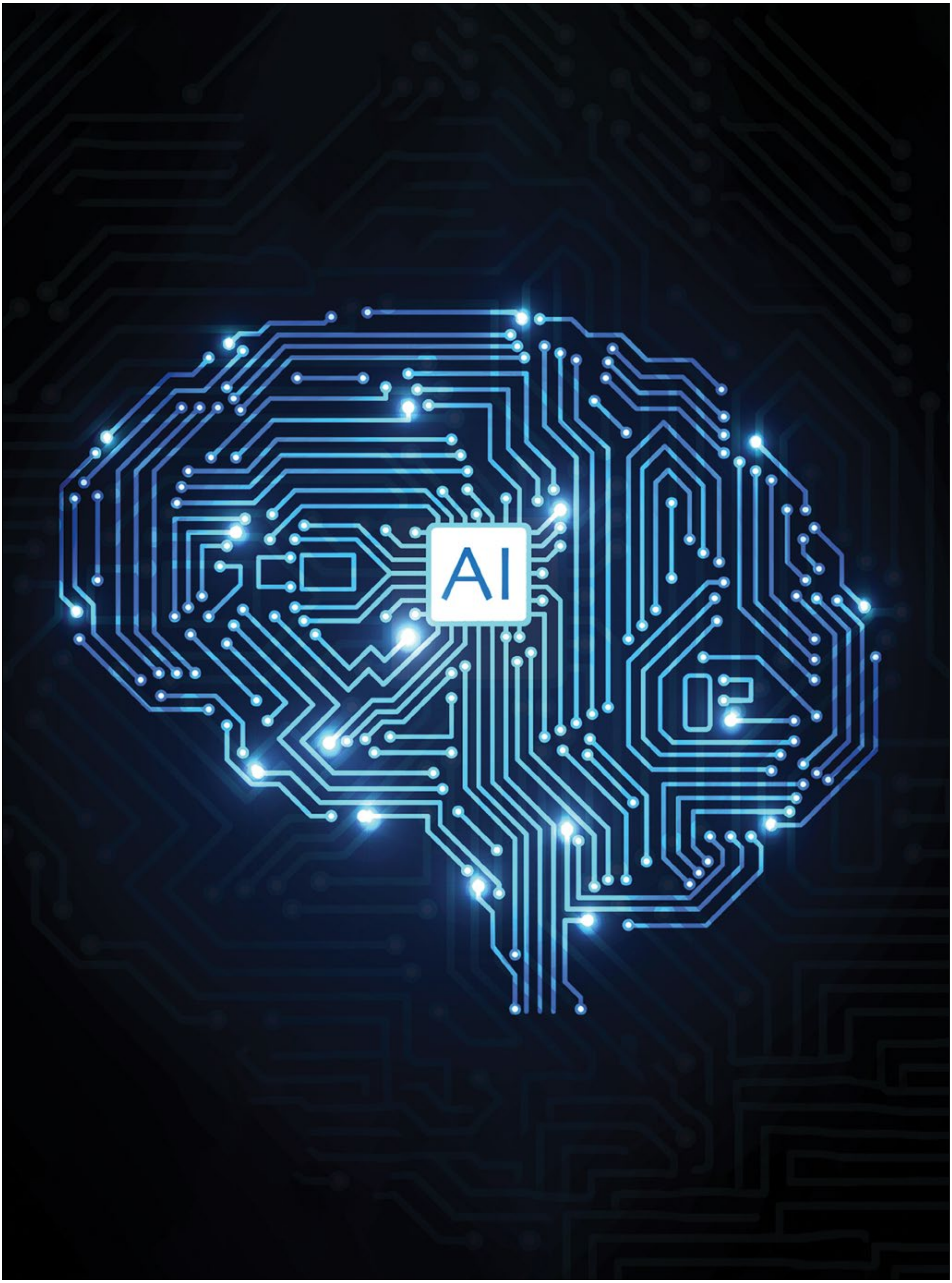
STANFORD UNIVERSITY

THE STANFORD EMERGING TECHNOLOGY REVIEW 2023

A Report on Ten Key Technologies and Their Policy Implications

CHAIRED BY Condoleezza Rice, John B. Taylor, Jennifer Widom, and Amy Zegart
DIRECTED BY Herbert S. Lin


AI

01

ARTIFICIAL INTELLIGENCE

KEY TAKEAWAYS

- AI is a foundational technology that is advancing other scientific fields and, like electricity and the internet, has the potential to transform how society operates.
- Even the most advanced AI has many failure modes that are unpredictable, not widely appreciated, not easily fixed, not explainable, and capable of leading to unintended consequences.
- There is substantial debate among AI experts about whether AI poses a long-term existential risk to humans, and whether the most important risks are current AI weaknesses.

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 on 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 recent years. The science of computing, worldwide availability of networks, and civilization-scale data—all that collectively underlies the AI of today and tomorrow—promises 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 interact directly with sophisticated AI applications for a multitude of everyday activities.

The global AI market was worth $136.55 billion in 2022, with North America receiving 36.8 percent of total AI revenues.[2] A Stanford University study found that total private investment in artificial intelligence exceeded $93 billion in 2021, a twofold increase in capital from 2020.[3] While artificial intelligence start-ups received roughly 9 to 10 percent of global venture capital investment in recent years,[4] global AI start-up funding slowed considerably in 2022, dropping from an all-time high of roughly $18 billion in the third quarter of 2021 to approximately $8.3 billion in quarter three of the following year.[5] Generative AI, discussed below, is estimated to raise global GDP by $7 trillion and lift productivity growth by 1.5 percent over a ten-year period, if adopted widely.[6]

What subfields are considered part of AI is a matter of ongoing discussion, and the boundaries between these fields are often fluid. Some of the core subfields include:

- Computer vision, enabling machines to recognize and understand visual information from the world, converting it into digital data and making decisions based on it

- 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 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 machine learning, though it draws on other subfields as well. Machine learning requires data and computing power—often called compute[7]—and much of today's AI research requires access to these on an enormous scale.

Artificial intelligence requires large amounts of data from which it can learn. These data can take various forms, including text, images, videos, sensor readings, and more. The quality and quantity of data play a crucial role in determining the performance and capabilities of AI models. Without sufficient and high-quality data, AI models may generate inaccurate or biased outcomes. (Roughly speaking, a model is developed to solve a particular problem—different problems call for different models, and for problems that are sufficiently different from each other, entirely new models need to be developed.) Research continues today on how to train models incrementally, starting from an existing model and then using a much smaller amount of specially curated data to refine the performance of those models for specialized purposes.

For a sense of scale, one AI model recently in the news is GPT-4. Estimates of the data required to train this model suggest that around a million books (hundreds of gigabytes of text) were drawn from billions of web pages and scanned books. The hardware requirements for computing power are also substantial. The costs to compute the training of GPT-4, for example, were enormous. Reports indicate the training of this application took about twenty-five thousand Nvidia A100 GPU deep-learning chips—at a cost of $10,000 each—running for about one hundred days.[8] 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 generally fabricated offshore.[9] (Chapter 8 on semiconductors discusses this point at greater length.)

Last, training AI models is an energy-intensive activity. One estimate of the electricity costs of training a large language model such as GPT-4 pegs the figure at about fifty million kilowatt-hours.[10] Then once it's up and running, the energy cost of a query on ChatGPT is around 0.002 of a kilowatt-hour.[11] Given hundreds of millions of queries per day, the operating energy cost of ChatGPT might be a few hundred thousand kilowatt-hours, at a cost of several tens of thousands of dollars, per day.

AI can automate a wide range of tasks. But AI also has particular promise in augmenting human capabilities and further enabling people to do what people are best at doing.[12] AI systems can work alongside people, complementing and assisting rather than replacing them. Some present-day examples include:

Health Care

- **Medical diagnostics** AI systems that can predict and detect the onset of strokes and subsequently perform automated triage, mobile viewing, and secure communication across several specialties and diseases qualified for Medicare reimbursement in 2020.[13]

- **Drug discovery** An AI-enabled search identified a compound that inhibits the growth of a bacterium responsible for many drug-resistant infections (e.g., pneumonia, meningitis) by sifting through a library of seven thousand potential drug compounds for an appropriate chemical structure.[14]

- **Robotic assistants** Mobile robots can carry out health care–related tasks such as making specialized deliveries, disinfecting hospital wards, and assisting physical therapists, thus supporting nurses and enabling them to spend more time with face-to-face human interactions.[15]

Agriculture

- **Production optimization** AI-enabled computer vision helps some salmon farmers sort fish into the right size to keep, thus off-loading the labor-intensive task of sorting fish.[16]

- **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.[17]

Logistics and Transportation

- **Resource allocation** AI enables some commercial shipping companies to predict ship arrivals five days in the future with high accuracy, thus allowing real-time allocations of personnel and schedule adjustments.[18]

Law

- **Legal transcription** AI enables the real-time transcription of legal proceedings and client meetings with reasonably high accuracy, and some such services are free of charge.[19]

- **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.[20]

Key Developments

Large Language Models

Large language models (LLMs) are AI systems trained on very large volumes of written text to recognize, summarize, and 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. A simple example might be that the word sequence "thank you" is far more likely to occur than "thank zebras." The resulting systems, which include chatbots such as ChatGPT, Bard, and Claude, generate output surprisingly similar to that of humans across a wide range of subjects, including computer code, poetry, legal case summaries, and medical advice. LLMs are examples of foundation models, which are machine-learning models trained on big datasets that can drive a large number of applications.[21] LLMs are also an example of generative AI, a type of AI that can produce new content (e.g., text, images, sounds, animation) based on how it has been trained and the inputs it is given.

Computer Vision

In recent years, computer vision has made substantial progress on a number of important problems, including:

- Image classification (categorizing objects in images)
- Facial detection and recognition (finding faces in images and then matching those faces to existing face images)
- Medical image segmentation (identifying an organ in an image and isolating the portions of the image associated with that organ)
- Object recognition (identifying and localizing instances of objects in images)
- Activity recognition (identifying human activity depicted in a video, e.g., a human being sitting or walking)[22]

AI-Enabled Scientific Discovery

Over the past few years, AI models using large amounts of scientific data have been the accelerant of several scientific discoveries. Prominent examples include protein structure predictions for multiple proteins associated with SARS-CoV-2,[23] the use of AI models to discover new antibodies,[24] and the improvement of plasma control procedures for nuclear fusion.[25]

Existential Concerns about AI

Large language models have generated considerable attention because of their apparent sophistication. Indeed, their capabilities have led some to suggest that LLMs are the initial sparks of artificial general intelligence (AGI).[26] AGI is artificial intelligence capable of performing any intellectual task that a human can perform, including learning. But because it would run on a computer, it is likely to learn much faster than humans—outstripping human capabilities in short order.

The prospect that artificial general intelligence will soon be achieved has raised substantial debate about its risks. In May 2023, a number of senior and respected AI researchers released a statement saying that "mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks, such as pandemics and nuclear war." Concerned about the speed at which the power of AI-enabled systems is growing, they worry that in the absence of good governance, future systems could pose existential risks to humanity.

Others suggest that focusing on low-probability doomsday scenarios distracts from the real and immediate risks that AI poses today.[27] Instead, they argue, AI researchers should prioritize addressing the harms AI systems are already causing, like biased decision making and job displacement. These problems are the ones on which governments and regulators should be focusing their efforts.

A National Artificial Intelligence Research Resource

As it stands today, AI models such as GPT-4 can be developed only by large industrial actors with the resources to build and operate large data and compute centers—companies such as Google, Microsoft, and Meta (previously Facebook). 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 cannot do so independently.

For this reason, a bipartisan group of legislators in July 2023 proposed a bill to establish the National Artificial Intelligence Research Resource (NAIRR) as a shared national research infrastructure that provides civil society researchers greater access to the complex resources, data, and tools needed to support research on safe and trustworthy artificial intelligence. Even so, the scale of government resources proposed is a factor of five or ten lower than what the private sector is willing and able to invest.[28]

Over the Horizon

Impact of New Technologies

New technologies often have positive and negative impacts. Potential positive impacts of new AI technologies are most likely to be seen in the applications they enable for societal use. These may include:

- Truck drivers can off-load to AI the most boring and time-consuming aspects of their jobs—the long-haul drives—and still retain those aspects of their jobs that require human-centered interactions, usually involving the first and last miles of their routes.
- Smart AI sensors and cameras can improve patient safety in intensive care units, operating rooms, and even at home by improving health-care providers' and caretakers' ability to monitor and react to patient health developments, including falls and injuries.[29]

Potential negative AI impacts likely will emerge from known problems with the current state-of-the-art AI and from unfettered success with AI in the future. Some of the known issues with today's leading AI models include:

Explainability This is the ability to explain the reasoning and describe the data underlying an AI system's conclusions. Explainability is useful for:

- Giving users confidence that an AI system works well
- Safeguarding against bias of various kinds
- Adhering to regulatory standards or policy requirements
- Helping developers understand why a system works a certain way, assess its vulnerabilities, or verify its outputs
- Meeting society's expectations about how individuals are afforded agency in a decision-making process[30]

Today's AI is for the most part 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.

Bias and fairness Because machine-learning models are trained on existing datasets, such models are likely to encode any biases present in such 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.[31] 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 history, they will also reflect the biases embedded in history, and a machine-learning model based on such datasets will also reflect such biases.

Vulnerability to spoofing For many AI models, it is possible to tweak data inputs to fool them into drawing false conclusions. For example, in figure 1.1, the addition of a small amount of noise to the "panda" image results in its being classified as a gibbon with very high confidence.

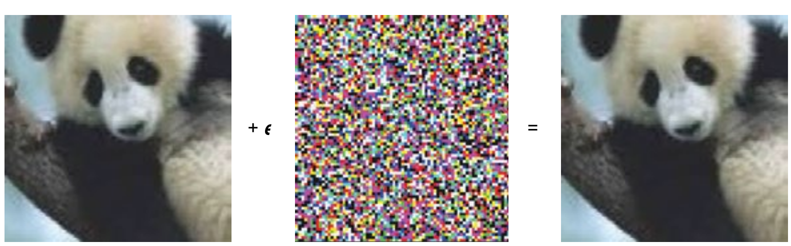
Deepfakes AI provides the capability for generating highly realistic but entirely inauthentic audio and video imagery. Such capability has obvious implications for evidence presented in a courtroom and political deception.

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

Overtrust If AI systems become a common presence in society, their novelty will inevitably diminish for users. The level of trust in computer outputs often increases with familiarity. On the other hand, skepticism about answers received from a system is necessary if one is to challenge the correctness of these outputs. As trust in AI grows due to reduced skepticism, there's a higher risk of overlooking errors, mishaps, and unforeseen incidents. One experiment recently showed that those with access to an AI-based coding assistant wrote code significantly less secure than those without an AI-based assistant—even though the former were more likely to believe they had written secure code.[32]

Hallucinations AI hallucinations refer to situations where an AI model generates results or answers that are plausible but nevertheless do not correspond to reality. That is, AI models can simply make things up seemingly from whole cloth. 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.

FIGURE 1.1 A panda turns into a gibbon



"panda"
57.7% confidence

"gibbon"
99.3% confidence

Source: Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy, "Explaining and Harnessing Adversarial Examples," paper presented at International Conference on Learning Representations, San Diego, May 2015.

Out-of-distribution (OOD) inputs All machine-learning systems must be trained on a large volume of data. If the inputs to a system are substantially different from the training data—or out of distribution—the system may well draw conclusions that are more unreliable than if the inputs were like the training data.

Copyright violations Some AI-based models have been trained on large volumes of data that have been 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 these data owners.

AI researchers are cognizant of issues such as those described above, and in many cases, work has been or is being done to develop corrective measures for them. Nevertheless, it's fair to say that in most cases, such defenses do not generalize very far beyond the specific problems against which these defenses are deployed.

Challenges of Innovation and Implementation

The primary challenge of bringing AI innovation into operation is risk management. It is often said that AI, especially machine learning, 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 it is new, the people who would make the decision to deploy AI-based systems do not have a good understanding of the risks that might accompany such deployment.

Consider, for example, artificial intelligence 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 proceeds at a slow pace. One important reason for this outcome 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. In this environment, it is 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 risk acceptance may well be necessary.

Policy, Legal, and Regulatory Issues

THE FUTURE OF WORK

Large language models such as GPT-4 have already demonstrated how they can be used in a variety of diverse 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, but uncertainty abounds. What and how many present-day jobs will disappear or be created? Which tasks could best be handled by AI?

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 feel painful shifts in the short term.[33]

- AI is helping some workers to increase productivity and job satisfaction.[34] At the same time, other workers are already losing their jobs as AI—despite potentially underperforming humans—demonstrates adequate competence for business operations.[35]

- 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 jobs 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.[36]

Research on foundational AI technologies is difficult if not impossible to regulate. . . . Regulation of specific applications of AI may be more easily implemented.

REGULATION OF AI

Governments around the world have been increasingly focused on establishing regulations and guidelines on AI. Research on foundational AI technologies is difficult if not impossible to regulate, especially when other nations have strong incentives to carry on regardless of actions taken by US policymakers. The same applies to voluntary restrictions on research by companies concerned about the competition. Regulation of specific applications of AI may be more easily implemented, in part because of existing regulatory frameworks in application domains such as health care, finance, and law.

The European Union is advancing comprehensive legislation that would provide for harmonized rules for the governance of artificial intelligence to mitigate new risks or negative consequences for individuals or society.[37] In the United States, more nascent federal and state discussions have advanced ideas such as the creation of a centralized AI agency, licensing of AI companies, and registration and transparency requirements of AI models.[38] Whether regulatory efforts will in the end prove consistent with technical realities remains to be seen.

NATIONAL SECURITY

AI is expected to have a profound impact on the military.[39] Weapons systems, command and control, logistics, acquisition, and training will leverage multiple AI technologies to operate more effectively and efficiently, at lower cost, and with less risk to friendly forces. The DOD is therefore dedicating billions of dollars to institutional reforms and research advancements aimed at integrating AI into its warfighting and war preparation strategies. 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.

The DOD is also cognizant of its obligation to proceed ethically with the development of AI capabilities; it has adopted a set of guiding principles that address responsibility, equity, traceability, reliability, and governability in and for AI.[40] An additional important concern, subsumed under these principles but worth calling out explicitly, is determining where the use of AI may or may not be appropriate, such as AI in nuclear command and control.

TALENT

The United States is eating its seed corn with respect to the AI talent pool. Faculty at Stanford and other universities report that the number of AI students 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 obtain funding to pay for research equipment and personnel like staff scientists, technicians, and programmers. They must search 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.

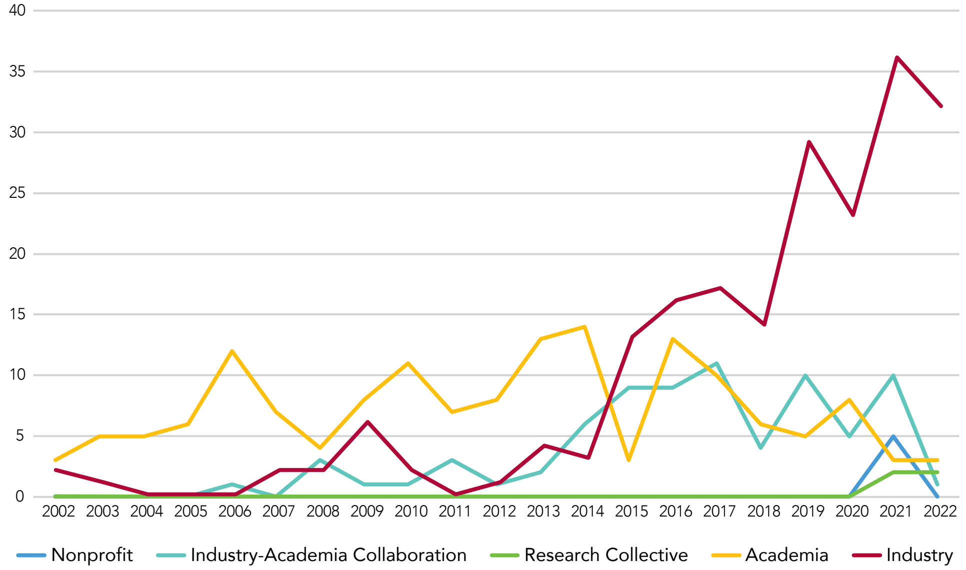
Industry often makes decisions more rapidly than government grant makers and imposes fewer regulations on the conduct of research. Large companies have the advantage of having research-supporting infrastructure in place, such as compute facilities and data storehouses.

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 not if their universities do not have access to the facilities of industry. Figure 1.2 shows how most significant machine-learning systems are now released by industry, with very few released by academic institutions, research collectives, or nonprofits.

At the same time, China's efforts to recruit top scientific talent offer further temptations for scientists to leave the United States. Although these efforts are often targeted toward ethnic Chinese in the United States—ranging from the well established to those finishing graduate degrees—China offers 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 these factors are leading to an AI brain drain that does not favor the US research enterprise.

FIGURE 1.2 Number of significant machine-learning systems by sector



Source: Nestor Maslej et al., *The AI Index 2023 Annual Report*, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2023.

NOTES

1. Christopher Manning, "Artificial Intelligence Definitions," Stanford Institute for Human-Centered AI, 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 2023–2030," accessed August 15, 2023, https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market.

3. Daniel Zhang et al., *The AI Index 2022 Report*, AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University, March 2022, https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf.

4. Gené Teare, "Special Series Launch: The Promises and Perils of a Decade of AI Funding," Crunchbase News, November 22, 2022, https://news.crunchbase.com/ai-robotics/ai-venture-cybersecurity-health-care-series-launch.

5. Shubham Sharma, "Report: AI Startup Funding Hits Record High of $17.9B in Q3," VentureBeat, November 11, 2021, https://venturebeat.com/business/report-ai-startup-funding-hits-record-high-of-17-9b-in-q3.

6. 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.

7. 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.

8. 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.

9. 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.

10. Ludvigsen, "The Carbon Footprint."

11. Ludvigsen, "ChatGPT's Electricity Consumption," Medium, July 12, 2023, https://towardsdatascience.com/chatgpts-electricity-consumption-7873483feac4.

12. Hope Reese, "A Human-Centered Approach to the AI Revolution," Stanford University Institute for Human-Centered AI, October 17, 2022, https://hai.stanford.edu/news/human-centered-approach-ai-revolution.

13. 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.

14. Gary Liu 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.

15. 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.

16. 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.

17. Itransition, "Machine Learning in Agriculture: Use Cases and Applications," February 1, 2023, https://www.itransition.com/machine-learning/agriculture.

18. GateHouse Maritime, "Vessel Tracking Giving Full Journey Visibility," accessed August 15, 2023, https://gatehousemaritime.com/data-services/vessel-tracking.

19. 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.

20. 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.

21. Rishi Bommasani et al., "On the Opportunities and Risks of Foundation Models," arXiv, Cornell University, August 16, 2021, https://arxiv.org/abs/2108.07258.

22. Nestor Maslej et al., *The AI Index 2023 Annual Report*, AI Index Steering Committee, Stanford Institute for Human-Centered AI, April 2023, 81–98, https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf.

23. John Jumper et al., "Computational Predictions of Protein Structures Associated with COVID-19," DeepMind, Google, August 4, 2020, https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19.

24. Amir Shanehsazzadeh et al., "Unlocking de Novo Antibody Design with Generative Artificial Intelligence," bioRxiv, March 29, 2023, https://doi.org/10.1101/2023.01.08.523187.

25. Jonas Degrave et al., "Magnetic Control of Tokamak Plasmas through Deep Reinforcement Learning," *Nature* 602 (2022): 414–19, https://doi.org/10.1038/s41586-021-04301-9.

26. Sébastien Bubeck 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.

27. "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.

28. Dina Bass, "Microsoft Invests $10 Billion in ChatGPT Maker OpenAI," Bloomberg, January 23, 2023, https://www.bloomberg.com/news/articles/2023-01-23/microsoft-makes-multibillion-dollar-investment-in-openai; Office of Management and Budget, Executive Office of the President, "Proposed Funding for the National AI Initiative as a Whole Is Measured in a Few Billion," *Budget of the U.S. Government, Fiscal Year 2024*, 2023, 132–33, https://www.whitehouse.gov/wp-content/uploads/2023/03/budget_fy2024.pdf.

29. 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.

30. Royal Society, "Explainable AI: The Basics," November 28, 2019, https://royalsociety.org/topics-policy/projects/explainable-ai.

31. Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," Conference on Fairness, Accountability, and Transparency,

Proceedings of Machine Learning Research 81 (February 2018): 1–15, https://www.media.mit.edu/publications/gender-shades-intersectional-accuracy-disparities-in-commercial-gender-classification

32. Neil Perry et al., "Do Users Write More Insecure Code with AI Assistants?," arXiv, Cornell University, December 16, 2022, https://doi.org/10.48550/arXiv.2211.03622.

33. 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.

34. Martin Neil Baily, 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.

35. 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.

36. David Autor et al., "New Frontiers: The Origins and Content of New Work, 1940–2018," National Bureau of Economic Research, August 2022, https://doi.org/10.3386/w30389.

37. European Commission, Proposal for a Regulation of the European Parliament and the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, European Union, June 1, 2021, https://eur-lex.europa.eu/legal-content/EN/TXT.

38. Dylan Matthews, "The AI Rules That US Policymakers Are Considering, Explained," Vox, August 1, 2023, https://www.vox.com/future-perfect/23775650/ai-regulation-openai-gpt-anthropic-midjourney-stable.

39. National Security Commission on Artificial Intelligence, Final Report, March 19, 2021, https://apps.dtic.mil/sti/pdfs/AD1124333.pdf.

40. 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.