

STANFORD UNIVERSITY

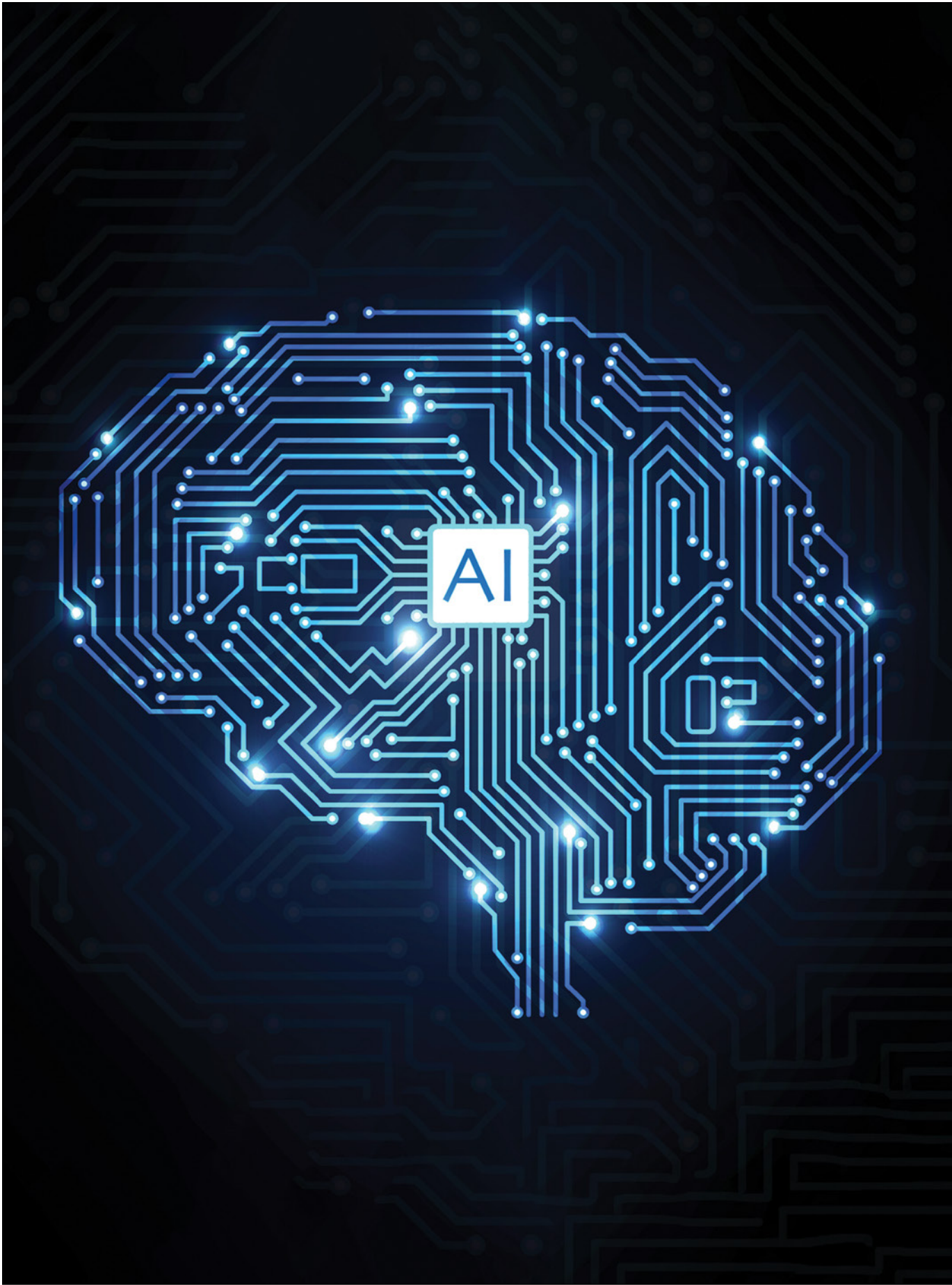
THE STANFORD EMERGING TECHNOLOGY REVIEW 2025

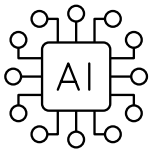
A Report on Ten Key Technologies and Their Policy Implications

CO-CHAIRS Condoleezza Rice, John B. Taylor, Jennifer Widom, and Amy Zegart

DIRECTOR AND EDITOR IN CHIEF Herbert S. Lin | **MANAGING EDITOR** Martin Giles







ARTIFICIAL INTELLIGENCE

KEY TAKEAWAYS

- Artificial intelligence (AI) is a foundational technology that is supercharging other scientific fields and, like electricity and the internet, has the potential to transform societies, economies, and politics worldwide.
- Despite rapid progress in the past several years, even the most advanced AI still has many failure modes that are unpredictable, not widely appreciated, not easily fixed, not explainable, and capable of leading to unintended consequences.
- Mandatory governance regimes for AI, even those to stave off catastrophic risks, will face stiff opposition from AI researchers and companies, but voluntary regimes calling for self-governance are more likely to gain support.

Overview

Artificial intelligence (AI), a term coined by computer scientist and Stanford professor John McCarthy in 1955, was originally defined as “the science and engineering of making intelligent machines.” In turn, intelligence might be defined as the ability to learn and perform suitable techniques to solve problems and achieve goals, appropriate to the context in an uncertain, ever-varying world.¹ AI could be said to refer to a computer’s ability to display this type of intelligence.

The emphasis today in AI is on machines that can learn as well as humans can learn, or at least somewhat comparably so. However, because machines are not limited by the constraints of human biology, AI systems may be able to run at much higher speeds and digest larger volumes and types of information than are possible with human capabilities.

Today, AI promises to be a fundamental enabler of technological advancement in many fields, arguably of comparable importance to electricity in an earlier era or the internet in more recent years. The science of computing, worldwide availability of networks, and civilization-scale data—all that collectively underlies the AI of today and tomorrow—are poised to have similar impact on technological progress in the future. Moreover, the users of AI will not be limited to those with specialized training; instead, the average person on the street will increasingly interact directly with sophisticated AI applications for a multitude of everyday activities.

The global AI market was worth \$196.63 billion in 2023, with North America receiving 30.9 percent of total AI revenues.² The Stanford Institute for Human-Centered Artificial Intelligence (HAI) *AI Index 2024 Annual Report* found that private investment in all AI start-ups totaled \$95.99 billion in 2023, marking the second consecutive year of decline since a record high of over \$120 billion in 2021.³ Amid a 42 percent fall in overall global venture funding across all sectors in 2023,⁴ AI start-ups raised \$42.5 billion in venture capital that year, marking only a 10 percent decrease⁵ from 2022.⁶

Many tech companies are significantly ramping up investments in AI infrastructure, such as larger and more powerful computing clusters to meet the growing demand for AI capabilities. Companies such as Amazon and Meta have begun revamping their data centers,⁷ and BlackRock, Microsoft, and the technology investor MGX, which is backed by the United Arab Emirates, announced in September 2024 the new Global AI Infrastructure Investment Partnership fund, which seeks to raise \$30 billion in private equity capital to finance data centers and other projects that span the AI infrastructure ecosystem.⁸ The fund may ultimately invest up to \$100 billion over time.⁹

One estimate forecasts that generative AI—which can create novel text, images, and audio output and is discussed in more detail later in this chapter—could raise global GDP by \$7 trillion and raise

productivity growth by 1.5 percent over a ten-year period if it is adopted widely.¹⁰ Private funding for generative AI start-ups surged to \$25.2 billion in 2023, a nearly ninefold increase from 2022, and accounted for around a quarter of all private investments related to AI in 2023.¹¹

The question of what subfields are considered part of AI is a matter of ongoing debate, and the boundaries between these fields are often fluid. Some of the core subfields are the following:

- Computer vision, enabling machines to recognize and understand visual information from the world, convert it into digital data, and make decisions based on these data
- Machine learning (ML), enabling computers to perform tasks without explicit instructions, often by generalizing from patterns in data. This includes deep learning that relies on multilayered artificial neural networks—which process information in a way inspired by the human brain—to model and understand complex relationships within data.
- Natural language processing, equipping machines with capabilities to understand, interpret, and produce spoken words and written texts

Most of today's AI is based on ML, though it draws on other subfields as well. ML requires data and computing power—often called compute¹²—and much of today's AI research requires access to these on an enormous scale.

In October 2024, the Royal Swedish Academy of Sciences awarded the Nobel Prize in Physics for 2024 to John Hopfield and Geoffrey Hinton for their work in applying tools and concepts from statistical mechanics to develop “foundational discoveries and inventions that enable machine learning with artificial neural networks”¹³ (further discussed below). Underscoring the importance of AI-based techniques in advancing science, it also awarded the

Nobel Prize in Chemistry for 2024 to Demis Hassabis and John M. Jumper for AI-based protein structure prediction,¹⁴ an important and long-standing problem in biology and chemistry involving the prediction of the three-dimensional shape a protein would assume given only the DNA sequence associated with it.

Machine learning also requires large amounts of data from which it can learn. These data can take various forms, including text, images, videos, sensor readings, and more. Learning from these data is called training the AI model.

The quality and quantity of data play a crucial role in determining the performance and capabilities of AI systems. Without sufficient and high-quality data, models may generate inaccurate or biased outcomes. (Roughly speaking, a traditional ML model is developed to solve a particular problem—different problems call for different models; for problems sufficiently different from each other, entirely new models need to be developed. Foundation models, discussed below, break this tradition to some extent.) Research continues on how to train systems incrementally, starting from existing models and using a much smaller amount of specially curated data to refine those models' performance for specialized purposes.

For a sense of scale, estimates of the data required to train GPT-4, OpenAI's large language model (LLM) released in March 2023 and the base on which some versions of ChatGPT were built, suggest that its training database consisted of the textual equivalent of around 100 million books, or about 10 trillion words, drawn from billions of web pages and scanned books. (LLMs are discussed further below.) The hardware requirements for computing power are also substantial. The costs to compute the training of GPT-4, for example, were enormous. Reports indicate that the training took about twenty-five thousand Nvidia A100 GPU deep-learning chips—at a cost of \$10,000 each—running for about one hundred days.¹⁵ Doing the math—and noting that other

hardware components were likely also needed—suggests the overall hardware costs for GPT-4 were at least a few hundred million dollars. And the chips underlying this hardware are specialty chips often fabricated offshore.¹⁶ (Chapter 8 on semiconductors discusses this point at greater length.)

Lastly, AI models consume a lot of energy. Consider first the training phase: One estimate of the electricity required to train a foundation model such as GPT-4 pegs the figure at about fifty million kilowatt-hours (kWh).¹⁷ The average American household uses about 11,000 kWh per year, meaning the energy needed to train GPT-4 was approximately the same as that used by 4,500 average homes in a year. Paying for this energy adds significant cost, even before a single person actually uses a model.

Then, once a model is up and running, the cost of energy used to power queries can add up fast. This is known as the inference phase. For ChatGPT, the energy used per query is around 0.002 of a kilowatt-hour, or 2 watt-hours.¹⁸ (For comparison, a single Google search requires about 0.3 watt-hours,¹⁹ and an alkaline AAA battery contains about 2 watt-hours of energy.) Given hundreds of millions of queries per day, the operating energy requirement of ChatGPT might be a few hundred thousand kilowatt-hours per day, at a cost of several tens of thousands of dollars.

AI can automate a wide range of tasks. But it also has particular promise in augmenting human capabilities and further enabling people to do what they are best at doing.²⁰ AI systems can work alongside humans, complementing and assisting their work rather than replacing them. Some present-day examples are discussed below.

Healthcare

- **Medical diagnostics** An AI system that can predict and detect the onset of strokes qualified for Medicare reimbursement in 2020.²¹

- **Drug discovery** An AI-enabled search identified a compound that inhibits the growth of a bacterium responsible for many drug-resistant infections, such as pneumonia and meningitis, by sifting through a library of seven thousand potential drug compounds for an appropriate chemical structure.²²
- **Patient safety** Smart AI sensors and cameras can improve patient safety in intensive care units, operating rooms, and even at home by improving healthcare providers' and caregivers' ability to monitor and react to patient health developments, including falls and injuries.²³
- **Robotic assistants** Mobile robots using AI can carry out healthcare-related tasks such as making specialized deliveries, disinfecting hospital wards, and assisting physical therapists, thus supporting nurses and enabling them to spend more time having face-to-face human interactions.²⁴

Agriculture

- **Production optimization** AI-enabled computer vision helps some salmon farmers pick out fish that are the right size to keep, thus off-loading the labor-intensive task of sorting them.²⁵
- **Crop management** Some farmers are using AI to detect and destroy weeds in a targeted manner, significantly decreasing environmental harm by using herbicides only on undesired vegetation rather than entire fields, in some cases reducing herbicide use by as much as 90 percent.²⁶

Logistics and Transportation

- **Resource allocation** AI enables some commercial shipping companies to predict ship arrivals five days into the future with high accuracy, thus allowing real-time allocations of personnel and schedule adjustments.²⁷

- **Autonomous trucking** Multiple companies collaborated in a consortium that arranged for trucks carrying tires to drive autonomously for over fifty thousand long-haul trucking miles in the period from January to August 2024.²⁸ If this and other demonstrations continue to be successful, it is possible that long-haul drives—the most boring and time-consuming aspect of a truck driver's job—can be automated; at the same time, aspects of such jobs requiring human-centered interactions, including navigating the first miles out of the factory and the last miles of delivering goods to customers, could be retained.

Law

- **Legal transcription** AI enables the real-time transcription of legal proceedings and client meetings with reasonably high accuracy, and some of these services are free of charge.²⁹
- **Legal review** AI-based systems can reduce the time lawyers spend on contract review by as much as 60 percent. Further, such systems can enable lawyers to search case databases more rapidly than online human searches—and even write case summaries.³⁰

Key Developments

Foundation Models

Foundation models dominated the conversation about AI in both 2023 and 2024. These models are large-scale systems trained on vast amounts of diverse data that can handle a variety of tasks.³¹ They often contain billions or trillions of parameters,³² and their massive size allows them to capture more complex patterns and relationships. Trained on these datasets, foundation models can develop broad capabilities³³ and are thus sometimes called general-purpose models. They excel at transfer

learning—applying knowledge learned in one context to another—making them more flexible and efficient than traditional task-specific models. A single foundation model is often fine-tuned for various tasks, reducing the need to train separate models from scratch.

These models are generally classified as closed source or open source. A closed-source model is a proprietary one developed and maintained by a specific organization, usually a for-profit company, with its source code, data, and architecture kept confidential. Access to these models is typically restricted through technically enforced usage permissions, such as application programming interfaces, allowing the developers to control the model's distribution, usage, and updates. By contrast, an open-source model is one whose code, data, and underlying architecture are publicly accessible, allowing anyone to use, modify, and distribute it freely.

The most familiar type of foundation model is an LLM—a system trained on very large volumes of textual content. LLMs are an example of generative AI, a type of AI that can produce new material based on how it has been trained and the inputs it is given. Models trained on text can generate new text based on a statistical analysis that makes predictions about what other words are likely to be found immediately after the occurrence of certain words.

These models do not think or feel like humans do, even though their responses may make it seem like they do. Instead, LLMs use statistical analysis based on training data. For example, because the word sequence “thank you” is far more likely to occur than “thank zebras,” a person’s query to an LLM asking it to draft a thank-you note to a colleague is unlikely to generate the response “thank zebras.”

These models generate linguistic output surprisingly similar to that of humans across a wide range of subjects. For example, LLMs can generate useful computer code, poetry, legal case summaries, and

medical advice, and they outscore the median human performance on clinical examination in obstetrics and gynecology,³⁴ on standardized tests of divergent thinking,³⁵ and on other standardized tests such as the LSAT, sections of the GRE, and various AP exams.³⁶ However, models do not necessarily excel at the actual tasks or skills that these tests are trying to capture and, as discussed below, still produce errors and fail in all sorts of other ways, many of them unexpected.

Well-known closed-source LLMs include OpenAI’s GPT models (e.g., GPT-3, GPT-3.5, and GPT-4), Anthropic’s Claude, and Google’s Gemini. Well-known open-source LLMs include Meta’s Llama, Big Science’s BLOOM, EleutherAI’s GPT-J, and Google’s BERT and T5.

Specialized foundation models have also been developed in other modalities such as audio, video, and images:

- Foundation models for images are able to generate new images based on a user’s text input. Novel methods for handling images, combined with using very large collections of pictures and text for training, have led to models that can turn written descriptions into images that are quickly becoming comparable to—and sometimes indistinguishable from—real-life photographs and artwork created by humans. Examples include OpenAI’s DALL-E 3, the open-source Stable Diffusion, Google’s Imagen, Adobe Firefly, and Meta’s Make-A-Scene.
- An example of a foundation model for audio is UniAudio, which handles all audio types and employs predictive algorithms to generate high-quality speech, sound, and music, surpassing leading methods in tasks such as text to speech, speech enhancement, and voice conversion.
- Foundation models in video such as Meta’s Emu Video represent a significant advancement in

video generation. Emu first generates an image from text input and then creates a video based on both the text and the generated image. Emu Video has demonstrated superior performance over previous state-of-the-art methods in terms of image quality, faithfulness to text instructions, and evaluations from humans.

Multimodal Models

AI systems that incorporate multiple modalities—text, images, and sound—within single models are becoming increasingly popular. This multimodal approach, shown in figure 1.1, aims to create more humanlike experiences by leveraging various senses such as sight, speech, and hearing to mirror how humans interact with the world.

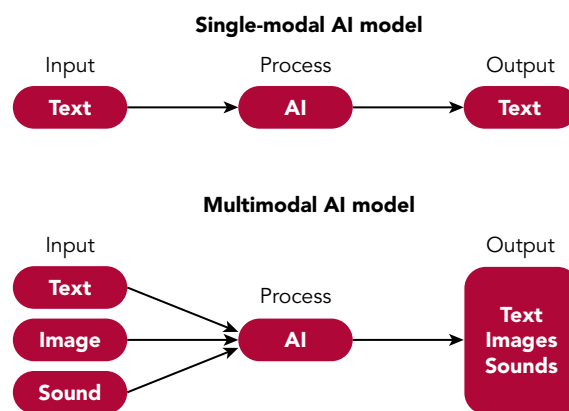
Multimodal AI systems have diverse applications across sectors. They can enhance accessibility for people with disabilities through real-time transcription, sign language translation, and detailed image descriptions. They can also eliminate language barriers via cost-effective, near-real-time translation services. In education, multimodal AI can support personalized learning by adapting content

to various formats and learner types, improving engagement and comprehension. When integrated with virtual and augmented reality, it can create immersive, highly realistic training environments that are particularly valuable in fields like healthcare. The advent of multimodal AI is also set to further transform human-computer interactions, enabling more intuitive communication and expanding the range of tasks that AI systems can handle.

Embodied AI

Embodied AI involves integrating AI systems into robots or other physical devices. This approach aims to bridge the gap between the digital and physical realms. Embodied AI has the potential to enhance robotic capabilities and expand the range of interactions robots have with the physical world. These robot-plus-AI systems could potentially address knowledge tasks, physical tasks, or combinations of both. (This topic is explored further in chapter 7 on robotics.) As research progresses in AI autonomy and reasoning, embodied AI systems may be able to handle increasingly complex tasks with greater independence. This could lead to applications in various fields such as logistics and domestic assistance.

FIGURE 1.1 Multimodal AI systems can transform one type of input into a different type of output



The advent of multimodal AI is . . . set to further transform human-computer interactions, enabling more intuitive communication and expanding the range of tasks that AI systems can handle.

Existential Concerns About AI

LLMs have generated considerable attention because of their apparent sophistication. Indeed, their capabilities have led some to suggest that they are the initial sparks of artificial general intelligence (AGI).³⁷ AGI is AI that is capable of performing any intellectual task that a human can perform, including learning. But, according to this argument, because an electronic AGI would run on electronic circuits rather than biological ones, it is likely to learn much faster than biological human intelligences—rapidly outstripping their capabilities.

The belief in some quarters that AGI will soon be achieved has led to substantial debate about its risks. Scholars have continued to argue over the past year about whether current models present initial sparks of AGI,³⁸ although there hasn't been substantial evidence presented that proves they possess such capabilities.

Others suggest that focusing on low-probability doomsday scenarios distracts from the real and immediate risks AI poses today.³⁹ Instead, society should be prioritizing efforts to address the harms that AI systems are already causing, like biased decision-making, hallucinations (error-ridden responses that appear to provide accurate information), and job displacement. Those who support this view argue that these problems are the ones on which governments and regulators should be concentrating their efforts.

A National AI Research Resource

LLMs such as GPT-4, Claude, Gemini, and Llama can be developed only by large companies with the resources to build and operate very large data and compute centers. For a sense of scale, Princeton University announced in March 2024 that it would dip into its endowment to purchase 300 advanced Nvidia chips to use for research at a total estimated cost of about \$9 million.⁴⁰ By contrast, Meta announced at the start of 2024 that it intended to purchase 350,000 such chips by the end of the year⁴¹—over one thousand times as many chips as Princeton and with a likely price tag of nearly \$10 billion.

Traditionally, academics and others in civil society have undertaken research to understand the potential societal ramifications of AI, but with large companies controlling access to these AI systems, they can no longer do so independently. In July 2023, a bipartisan bill (S.2714, the CREATE AI Act of 2023)⁴² was proposed to establish the National Artificial Intelligence Research Resource (NAIRR) as a shared national research infrastructure that would provide civil society researchers greater access to the complex resources, data, and tools needed to support research on safe and trustworthy AI. The bill's text did not mention funding levels, but the final NAIRR task force report, released in January 2023, indicated that NAIRR should be funded at a level of \$2.6 billion over its initial six-year span.⁴³ In January 2024, the National Science Foundation established the NAIRR pilot to establish proof of concept for the full-scale NAIRR.

As a point of comparison to the fledgling NAIIR effort, investments from high-tech companies for AI exceeded \$27 billion in 2023 alone.⁴⁴

Over the Horizon

Impact of New AI Technologies

Potential positive impacts of new AI technologies are most likely to be seen in the applications they enable for societal use, as described in detail above. On the other hand, no technology is an unalloyed good. Potential negative impacts from AI will likely emerge from known problems with current state-of-the-art AI and from technical advances in the future. Some of the known issues with today's leading AI models include the following:

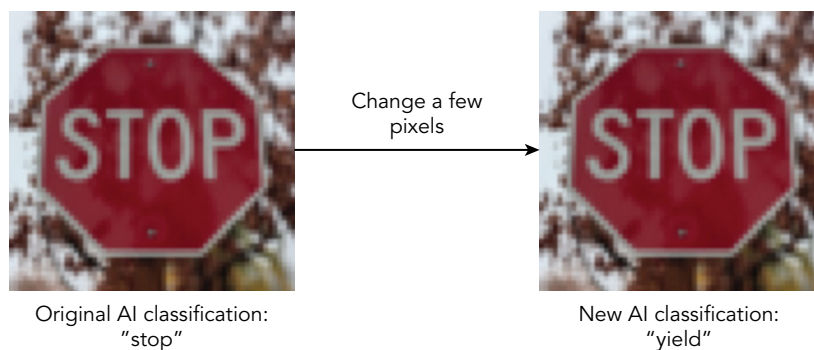
Explainability This is the ability to explain the reasoning behind—and describe the data underlying—an AI system's conclusions. Today's AI is largely incapable of explaining the basis on which it arrives at any particular conclusion. Explanations are not

always relevant, but in certain cases, such as medical decision-making, they may be critical so that users can have confidence in an AI system's output.

Bias and fairness Because ML models are trained on existing datasets, they are likely to encode any biases present in these datasets. (Bias should be understood here as a property of the data that is commonly regarded as societally undesirable.) For example, if a facial recognition system is primarily trained on images of individuals from one ethnic group, its accuracy at identifying people from other ethnic groups may be reduced.⁴⁵ Use of such a system could well lead to disproportionate singling out of individuals in those other groups. To the extent that these datasets reflect historical approaches, they will also reflect the biases embedded in that history, and an ML model based on such datasets will also reflect these biases.

Vulnerability to spoofing It is possible to tweak data inputs to fool many AI models into drawing false conclusions. For example, in figure 1.2, changing a small number of pixels in a visual image of a traffic stop sign can lead to its being classified as

FIGURE 1.2 Changing a few pixels can fool AI into thinking a picture of a stop sign is a picture of a yield sign



Source: Derived from figure 1 in Fabio Carrara, Fabrizio Falchi, Giuseppe Amato, Rudy Becarelli, and Roberto Caldelli, "Detecting Adversarial Inputs by Looking in the Black Box," in "Transparency in Algorithmic Decision Making," special issue, *ERCIM News* 116 (January 2019): 16–19.

a yield sign, even though this fuzzing of the image is invisible to the naked eye. That example seems innocuous, but as AI models are used increasingly in applications from medical treatment to intelligence and military operations, the potential harms could be substantial. It is also possible that an attack targeting one AI model could work against other models performing the same task—a phenomenon known as transferability. One study reports that as often as 80 percent of the time, transferability allows attackers to create an attack on a surrogate model and then apply it to their intended target, too.⁴⁶

Data poisoning An attacker manipulating the dataset used to train an ML model can damage its performance and even create predictable errors.

Deepfakes AI provides the capability for generating highly realistic but entirely inauthentic audio and video imagery. This has obvious implications for evidence presented in courtrooms and for efforts to manipulate political contests. In September 2023, just before elections took place in Slovakia, a deepfake audio was posted to Facebook in which a candidate was heard discussing with a journalist how to rig the

election by buying votes.⁴⁷ In January 2024, voters in New Hampshire received robocalls that used a voice sounding like President Biden’s telling them not to vote in the state’s presidential primary.⁴⁸ In elections in India in early 2024, deepfake videos were used to depict deceased politicians as though they were still alive (see figure 1.3).⁴⁹ All of these deepfakes are much more sophisticated than attempts such as the “dumbfake” video of Representative Nancy Pelosi (D-CA) that involved merely slowing down an existing video of her to make her look drunk.⁵⁰

Privacy Many LLMs are trained on data found on the internet rather indiscriminately, and such data may include personal information of individuals. When incorporated into LLMs, this information could be publicly disclosed more often.

Overtrust and overreliance If AI systems become commonplace in society, their novelty will inevitably diminish for users. The level of trust in computer outputs often increases with familiarity. But skepticism about answers received from a system is essential if one is to challenge the correctness of these outputs. As trust in AI grows, reducing skepticism, there’s a

FIGURE 1.3 Deepfake videos of deceased Indian politicians speaking as if they were alive were used in India’s 2024 elections



Photo from the late Indian Congress leader H. Vasanthakumar’s funeral in 2020



Screenshot from a deepfake video of H. Vasanthakumar endorsing his son’s parliamentary candidacy in 2024

Source: (Left) PTI Photo / R. Senthil Kumar; (right) “H Vasantha Kumar,” posted April 16, 2024, by Vasanth TV, YouTube, https://www.youtube.com/watch?v=98_K-Ag7p2M

higher risk that errors, mishaps, and unforeseen incidents will be overlooked. One recent experiment showed that developers with access to an AI-based coding assistant wrote code that was significantly less secure than those without an AI-based assistant—even though the former were more likely to believe they had written secure code.⁵¹

Hallucinations As noted earlier, AI hallucinations refer to situations where an AI model generates results or answers that are plausible but do not correspond to reality. In other words, models can simply make things up, but human users will not be aware they have done this. The results are plausible because they are constructed based on statistical patterns that the model has learned to recognize from its training data. But they may not correspond to reality because the model does not have an understanding of the real world. For example, in September 2024, a Stanford professor asked an AI model to name ten publications she had written. The AI responded with five correct publications and five that she had never actually written—but the AI results included titles and summaries that made them seem real. When she told the model that “the last two entries are hallucinations,” it simply provided two new results that were also hallucinations.

Out-of-distribution inputs All ML systems must be trained on a large volume of data. If the inputs subsequently given to a system are substantially different from the training data—a situation known as being out-of-distribution—the system may draw conclusions that are more unreliable than if the inputs were similar to the training data.

Copyright violations Some AI-based models have been trained on large volumes of data found online. These data have generally been used without the consent or permission of their owners, thereby raising important questions about appropriately compensating and acknowledging those owners. For example, in January 2023, Getty Images sued Stability AI in an English court for infringing on the copyrights of millions of photographs, their associated captions,

and metadata in building and offering the products Stable Diffusion (an application that generates images from text) and DreamStudio (the app that serves as a user interface to Stable Diffusion).⁵² In late 2023, the *New York Times* sued OpenAI and Microsoft over their alleged use of millions of articles published by the *Times* to train the companies’ LLMs.⁵³ In June 2024, music labels Sony Music, Universal Music Group, and Warner Records sued AI start-ups Suno and Udio for copyright infringement, alleging that the companies had trained their music-generation systems on protected content.⁵⁴

AI researchers are cognizant of issues such as these, and in many cases work has been done—or is being done—to develop corrective measures. However, in most cases, these defenses don’t apply very well to instances beyond the specific problems that they were designed to solve.

Challenges of Innovation and Implementation

The primary challenge of bringing AI innovation into operation is risk management. It is often said that AI, and especially ML, brings a new conceptual paradigm for how systems can exploit information to gain advantage, relying on pattern recognition in the broadest sense rather than on explicit understanding of situations that are likely to occur. Because there have been significant recent advances in AI, the people who would make decisions to deploy AI-based systems may not have a good understanding of the risks that could accompany such deployment.

Consider, for example, AI as an important approach for improving the effectiveness of military operations. Despite broad agreement by the military services and the US Department of Defense (DOD) that AI would be of great benefit, the actual integration of AI-enabled capabilities into military forces has proceeded at a slow pace. Certainly, it is well understood that technical risks of underperformance and error in new technologies take time to mitigate.

But another important reason for the slow pace is that the DOD acquisition system has largely been designed to minimize the likelihood of programmatic failure, fraud, unfairness, waste, and abuse—in short, to minimize risk. It is therefore not surprising that the incentives at every level of the bureaucracy are aligned in that manner. For new approaches like AI to take root, a greater degree of programmatic risk acceptance may be necessary, especially in light of the possibility that other nations could adopt the technology faster, achieving military advantages over US forces.

Policy, Legal, and Regulatory Issues

THE FUTURE OF WORK

Some researchers expect that, within the next five to ten years, more and more workers will have AI added to their workflows to enhance productivity or will even be replaced by AI systems, which may cause significant disruptions to the job market in the near future.⁵⁵ LLMs have already demonstrated how they can be used in a wide variety of fields, including law, customer support, coding, and journalism. These demonstrations have led to concerns that the impact of AI on employment will be substantial, especially on jobs that involve knowledge work. However, uncertainty abounds. What and how many present-day jobs will disappear? Which tasks could best be handled by AI? And what new jobs might be created by the technology today and in the future?

Some broad outlines and trends are clear:

- Individuals whose jobs entail routine white-collar work may be more affected than those whose jobs require physical labor; some will experience painful shifts in the short term.⁵⁶
- AI is helping some workers to increase their productivity and job satisfaction.⁵⁷ At the same time, other workers are already losing their jobs as AI demonstrates adequate competence for business operations, despite potentially underperforming

the humans it replaces.⁵⁸ In at least some cases, companies are deciding that the cost savings of eliminating human workers outweigh the drawbacks of mediocre AI performance.

- Training displaced workers to be more competitive in an AI-enabled economy does not solve the problem if new jobs are not available. The nature and extent of new roles resulting from widespread AI deployment are not clear at this point, although historically the introduction of new technologies has not resulted in a long-term net loss of jobs.⁵⁹

GOVERNANCE AND REGULATION OF AI

Governments around the world have been increasingly focused on establishing regulations and guidelines for AI. Research on foundational AI technologies is difficult to regulate across international boundaries even among like-minded nations, especially when other nations have strong incentives to carry on regardless of actions taken by US policymakers. It is even more difficult, and may well be impossible, to reach agreement between nations that regard each other as strategic competitors and adversaries. The same applies to voluntary restrictions on research by companies concerned about competition from less constrained foreign rivals. Regulation of specific applications of AI may be more easily implemented, in part because of existing regulatory frameworks in domains such as healthcare, finance, and law.

The most ambitious attempt to regulate AI came into force in August 2024 with the European Union's AI Act. This forbids certain applications of AI, such as individual predictive policing based solely on a person's data profile or tracking of their emotional state in the workplace and educational institutions, unless for medical or safety reasons.⁶⁰ Additionally, it imposes a number of requirements on what the AI Act calls "high-risk" systems. (The legislation provides a very technical definition of such systems, but generally they include those that could pose a significant risk to health, safety, or fundamental rights.)

The resources needed to train GPT-4 far exceed those available through grants or any other sources to any reasonably sized group of the top US research universities.

These requirements address data quality, documentation and traceability, transparency and explainability, human oversight, accuracy, cybersecurity, and robustness.

In the United States, the president's Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence was issued on October 30, 2023.⁶¹ The order addressed actions to advance AI safety and security; privacy; equity and civil rights; consumer, patient, student, and worker interests; the promotion of innovation and competition, as well as American leadership; and government use of AI. Of particular note is the order's requirement that developers of advanced AI systems posing a serious risk to national security, national economic security, or national public health and safety inform the US government when training them and share with it all results from internal safety testing conducted by red teams. (A red team is a team of experts that attempts to subvert or break the system it is asked to test. It then reports its findings to the owner of the system so that the owner can take corrective action.) The order also requires government actions to develop guidance to help protect against the use of AI to develop biological threats and to advance the use of AI to protect against cybersecurity threats, to help detect AI-generated content, and to authenticate official content.

At the state level in the United States, an attempt to pass an AI regulatory bill in California (SB 1047, the Safe and Secure Innovation for Frontier Artificial Intelligence Models Act) was vetoed by the governor

in September 2024. The act sought to hold the creators of advanced AI models liable in civil court for causing catastrophic harms unless they had taken certain advance measures to forestall such an outcome. Opposition to the bill was based on concerns about a technologically deficient definition of advanced AI models, the burden that the bill would place on small start-ups and academia, and the unfairness of holding model developers responsible for harmful applications that others build using the developers' models.

Other important developments regarding AI governance include the AI Safety Summit, held on November 1–2, 2023, at Bletchley Park in the United Kingdom,⁶² which issued the Bletchley Declaration, and the AI Seoul Summit of May 2024. In the Bletchley Declaration, the European Union and twenty-eight nations collectively endorsed international cooperation to manage risks associated with highly capable general-purpose AI models. Signatories committed to ensuring that AI systems are developed and deployed safely and responsibly. The summit also led to the establishment of the United Kingdom's AI Safety Institute and the US Artificial Intelligence Safety Institute, located within the National Institute of Standards and Technology.

The Seoul Declaration from the AI Seoul Summit 2024 built on the Bletchley Declaration to acknowledge the importance of interoperability between national AI governance frameworks to maximize benefits and minimize risks from advanced AI systems. In addition, sixteen major AI organizations

agreed on the Frontier AI Safety Commitments, a set of voluntary guidelines regarding the publication of safety frameworks for frontier AI models and the setting of thresholds for intolerable risks, among other things.

NATIONAL SECURITY

AI is expected to have a profound impact on militaries worldwide.⁶³ Weapons systems, command and control, logistics, acquisition, and training will all seek to leverage multiple AI technologies to operate more effectively and efficiently, at lower cost and with less risk to friendly forces. Trying to overcome decades of institutional inertia, the DOD is dedicating billions of dollars to institutional reforms and research advances aimed at integrating AI into its warfighting and war preparation strategies. Senior military officials recognize that failure to adapt to the emerging opportunities and challenges presented by AI would pose significant national security risks, particularly considering that both Russia and China are heavily investing in AI capabilities.

In adopting a set of guiding principles that address responsibility, equity, traceability, reliability, and governability in and for AI,⁶⁴ the DOD has taken an important first step in meeting its obligation to proceed ethically with the development of AI capabilities; eventually, these principles will have to be operationalized in specific use cases. An additional important concern, subsumed under these principles but worth calling out, is determining where the use of AI may or may not be appropriate—for example, whether AI is appropriate in nuclear command and control. The United States, the United Kingdom, and France have made explicit commitments to maintain human control over nuclear weapons.⁶⁵

Meanwhile, other countries are also adopting AI, and nations such as Russia and China are unlikely to make the same operational and ethical decisions as Western countries about the appropriate roles of AI vis-à-vis humans in controlling the operation

of weapons or in making decisions about the use of deadly force. Notably, in late 2023, press reports indicated that President Biden and Chinese President Xi Jinping had considered entering into a dialogue about AI in nuclear command and control, but such an arrangement was never formalized.⁶⁶

TALENT

The United States is eating its seed corn with respect to the AI talent pool. As noted in the Foreword, faculty at Stanford and other universities report that the number of students studying in AI who are joining the industry, particularly start-ups, is increasing at the expense of those pursuing academic careers and contributing to foundational AI research.

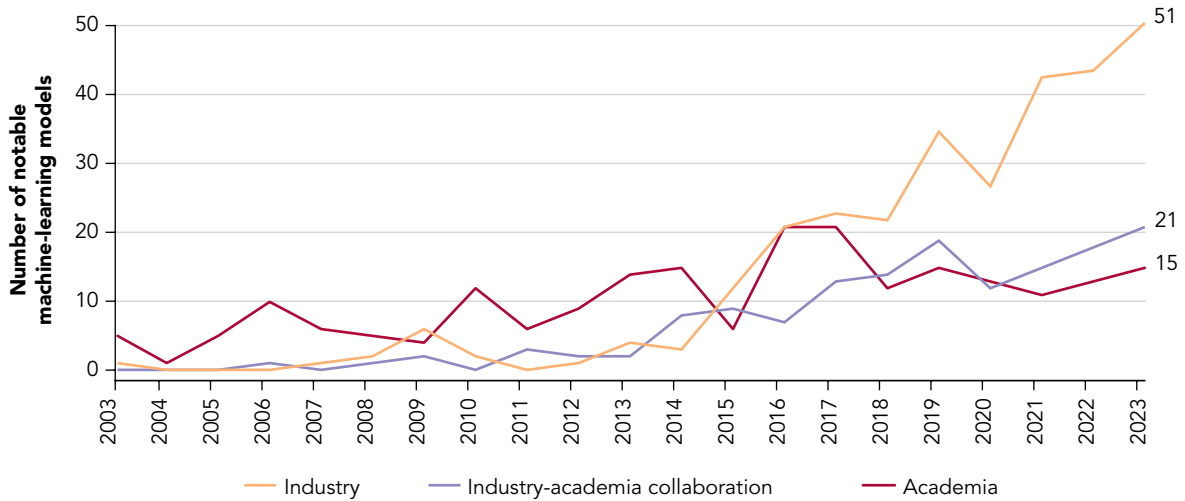
Many factors are contributing to this trend. One is that industry careers come with compensation packages that far outstrip those offered by academia. Academic researchers must also obtain funding to pay for research equipment, computing capability, and personnel like staff scientists, technicians, and programmers. This involves searching for government grants, which are typically small compared to what large companies might be willing to invest in their own researchers. Consider, for example, that the resources needed to build and train GPT-4 far exceed those available through grants or any other sources to any reasonably sized group of the top US research universities, let alone any single university.

Industry often makes decisions more rapidly than government grant makers and imposes fewer regulations on the conduct of research. Large companies are at an advantage because they have research-supporting infrastructure in place, such as compute facilities and data warehouses.

One important consequence is that academic access to research infrastructure is limited, so US-based students are unable to train on state-of-the-art systems—at least this is the case if their universities do not have access to the facilities of the corporate sector.

FIGURE 1.4 Most notable machine-learning models are now released by industry

Number of notable machine-learning models by sector, 2003–23



Source: Adapted from Nestor Maslej, Loredana Fattorini, Raymond Perrault, et al., *The AI Index 2024 Annual Report*, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024. Data from Epoch, 2023

Figure 1.4 shows that most notable ML systems are now released by industry, while very few are released by academic institutions.

At the same time, China’s efforts to recruit top scientific talent offer further temptations for scientists to leave the United States. These efforts are often targeted toward ethnic Chinese in the US—ranging from well-established researchers to those just finishing graduate degrees—and offer recruitment packages that promise benefits comparable to those available from private industry, such as high salaries, lavish research funding, and apparent freedom from bureaucracy.

All of these factors are leading to an AI “brain drain” that does not favor the US research enterprise.

NOTES

1. Christopher Manning, “Artificial Intelligence Definitions,” Institute for Human-Centered AI, Stanford University, September 2020, <https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf>.
2. Grand View Research, “Artificial Intelligence Market Size, Share, Growth Report 2024–2030,” accessed September 23, 2024, <https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market>.
3. Nestor Maslej, Loredana Fattorini, Raymond Perrault, et al., *The AI Index 2024 Annual Report*, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024, https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf.
4. CB Insights, “State of Venture 2023 Report,” February 1, 2024, <https://www.cbinsights.com/research/report/venture-trends-2023/>.
5. Reputable sources disagree as to whether global venture funding for AI start-ups increased or decreased in 2023. We have followed the majority view.
6. CB Insights, “State of AI 2023 Report,” February 1, 2024, <https://www.cbinsights.com/research/report/ai-trends-2023/>.
7. Karen Weise, “In Race to Build A.I., Tech Plans a Big Plumbing Upgrade,” *New York Times*, April 27, 2024, <https://www.nytimes.com/2024/04/27/technology/ai-big-tech-spending.html>.

8. Jack Pitcher and Connor Hart, "BlackRock, Microsoft, Others Form AI and Energy Infrastructure Investment Partnership," *Wall Street Journal*, September 17, 2024, <https://www.wsj.com/tech/ai/blackrock-global-infrastructure-partners-microsoft-mgx-launch-ai-partnership-1d00e09f>.
9. BlackRock, "BlackRock, Global Infrastructure Partners, Microsoft, and MGX Launch New AI Partnership to Invest in Data Centers and Supporting Power Infrastructure," September 17, 2024, <https://ir.blackrock.com/news-and-events/press-releases/press-releases-details/2024/BlackRock-Global-Infrastructure-Partners-Microsoft-and-MGX-Launch-New-AI-Partnership-to-Invest-in-Data-Centers-and-Supporting-Power-Infrastructure/default.aspx>.
10. Goldman Sachs, "Generative AI Could Raise Global GDP by 7%," April 5, 2023, <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>.
11. Maslej et al., *The AI Index 2024 Report*.
12. Jafar Alzubi, Anand Nayyar, and Akshi Kumar, "Machine Learning from Theory to Algorithms: An Overview," *Journal of Physics: Conference Series* 1142, Second National Conference on Computational Intelligence (December 2018), <https://doi.org/10.1088/1742-6596/1142/1/012012>.
13. The Nobel Prize in Physics 2024, "Summary," The Nobel Prize, October 12, 2024, <https://www.nobelprize.org/prizes/physics/2024/summary/>.
14. The Nobel Prize in Chemistry 2024, "Summary," The Nobel Prize, October 12, 2024, <https://www.nobelprize.org/prizes/chemistry/2024/summary/>.
15. Kif Leswing, "Meet the \$10,000 Nvidia Chip Powering the Race for A.I.," CNBC, February 23, 2023, <https://www.cnbc.com/2023/02/23/nvidias-a100-is-the-10000-chip-powering-the-race-for-ai-.html>; Kasper Groes Albin Ludvigsen, "The Carbon Footprint of GPT-4," Medium, July 18, 2023, <https://towardsdatascience.com/the-carbon-footprint-of-gpt-4-d6c676eb21ae>.
16. Darian Woods and Adrian Ma, "The Semiconductor Founding Father," December 21, 2021, in *The Indicator from Planet Money*, podcast produced by NPR, MP3 audio, 10:14, <https://www.npr.org/transcripts/1066548023>.
17. Ludvigsen, "The Carbon Footprint."
18. Kasper Groes Albin Ludvigsen, "ChatGPT's Electricity Consumption," Medium, July 12, 2023, <https://towardsdatascience.com/chatgpts-electricity-consumption-7873483feac4>. Different sources provide somewhat different numbers for the energy cost per query, but they all are in the range of a few watt-hours.
19. EPRI, "Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption," Technology Innovation, accessed September 23, 2023, <https://www.epri.com/research/products/3002028905>.
20. Hope Reese, "A Human-Centered Approach to the AI Revolution," Institute for Human-Centered AI, Stanford University, October 17, 2022, <https://hai.stanford.edu/news/human-centered-approach-ai-revolution>.
21. Viz.ai, "Viz.ai Receives New Technology Add-on Payment (NTAP) Renewal for Stroke AI Software from CMS," August 4, 2021, <https://www.viz.ai/news/ntap-renewal-for-stroke-software>.
22. Gary Liu, Denise B. Catacutan, Khushi Rathod, et al., "Deep Learning-Guided Discovery of an Antibiotic Targeting *Acinetobacter baumannii*," *Nature Chemical Biology* (2023), <https://doi.org/10.1038/s41589-023-01349-8>.
23. Albert Haque, Arnold Milstein, and Fei-Fei Li, "Illuminating the Dark Spaces of Healthcare with Ambient Intelligence," *Nature* 585 (2020): 193–202, <https://doi.org/10.1038/s41586-020-2669-y>.
24. Khari Johnson, "Hospital Robots Are Helping Combat a Wave of Nurse Burnout," *Wired*, April 19, 2022, <https://www.wired.com/story/moxi-hospital-robot-nurse-burnout-health-care>.
25. Fish Site, "Innovasea Launches AI-Powered Biomass Camera for Salmon," August 17, 2023, <https://thefishsite.com/articles/innovasea-launches-ai-powered-biomass-camera-for-salmon>.
26. Itransition, "Machine Learning in Agriculture: Use Cases and Applications," February 1, 2023, <https://www.itransition.com/machine-learning/agriculture>.
27. GateHouse Maritime, "Vessel Tracking Giving Full Journey Visibility," accessed August 15, 2023, <https://gatehousemaritime.com/data-services/vessel-tracking>.
28. Kodiak, "J.B. Hunt, Bridgestone and Kodiak Surpass 50,000 Autonomous Long-Haul Trucking Miles in Delivery Collaboration," August 7, 2024, <https://kodiak.ai/news/jb-hunt-and-kodiak-collaborate>.
29. JD Supra, "Artificial Intelligence in Law: How AI Can Reshape the Legal Industry," September 12, 2023, <https://www.jdsupra.com/legalnews/artificial-intelligence-in-law-how-ai-8475732>.
30. Steve Lohr, "A.I. Is Doing Legal Work. But It Won't Replace Lawyers, Yet," *New York Times*, March 19, 2017, <https://www.nytimes.com/2017/03/19/technology/lawyers-artificial-intelligence.html>.
31. Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, et al., "On the Opportunities and Risks of Foundation Models," arXiv, Stanford University, July 12, 2022, <https://doi.org/10.48550/arXiv.2108.07258>.
32. Parameters are like the building blocks of knowledge that make up an AI system's understanding. You can think of them as tiny bits of information the AI uses to make sense of data and generate responses. When we say AI models have billions or trillions of parameters, it means they have an enormous number of these information pieces to work with. This allows them to understand and generate more sophisticated and nuanced content.
33. Bommasani et al., "On the Opportunities and Risks."
34. Sarah W. Li, Matthew W. Kemp, Susan J. S. Logan, Sebastian E. Illanes, and Mahesh A. Choolani, "ChatGPT Outscored Human Candidates in a Virtual Objective Structured Clinical Examination in Obstetrics and Gynecology," *American Journal of Obstetrics & Gynecology* 229, no. 2 (August 2023): 172.E1-172.E12, <https://doi.org/10.1016/j.ajog.2023.04.020>.
35. Kent F. Hubert, Kim N. Awa, and Darya L. Zabelina, "The Current State of Artificial Intelligence Generative Language Models Is More Creative Than Humans on Divergent Thinking Tasks," *Scientific Reports* 14, no. 3440 (February 2024), <https://doi.org/10.1038/s41598-024-53303-w>.
36. Josh Achiam, Steven Adler, Sandhini Agarwal, et al., "GPT-4 Technical Report," arXiv, March 4, 2024, <https://doi.org/10.48550/arXiv.2303.08774>.
37. Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, et al., "Sparks of Artificial General Intelligence: Early Experiments with GPT-4," arXiv, Cornell University, April 13, 2023, <https://doi.org/10.48550/arXiv.2303.12712>.
38. For an argument that they are, see Bubeck et al., "Sparks of Artificial General Intelligence"; for an argument that they are not, see Thomas Macaulay, "Meta's AI Chief: LLMs Will Never Reach

- Human-Level Intelligence," *The Next Web*, April 10, 2024, <https://thenextweb.com/news/meta-yann-lecun-ai-behind-human-intelligence>.
39. "Stop Talking about Tomorrow's AI Doomsday When AI Poses Risks Today," editorial, *Nature* 618 (June 2023): 885–86, <https://www.nature.com/articles/d41586-023-02094-7>.
40. Poornima Apte, "Princeton Invests in New 300-GPU Cluster for Academic AI Research," *AI at Princeton*, Princeton University, March 15, 2024, <https://ai.princeton.edu/news/2024/princeton-invests-new-300-gpu-cluster-academic-ai-research>.
41. Tae Kim, "Mark Zuckerberg Says Meta Will Own Billions Worth of Nvidia H100 GPUs by Year End," *Barrons*, January 19, 2024, <https://www.barrons.com/articles/meta-stock-price-nvidia-zuckerberg-b0632fed>.
42. The bill's full name is Creating Resources for Every American to Experiment with Artificial Intelligence Act of 2023. *Congress.gov*, "S.2714 – 118th Congress (2023–2024): CREATE AI Act of 2023," July 27, 2023, <https://www.congress.gov/bill/118th-congress/senate-bill/2714>.
43. White House, National Artificial Intelligence Research Resource Task Force Releases Final Report, Office of Science and Technology Policy, News & Updates, Press Releases, <https://www.whitehouse.gov/ostp/news-updates/2023/01/24/national-artificial-intelligence-research-resource-task-force-releases-final-report/>.
44. Karen Kwok, "AI Firms Lead Quest for Intelligent Business Model," *Reuters*, December 12, 2023, <https://www.reuters.com/breakingviews/ai-firms-lead-quest-intelligent-business-model-2023-12-12/>.
45. Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Proceedings of Machine Learning Research* 81, Conference on Fairness, Accountability, and Transparency (February 2018): 1–15, <https://www.media.mit.edu/publications/gender-shades-intersectional-accuracy-disparities-in-commercial-gender-classification>.
46. Nicolas Papernot, Patrick McDaniel, and Ian Goodfellow, "Transferability in Machine Learning: From Phenomena to Black-Box Attacks Using Adversarial Samples," *arXiv*, May 24, 2016, <https://doi.org/10.48550/arXiv.1605.07277>.
47. Ivana Kottasová, Sophie Tanno, and Heather Chen, "Pro-Russian Politician Wins Slovakia's Parliamentary Election," *CNN*, October 2, 2023, <https://www.cnn.com/2023/10/01/world/slovakia-election-pro-russia-robot-fico-win-intl-hnk/index.html>.
48. Jeongyoon Han, "New Hampshire Is Investigating a Robocall That Was Made to Sound Like Biden," *NPR*, January 22, 2024, <https://www.npr.org/2024/01/22/1226129926/nh-primary-biden-ai-robocall>.
49. Samriddhi Sakunia, "AI and Deepfakes Played a Big Role in India's Elections," *New Lines Magazine*, July 12, 2024, <https://newlinesmag.com/spotlight/ai-and-deepfakes-played-a-big-role-in-indias-elections/>.
50. *Reuters*, "Fact Check: 'Drunk' Nancy Pelosi Video Is Manipulated," August 3, 2020, <https://www.reuters.com/article/world/fact-check-drunk-nancy-pelosi-video-is-manipulated-idUSKCN24Z2B1/>.
51. Neil Perry, Megha Srivastava, Deepak Kumar, and Dan Boneh, "Do Users Write More Insecure Code with AI Assistants?," *arXiv*, Cornell University, December 18, 2023, <https://doi.org/10.48550/arXiv.2211.03622>.
52. Charlotte Hill, Charlotte Allen, Tom Perkins, and Harriet Campbell, "Generative AI in the Courts: Getty Images v Stability AI," *Penningtons Manches Cooper*, February 16, 2024, <https://www.penningtonslaw.com/news-publications/latest-news/2024/generative-ai-in-the-courts-getty-images-v-stability-ai>.
53. Michael M. Grynbaum and Ryan Mac, "The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work," *New York Times*, December 27, 2023, <https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html>.
54. Blake Brittain, "Music Labels Sue AI Companies Suno, Udio for US Copyright Infringement," *Reuters*, June 24, 2024, <https://www.reuters.com/technology/artificial-intelligence/music-labels-sue-ai-companies-suno-udio-us-copyright-infringement-2024-06-24/>.
55. Maja S. Svanberg, Wensu Li, Martin Fleming, Brian C. Goehring, and Neil C. Thompson, "Beyond AI Exposure: Which Tasks Are Cost-Effective to Automate With Computer Vision?," *Future Tech*, Working Paper, January 18, 2024, https://futuretech-site.s3.us-east-2.amazonaws.com/2024-01-18+Beyond_AI_Exposure.pdf.
56. Claire Cain Miller and Courtney Cox, "In Reversal Because of A.I., Office Jobs Are Now More at Risk," *New York Times*, August 24, 2023, <https://www.nytimes.com/2023/08/24/upshot/artificial-intelligence-jobs.html>.
57. Martin Neil Bailey, Erik Brynjolfsson, and Anton Korinek, "Machines of Mind: The Case for an AI-Powered Productivity Boom," *Brookings Institution*, May 10, 2023, <https://www.brookings.edu/articles/machines-of-mind-the-case-for-an-ai-powered-productivity-boom>.
58. Pranshu Verma and Gerrit De Vynck, "ChatGPT Took Their Jobs: Now They Walk Dogs and Fix Air Conditioners," *Washington Post*, June 5, 2023, <https://www.washingtonpost.com/technology/2023/06/02/ai-taking-jobs/>; Challenger, Gray & Christmas, Inc., "Challenger Report," May 2023, <https://omscgcinc.wpenginepowered.com/wp-content/uploads/2023/06/The-Challenger-Report-May23.pdf>.
59. David Autor, Caroline Chin, Anna M. Salomons, and Bryan Seegmiller, "New Frontiers: The Origins and Content of New Work, 1940–2018," *National Bureau of Economic Research*, Working Paper 30389, August 2022, <https://doi.org/10.3386/w30389>.
60. Parliament and Council Regulation 2024/1689 Artificial Intelligence Act, art. 6-7, 2024 O.J. L 2024/1689, <https://artificialintelligenceact.eu/section/3-1/>.
61. White House, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Presidential Actions, Briefing Room, White House, October 30, 2023, <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>.
62. Department for Science, Innovation, and Technology, "The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023," Foreign, Commonwealth, and Development Office, Prime Minister's Office, November 1, 2023, <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>.
63. National Security Commission on Artificial Intelligence, Final Report, March 19, 2021, <https://apps.dtic.mil/sti/pdfs/AD1124333.pdf>.
64. C. Todd Lopez, "DOD Adopts 5 Principles of Artificial Intelligence Ethics," *DOD News*, US Department of Defense,

February 25, 2020, <https://www.defense.gov/News/News-Stories/article/article/2094085/dod-adopts-5-principles-of-artificial-intelligence-ethics>.

65. US Department of Defense, 2022 Nuclear Posture Review, October 27, 2022, 13, <https://apps.dtic.mil/sti/trecms/pdf/AD1183539.pdf>; Ministry of Defence, "Defence Artificial Intelligence Strategy," GOV.UK, June 15, 2022, <https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy/defence-artificial-intelligence-strategy>; United Nations Parties to the Treaty on Non-Proliferation of Nuclear Weapons, *Principles and Responsible Practices for Nuclear Weapon States*, NPT/CONF.2020/WP.70, July 2022, <https://undocs.org/NPT/CONF.2020/WP.70>.

66. Ashely Deeks, "Too Much Too Soon: China, the U.S., and Autonomy in Nuclear Command and Control," *Lawfare*, December 4, 2023, <https://www.lawfaremedia.org/article/too-much-too-soon-china-the-u.s.-and-autonomy-in-nuclear-command-and-control>.

STANFORD EXPERT CONTRIBUTORS

Dr. Fei-Fei Li

SETR Faculty Council, Sequoia Professor in the Computer Science Department, and Professor, by courtesy, of Operations, Information, and Technology at the Graduate School of Business

Dr. Christopher Manning

Thomas M. Siebel Professor of Machine Learning, and Professor of Linguistics and of Computer Science

Anka Reuel

SETR Fellow and PhD Student in Computer Science