 public companies per quarter mentioning such usage. This increased integration of AI with R&D illustrates the role it plays in corporate innovation.

#### 4. Tailwinds and Headwinds for Productivity Growth from GenAI

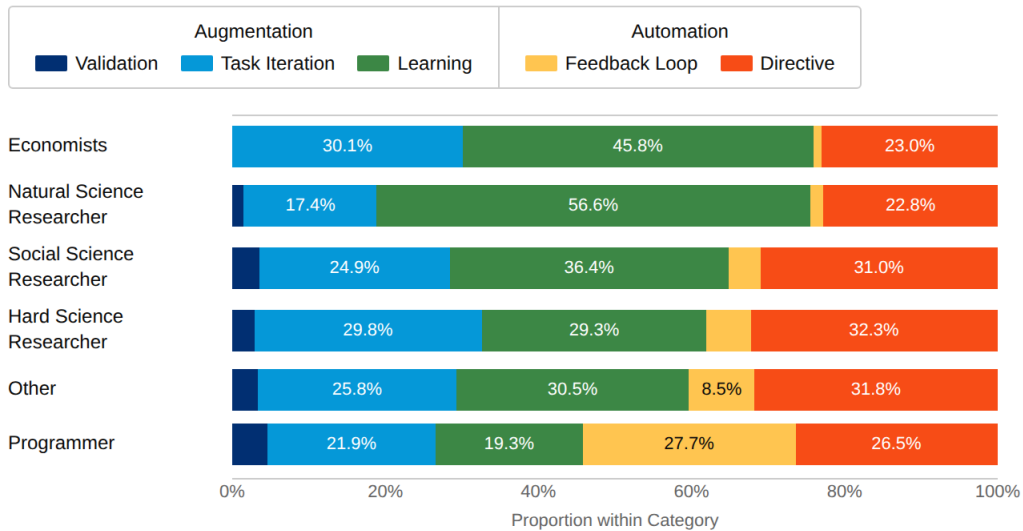
The qualities of GenAI and the limited evidence on its application suggest that two substantial tailwinds support a forecast of a noteworthy increase in productivity from the technology. GenAI has features typical of both a GPT — headed toward being widely used, stimulating related

innovation, and displaying ongoing improvement in (economic) performance — and an IMI — raising the efficiency of R&D through improvements to observation, analysis, communication, or organization. Because both GPTs and IMIs promote productivity growth for extended periods, it is reasonable to expect GenAI will have a noteworthy impact on productivity. That being said, we note several headwinds that should be taken into account.

First, whether the organizational change needed for GenAI to be a true GPT will take place is an open question. AI systems that preceded GenAI demonstrated the need for cross-functional teams with access to data that spans the enterprise, breaking down barriers between business units, optimizing supply chains, and re-allocating employees to de-emphasize repetitive writing tasks (Iansiti and Lakhani, 2020). Bresnahan (2024) observed that adoption was concentrated in places where

## Figure 8: GenAI Automation vs. Augmentation in Researcher Roles

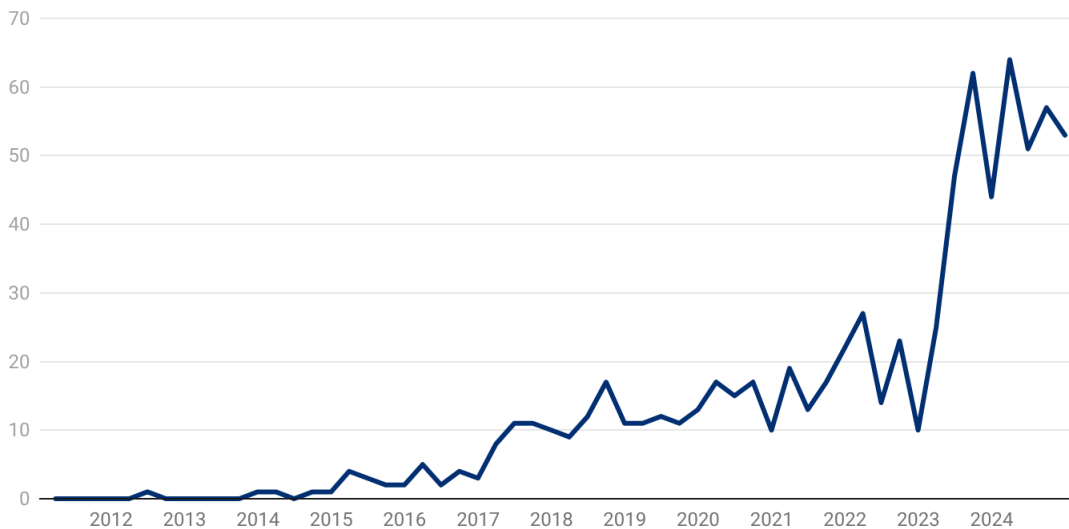
Proportion of AI interactions within each job category, by interaction type



Source: Authors' calculations using the Anthropic Economic Index.

## Figure 9: Mentions of Artificial Intelligence Use for Research in Conference Calls

Number of firms mentioning AI and research

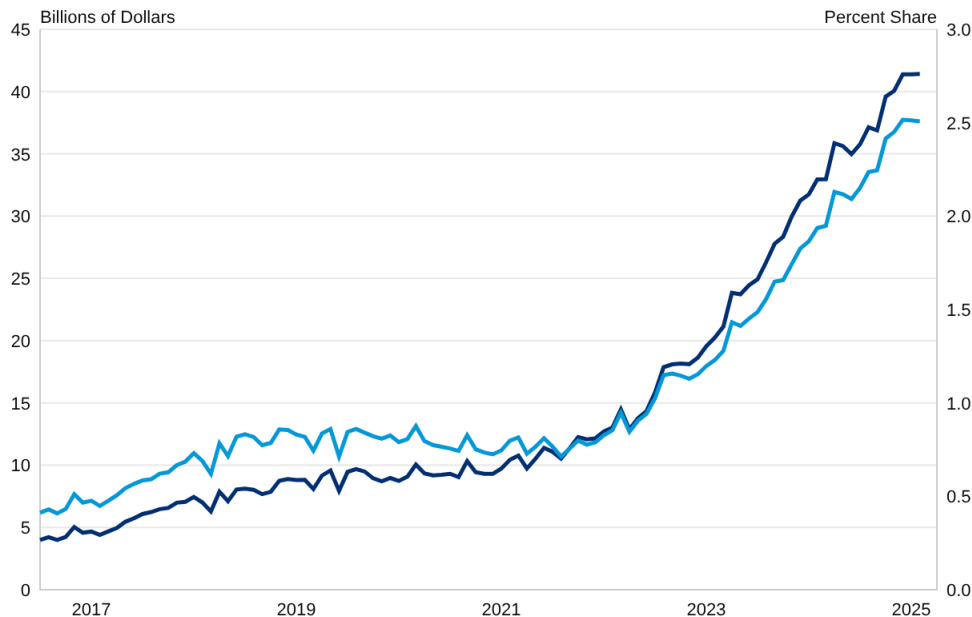


Note: Authors' estimates. Data through Q4 2024.

Source: S&P Capital database merged with Compustat. • Created with Datawrapper

**Figure 10: Data Center Construction**

— Construction Spending (Left Axis, Billions of Dollars) — Non-residential Investment (Right Axis, Percent Share)



Note: Excludes IT equipment. Nominal value of construction put in place.

Source: U.S. Census Bureau.

complementary innovation was less necessary, such as in firms that were highly digitized from their founding. These digital natives will surely lead the charge for GenAI as well. For other firms, the pace and success of reorganization innovation will be a key determinant of the scale and timing of productivity effects from GenAI.

Relatedly, the reliance of GPTs on complementary investment tends to damp the effect on labour productivity growth. For example, the effect on the productivity *level* of solid-state computing was large, but it played out over decades. Massive advances in computational technology, including the invention of the solid-state transistor and the fundamentals of system design had accumulated by the end of the 1940s and a steady decline in computing

costs had begun (Nordhaus, 2007). The surge in productivity attributed to information technology arrived some fifty years later.

Third, investment to develop and deploy new technologies is fraught with risk. If GenAI is a widely adopted “killer app” that defines a new era of IT, the computing capacity needed to deliver GenAI to millions of simultaneous users will be massive. Anticipation of this outcome helps explain the recent wave of irreversible investment in data centers and power generation (Figure 10). Historically, when such forecasts have proven wrong, the negative consequences of the resulting capital overhang have been substantial.

Fourth, the scope of application of GenAI as an IMI remains to be seen as well. For example, whether GenAI can uncover fundamental features of phenomena is a matter of some debate. Li *et al.* (2022) present evidence that GenAI does develop such knowledge in an “emergent world model.” Others argue that GenAI is employing a “bag of heuristics.”<sup>12</sup> This question is a crucial one in determining the capabilities of GenAI to contribute to science. Without a model of underlying structure, one cannot articulate fundamental laws. This limitation may be inherent to how GenAI is trained: humans learn scientific fundamentals from textbooks, but such laws may not form the rhetorical backbone of the verbal exchanges that dominate training corpora.

Fifth, we expect that GenAI will boost productivity growth *relative to the counterfactual* economy without it, but the growth effect of machine learning (and other IT innovations) may be waning. The impact of GenAI will have to match the impact of machine learning for the economy simply to match the recent history of productivity growth. In other words, the digital revolution may be baked into the productivity trend and GenAI is just its latest form.

## 5. Conclusion

The release of ChatGPT in late 2022 was a stark inflection point in public interest in GenAI and predictions of a first-order impact on productivity in the future soon followed, but its economic effects remain

uncertain. To complement the limited empirical evidence, we ask what the characteristics of GenAI suggest its future impact on productivity may be. We conclude there is strong evidence that GenAI has the potential to be both a GPT and an IMI. We therefore expect a noteworthy increase in labour productivity from GenAI, though the headwinds we cite suggest the range of plausible outcomes is wide with respect to both the magnitude and timing of the increase.

## References

- Abdin, M., J. Aneja, H. Awadalla, et al. (2024) “Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone,” arXiv Preprint arXiv:2404.14219. Available at: <https://arxiv.org/abs/2404.14219>.
- Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo (2020) “AI and Jobs: Evidence from Online Vacancies,” NBER Working Paper 28257, National Bureau of Economic Research.
- Acemoglu, D. and P. Restrepo (2020) “The Wrong Kind of AI? Artificial Intelligence and the Future of Labour Demand,” *Cambridge Journal of Regions, Economy and Society*, Vol. 13, No. 1, pp. 25–35.
- Agrawal, A., J. Gans, and A. Goldfarb (2018) *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business Review Press, Boston.
- Alemohammad, S., J. Casco-Rodriguez, L. Luzi, A.I. Humayun, H. Babaei, D. LeJeune, A. Siahkoohi, and R. Baraniuk (2023) “Self-consuming Generative Models Go Mad,” in *The Twelfth International Conference on Learning Representations*.
- Baily, M.N., D.M. Byrne, A.T. Kane, and P.E. Soto (2025) “Generative AI at the Crossroads: Light Bulb, Dynamo, or Microscope?” Working Paper, Center on Regulation and Markets, Brookings Institution.
- Baily, M.N. and A. Kane (2025a) “AI in the Finance Sector: Transforming Productivity and Risk Management,” Brookings Institution.

<sup>12</sup> A useful entry point to this ongoing debate is “LLMs and World Models,” by Melanie Mitchell, February 13, 2025, found at the AI: A Guide for Thinking Humans Substack blog.

- Baily, M.N. and A. Kane (2025b) “AI in the Healthcare Sector: Enhancing Care, Efficiency, and Innovation,” Brookings Institution. Available at: <https://www.brookings.edu/articles/harnessing-ai-for-economic-growth/>.
- Bick, A., A. Blandin, and D.J. Deming (2024) “The Rapid Adoption of Generative AI,” NBER Working Paper 32966, National Bureau of Economic Research.
- Bonney, K., C. Breaux, C. Buffington, E. Dinlersoz, L. Foster, N. Goldschlag, J. Haltiwanger, Z. Kroff, and K. Savage (2024) “Tracking Firm Use of AI in Real Time: A Snapshot from the Business Trends and Outlook Survey,” NBER Working Paper 32319, National Bureau of Economic Research.
- Bonney, K., C. Breaux, E. Dinlersoz, L. Foster, J. Haltiwanger, and A. Pande (2026) “The Microstructure of AI Diffusion: Evidence from Firms, Business Functions, and Worker Tasks,” NBER Working Paper 35141, National Bureau of Economic Research.
- Bordukova, M., N. Makarov, R. Rodriguez-Esteban, F. Schmich, and M.P. Menden (2024) “Generative Artificial Intelligence Empowers Digital Twins in Drug Discovery and Clinical Trials,” *Expert Opinion on Drug Discovery*, Vol. 19, No. 1, pp. 33–42.
- Bresnahan, T. (2019) “Artificial Intelligence Technologies and Aggregate Growth Prospects,” in J.W. Diamond and G.R. Zodrow (eds.), *Prospects for Economic Growth in the United States*, Cambridge University Press, Cambridge, England, pp. 132–172.
- Bresnahan, T. (2024) “What Innovation Paths for AI to become a GPT?” *Journal of Economics & Management Strategy*, Vol. 33, No. 2, pp. 305–316.
- Brohan, A., N. Brown, J. Carbajal, et al. (2023) “Rt-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control,” arXiv Preprint arXiv:2307.15818.
- Brynjolfsson, E., T. Mitchell, and D. Rock (2018) “What can Machines Learn and What Does it Mean for Occupations and the Economy?” *AEA Papers and Proceedings*, Vol. 108, pp. 43–47.
- Caliskan, A., J.J. Bryson, and A. Narayanan (2017) “Semantics Derived Automatically from Language Corpora contain Human-like Biases,” *Science*, Vol. 356, No. 6334, pp. 183–186.
- Carobene, A., A. Padoan, F. Cabitza, G. Banfi, and M. Plebani (2024) “Rising Adoption of Artificial Intelligence in Scientific Publishing: Evaluating the Role, Risks, and Ethical Implications in Paper Drafting and Review Process,” *Clinical Chemistry and Laboratory Medicine (CCLM)*, Vol. 62, No. 5, pp. 835–843.
- Cetin, E., Q. Sun, T. Zhao, and Y. Tang (2024) “An Evolved Universal Transformer Memory,” arXiv Preprint arXiv:2410.13166.
- Christiano, P.F., J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei (2017) “Deep Reinforcement Learning from Human Preferences,” *Advances in Neural Information Processing Systems*, Vol. 30.
- Cockburn, I.M., R. Henderson, and S. Stern (2019) “The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis,” in A. Agrawal, J. Gans, and A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, Chicago, pp. 115–146.
- Crane, L.D. and P.E. Soto (2026) “AI and Coder Employment: Compiling the Evidence,” Federal Reserve Board Working Paper, FEDS 2026018.
- DeepSeek-AI, A. Liu, B. Feng, B. Xue, et al. (2024) “DeepSeek-V3 Technical Report,” arXiv Preprint. Available at: <https://arxiv.org/abs/2412.19437>.
- Dell, M. (2025) “Deep Learning for Economists,” *Journal of Economic Literature*, Vol. 63, No. 1, pp. 5–58.
- Eloundou, T., S. Manning, P. Mishkin, and D. Rock (2024) “GPTs are GPTs: Labor Market Impact Potential of LLMs,” *Science*, Vol. 384, No. 6702, pp. 1306–1308.
- Felten, E.W., M. Raj, and R. Seamans (2019) “The Occupational Impact of Artificial Intelligence: Labor, Skills, and Polarization,” Working Paper, NYU Stern School of Business.
- Feng, D., X. Wei, L. Rosenbaum, A. Maki, and K. Dietmayer (2019) “Deep Active Learning for Efficient Training of a Lidar 3d Object Detector,” in *2019 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, pp. 667–674.
- Gentzkow, M., B. Kelly, and M. Taddy (2019) “Text as Data,” *Journal of Economic Literature*, Vol. 57, No. 3, pp. 535–574.
- Gerstgrasser, M., R. Schaeffer, A. Dey, R. Rafailov, H. Sleight, J. Hughes, T. Korbak, R. Agrawal, D. Pai, A. Gromov, D.A. Roberts, D. Yang, D.L. Donoho, and S. Koyejo (2024) “Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data,” arXiv Preprint arXiv:2404.01413.
- Goldfarb, A., B. Taska, and F. Teodoridis (2023) “Could Machine Learning be a General Purpose Technology? A Comparison of Emerging Technologies Using Data from Online Job Postings,” *Research Policy*, Vol. 52, No. 1, article 104653.
- Gu, A. and T. Dao (2023) “Mamba: Linear-time Sequence Modeling with Selective State Spaces,” arXiv Preprint arXiv:2312.00752.
- Handa, K., A. Tamkin, M. McCain, S. Huang, E. Durmus, S. Heck, J. Mueller, J. Hong, S. Ritchie, T. Belonax, K.K. Troy, D. Amodei, J. Kaplan, J. Clark, and D. Ganguli (2025) “Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations,” arXiv Preprint arXiv:2503.04761.

- Hart, G.L.W., T. Mueller, C. Toher, and S. Curtarolo (2021) “Machine Learning for Alloys,” *Nature Reviews Materials*, Vol. 6, No. 8, pp. 730–755.
- Hinton, G., O. Vinyals, and J. Dean (2015) “Distilling the Knowledge in a Neural Network,” arXiv Preprint arXiv:1503.02531. Available at: <https://arxiv.org/abs/1503.02531>.
- Ho, A., T. Besiroglu, E. Erdil, D. Owen, R. Rahman, Z.C. Guo, D. Atkinson, N. Thompson, and J. Sevilla (2024) “Algorithmic Progress in Language Models,” arXiv Preprint arXiv:2403.05812.
- Iansiti, M. and K.R. Lakhani (2020) *Competing in the Age of AI: Strategy and Leadership when Algorithms and Networks run the World*, Harvard Business Press, Cambridge, Massachusetts.
- Jacobs, R.A., M.I. Jordan, S.J. Nowlan, and G.E. Hinton (1991) “Adaptive Mixtures of Local Experts,” *Neural Computation*, Vol. 3, No. 1, pp. 79–87.
- Jiang, A.Q., A. Sablayrolles, A. Mensch, C. Bamford, D.S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L.R. Lavaud, M.-A. Lachaux, P. Stock, T. Le Scao, T. Lavril, T. Wang, T. Lacroix, and W. El Sayed (2023) “Mistral 7B,” arXiv Preprint arXiv:2310.06825.
- Johnson, S. (2024) “NotebookLM Gets a New Look, Audio Interactivity and a Premium Version,” *Google: The Keyword*, December 13, available at: <https://blog.google/innovation-and-ai/models-and-research/google-labs/notebooklm-new-features-december-2024/>.
- Kalidindi, S.R., M. Buzzy, B.L. Boyce, and R. Dingreville (2022) “Digital Twins for Materials,” *Frontiers in Materials*, Vol. 9, article 818535.
- Kane, A. and M.N. Baily (2025a) “AI in the Electricity Sector: Optimizing Grid Management and Energy Use,” Brookings Institution.
- Kane, A. and M.N. Baily (2025b) “AI in the Information Sector: Advancing Software, Customer Service, and Design,” Brookings Institution.
- Kaplan, J., S. McCandlish, T. Henighan, T.B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei (2020) “Scaling Laws for Neural Language Models,” arXiv Preprint arXiv:2001.08361.
- Korinek, A. (2023) “Generative AI for Economic Research: Use Cases and Implications for Economists,” *Journal of Economic Literature*, Vol. 61, No. 4, pp. 1281–1317.
- Krizhevsky, A., I. Sutskever, and G.E. Hinton (2012) “ImageNet Classification with Deep Convolutional Neural Networks,” *Advances in Neural Information Processing Systems*, Vol. 25.
- Li, K., A.K. Hopkins, D. Bau, F. Viégas, H. Pfister, and M. Wattenberg (2022) “Emergent World Representations: Exploring a Sequence Model Trained on a Synthetic Task,” arXiv Preprint arXiv:2210.13382.
- Lipsey, R.G., K.I. Carlaw, and C.T. Bekar (2005) *Economic Transformations: General Purpose Technologies and Long-term Economic Growth*, OUP Oxford.
- Liu, G., F.A. Reda, K.J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro (2018) “Image Inpainting for Irregular Holes using Partial Convolutions,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 85–100.
- Liu, R., J. Wei, F. Liu, C. Si, Y. Zhang, J. Rao, S. Zheng, D. Peng, D. Yang, D. Zhou, and A.M. Dai (2024) “Best Practices and Lessons Learned on Synthetic Data for Language Models,” arXiv Preprint arXiv:2404.07503.
- Lugmayr, A., M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool (2022) “Repaint: Inpainting using Denoising Diffusion Probabilistic Models,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11461–11471.
- McKinsey (2024) “The State of AI in Early 2024: Gen AI Adoption Spikes and Starts to Generate Value,” McKinsey & Company.
- Mokyr, J. (2004) *The Gifts of Athena: Historical Origins of the Knowledge Economy*, Princeton University Press.
- Mowery, D.C. and N. Rosenberg (1999) *Paths of Innovation: Technological Change in 20th-Century America*, Cambridge University Press, Cambridge, England.
- Nordhaus, W.D. (2007) “Two Centuries of Productivity Growth in Computing,” *The Journal of Economic History*, Vol. 67, No. 1, pp. 128–159.
- OpenAI (2023) “Introducing GPTs,” November 6, available at: <https://openai.com/index/introducing-gpts/>.
- Ouyang, L., J. Wu, X. Jiang, D. Almeida, C.L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe (2022) “Training Language Models to Follow Instructions with Human Feedback,” *Advances in Neural Information Processing Systems*, Vol. 35, pp. 27730–27744.
- Pairolero, N.A., A.V. Giczy, G. Torres, T.I. Erana, M.A. Finlayson, and A.A. Toole (2025) “The Artificial Intelligence Patent Dataset (AIPD) 2023 Update,” *The Journal of Technology Transfer*, pp. 1–24.
- Park, J.S., J. O’Brien, C.J. Cai, M.R. Morris, P. Liang, and M.S. Bernstein (2023) “Generative Agents: Interactive Simulacra of Human Behavior,” in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–22.
- Pope, R., S. Douglas, A. Chowdhery, J. Devlin, J. Bradbury, J. Heek, K. Xiao, S. Agrawal, and J. Dean (2023) “Efficiently Scaling Transformer Inference,” *Proceedings of Machine Learning and Systems*, Vol. 5, pp. 606–624.

- Porter, M.E. and S. Stern (2001) "Innovation: Location Matters," *MIT Sloan Management Review*.
- Reed, S., K. Zolna, E. Parisotto, S.G. Colmenarejo, A. Novikov, G. Barth-Maron, M. Gimenez, Y. Sulsky, J. Kay, J.T. Springenberg, T. Eccles, J. Bruce, A. Razavi, A. Edwards, N. Heess, Y. Chen, R. Hadsell, O. Vinyals, M. Bordbar, and N. de Freitas (2022) "A Generalist Agent," arXiv arXiv:2205.06175.
- Saadi, J.I. and M.C. Yang (2023) "Generative Design: Reframing the Role of the Designer in Early-Stage Design Process," *Journal of Mechanical Design*, Vol. 145, No. 4, article 041411.
- Serradilla, O., E. Zugasti, J. Rodriguez, and U. Zurutuza (2022) "Deep Learning Models for Predictive Maintenance: A Survey, Comparison, Challenges and Prospects," *Applied Intelligence*, Vol. 52, No. 10, pp. 10934–10964.
- Shazeer, N., A. Mirhoseini, K. Maziarz, A. Davis, Q.V. Le, G.E. Hinton, and J. Dean (2017) "Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer," in *Proceedings of the 5th International Conference on Learning Representations (ICLR)*. Available at: <https://arxiv.org/abs/1701.06538>.
- Shree, P. (2020) "The Journey of Open AI GPT Models," *Walmart Global Tech Blog*, November 9, available at: <https://medium.com/walmart-globaltech/the-journey-of-open-ai-gpt-model-s-32d95b7b7fb2>.
- Shumailov, I., Z. Shumaylov, Y. Zhao, Y. Gal, N. Papernot, and R. Anderson (2023) "The Curse of Recursion: Training on Generated Data Makes Models Forget," arXiv Preprint arXiv:2305.17493.
- Tao, F., H. Zhang, and C. Zhang (2024) "Advancements and Challenges of Digital Twins in Industry," *Nature Computational Science*, Vol. 4, No. 3, pp. 169–177.
- Taori, R., I. Gulrajani, T. Zhang, Y. Dubois, X. Li, C. Guestrin, P. Liang, and T.B. Hashimoto (2023) "Stanford Alpaca: An Instruction-following Llama Model," Stanford Center for Research on Foundation Models.
- Trajtenberg, M. (2018) "AI as the Next GPT: A Political-Economy Perspective," NBER Working Paper 24245, National Bureau of Economic Research.
- Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin (2017) "Attention is All You Need," *Advances in Neural Information Processing Systems*, Vol. 30.
- Wang, D. and A.-L. Barabási (2021) *The Science of Science*, Cambridge University Press.
- Wang, H., S. Ma, L. Dong, S. Huang, H. Wang, L. Ma, F. Yang, R. Wang, Y. Wu, and F. Wei (2023) "BitNet: Scaling 1-Bit Transformers for Large Language Models," arXiv Preprint arXiv:2310.11453.
- Webb, M. (2019) "The Impact of Artificial Intelligence on the Labor Market," Available at SSRN 3482150.
- Wei, J., X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou (2022) "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," *Advances in Neural Information Processing Systems*, Vol. 35, pp. 24824–24837.
- Whitehead, A.N. (1925) *Science and the Modern World: Lowell Lectures, 1925*, New American Library.
- Yin, M., H. Vu, and C. Persico (2026) "How (un)Stable Are LLM Occupational Exposure Scores? Evidence from Multi-Model Replication," NBER Working Paper w35110.
- Zeff, M. (2024) "Current AI Scaling Laws Are Showing Diminishing Returns, Forcing AI Labs to Change Course," *TechCrunch*, November 20, available at: <https://techcrunch.com/2024/11/20/ai-scaling-laws-are-showing-diminishing-returns-forcing-ai-labs-to-change-course/>.