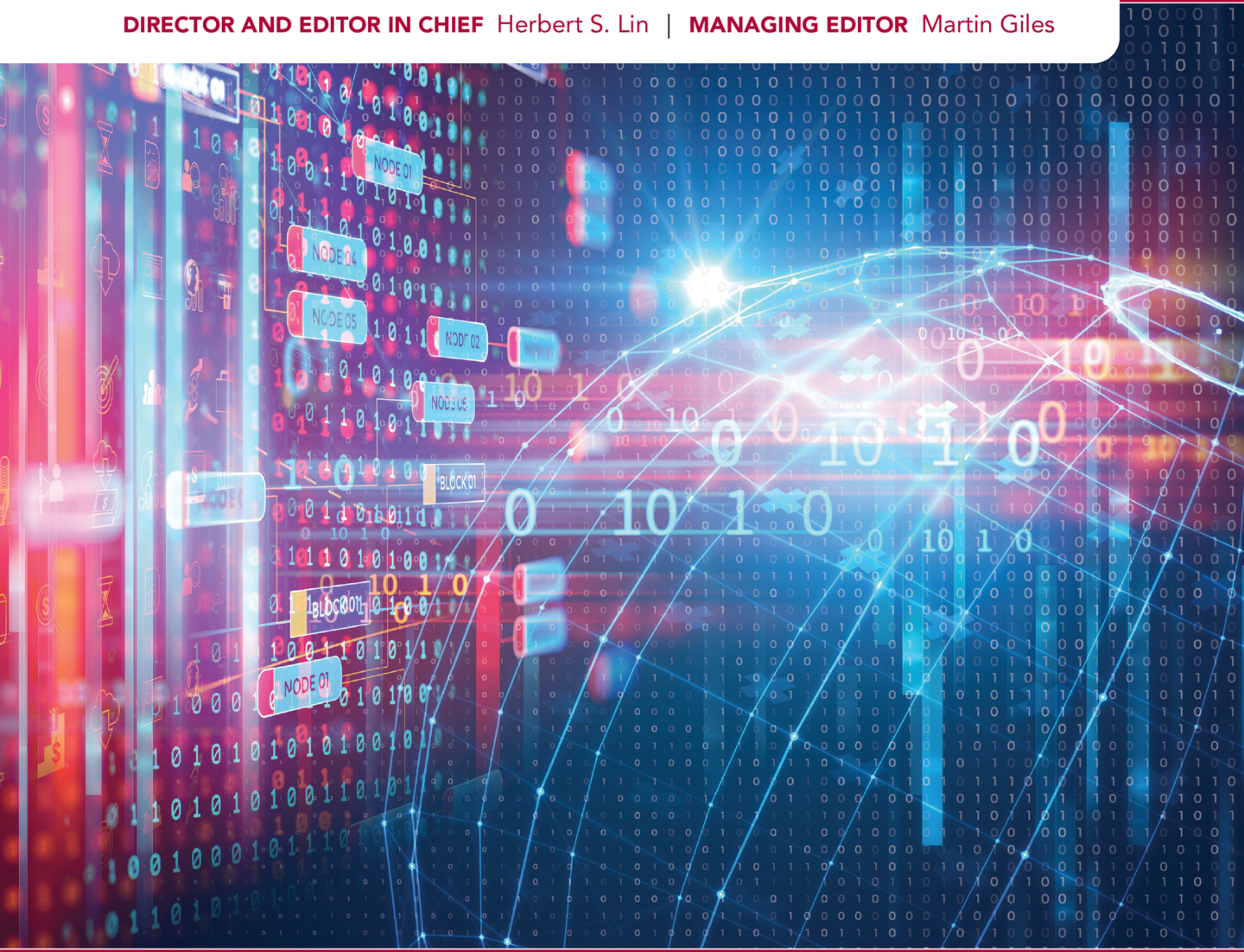
STANFORD UNIVERSITY

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



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
Amy Zegart

DIRECTOR AND EDITOR IN CHIEF
Herbert S. Lin

MANAGING EDITOR
Martin Giles

Stanford University
Stanford, California

CONTENTS

FOREWORD

In every era, technological discoveries bring both promise and risk. Rarely, however, has the world experienced technological change at the speed and scale we see today. From nanomaterials that are fifty thousand times smaller than the width of a human hair to commercial satellites and other private-sector technologies deployed in outer space, breakthroughs are rapidly reshaping markets, societies, and geopolitics. What's more, US technology policy isn't the unique province of government like it used to be. Instead, inventors and investors are making decisions with enormous policy consequences, even if they may not always realize it. Artificial intelligence (AI) algorithms are imbued with policy choices about which outcomes are desired and which are not. Nearly every new technology, from bioengineering new medicines to building underwater research drones, has both commercial and military applications. Private-sector investment, too, simultaneously generates both national advantages and vulnerabilities by developing new capabilities, supply chains, and dependencies and by pursuing commercial opportunities that may not serve long-term national interests.

While engineers and executives need to better understand the policy world, government leaders need to better understand the engineering and business worlds. Otherwise, public policies intended to protect against societal harms may end up accelerating them, and efforts to align innovation with the national interest could end up harming that interest by dampening America's innovation leadership and the geopolitical advantages that come with it.

In these complex times, the only certainties are that uncertainty is rampant and the stakes are high: Decisions made today in boardrooms, labs, and government offices are likely to set trajectories for the United States and the world for years to come.

Now more than ever, understanding the landscape of discovery and how to harness technology to forge a better future requires working across sectors, fields, and generations. Universities like Stanford have a vital role to play in this effort. In 2023, we launched the Stanford Emerging Technology Review (SETR), the first-ever collaboration between Stanford University's School of Engineering and the Hoover Institution. Our goal is ambitious: transforming technology education for decision makers in both the public and private sectors so that the United States can seize opportunities, mitigate risks, and ensure the American innovation ecosystem continues to thrive.

This is our latest report surveying the state of ten key emerging technologies and their implications. It harnesses the expertise of leading faculty in science and engineering fields, economics, international relations, and history to identify key technological developments, assess potential implications, and highlight what policymakers should know.

This report is our flagship product, but it is just one element of our continuous technology education campaign for policymakers that now involves nearly one hundred Stanford scholars across forty departments and research institutes. In the past year, SETR experts have briefed senior leaders across the US government—in Congress and in the White House, Commerce Department, Defense Department, and US intelligence community. We have organized and participated in fifteen Stanford programs, including multiday AI and biotechnology boot camps for congressional staff; SETR roundtables for national media and officials from European partners and allies; and workshops convening leaders across sectors in semiconductors, space technology, and bioengineering. And we are just getting started.

Our efforts are guided by three observations:

1. **America's global innovation leadership matters.**

American innovation leadership is not just important for the nation's economy and security. It is the linchpin for maintaining a dynamic global technology innovation ecosystem and securing its benefits.

International scientific collaboration has long been pivotal to fostering global peace, progress, and prosperity, even in times of intense geopolitical competition. During the Cold War, American and Soviet nuclear scientists and policymakers worked together to reduce the risk of accidental nuclear war through arms control agreements and safety measures. Today, China's rise poses many new challenges. Yet maintaining a robust global ecosystem of scientific cooperation remains essential—and it does not happen by magic. It takes work, leadership, and a fundamental commitment to freedom to sustain the openness essential for scientific discovery. Freedom is the fertile soil of innovation, and it takes many forms: the freedom to criticize a government; to admit failure in a research program as a step toward future progress; to share findings openly with others; to collaborate across geographical and technical borders with reciprocal access to talent, knowledge, and resources; and to work without fear of repression or persecution. In short, it matters whether the innovation ecosystem is led by democracies or autocracies. The United States has its flaws and challenges, but this country remains the best guarantor of scientific freedom in the world.

2. **Academia's role in American innovation is essential—and at risk.**

The US innovation ecosystem rests on three pillars: the government, the private sector, and the academy. Success requires robust research and development (R&D) in all three. But they are not the same, and evidence increasingly suggests that universities' role as the engines of innovation is at a growing risk.

Universities, along with the US National Laboratories, are the only institutions that conduct research on the frontiers of knowledge without regard for potential profit or foreseeable commercial application. This kind of research is called basic or fundamental research. It takes years, sometimes decades, to bear fruit. But without it, future commercial innovations would not be possible. Radar, the Global Positioning System (GPS), and the internet all stemmed from basic research done in universities. So did the recent "overnight success" of the COVID-19 mRNA vaccines, which relied on decades of university research that discovered mRNA could activate and block protein cells and figured out how to deliver mRNA to human cells to provoke an immune response. Similarly, the cryptographic algorithms protecting data on the internet today would not have been possible without decades of academic research in pure math. And many of the advances in AI, from ChatGPT to image recognition, build on pioneering work done in university computer science departments that also trained legions of students who have gone on to found, fund, and lead many of today's most important tech companies. In many ways and in nearly every field, America's innovation supply chain starts with research universities.

Yet evidence suggests that the engine of innovation in US research universities is not running as well as it could, posing long-term risks to the nation. In 2024, for the first time, the number of Chinese contributions surpassed those of the United States in the closely watched Nature Index, which tracks eighty-two of the world's premier science journals.[1] Funding trends are also heading in the wrong direction. The US government is the only funder capable of making large and risky investments in the basic science conducted at universities (as well as at national laboratories) that is essential for future applications. Yet federal R&D funding has plummeted in percentage terms since the 1960s, from 1.86 percent of GDP in 1964 to just 0.66 percent of GDP in 2016.[2] The Creating Helpful Incentives to Produce Semiconductors (CHIPS) and Science Act of 2022 was supposed to turn the tide by dramatically raising funding for basic research, but major increases were subsequently scrapped in budget negotiations. The United States still funds more basic research than China does, but Chinese investment is rising six times faster—and is expected to overtake US spending within a decade.[3]

Although private-sector investment in technology companies and associated university research has increased substantially, it is not a substitute for federal funding, which supports university R&D directed at national and public issues, not commercial viability.[4]

To be sure, the rising dominance of private industry in innovation brings significant benefits. But it is also generating serious and more hidden risks to the health of the entire American innovation ecosystem. Technology and talent are migrating from academia to the private sector, accelerating the development of commercial products while eroding the foundation for the future. We are already reaching a tipping point in AI. In 2022, more than 70 percent of students who received PhDs in artificial intelligence at US universities took industry jobs, leaving fewer faculty to teach the next generation.[5] As the bipartisan National Security Commission on Artificial Intelligence put it, "Talent follows talent."[6]

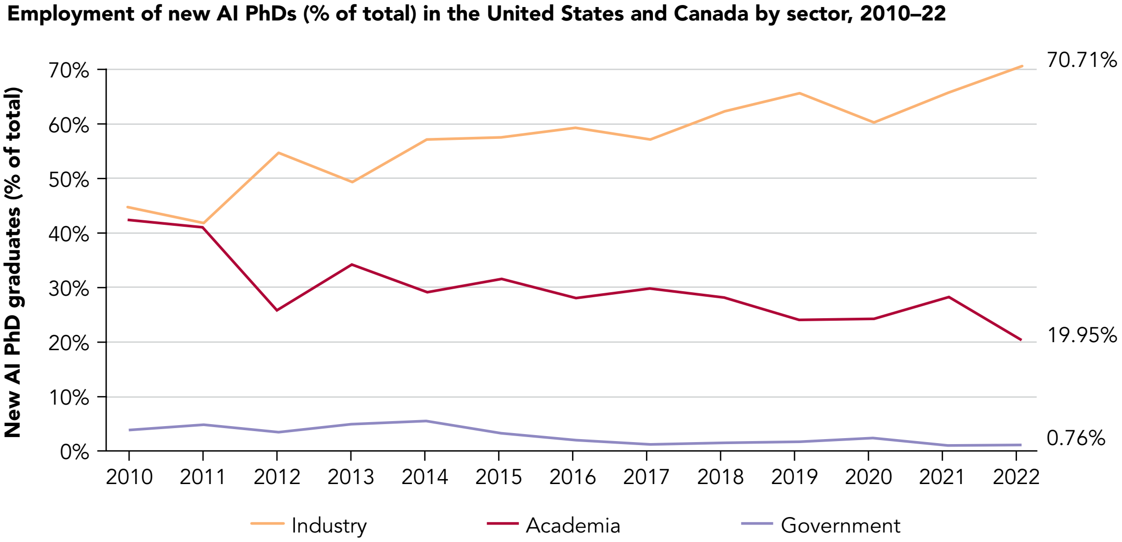
Today, only a handful of the world's largest companies have both the talent and the enormous compute power necessary for developing sophisticated large language models (LLMs) like ChatGPT. No university comes close. In 2024, for example, Princeton University announced that it would use endowment funds to purchase 300 advanced Nvidia chips to use for research, costing about $9 million, while Meta announced plans to purchase 350,000 of the same chips by year's end, at an estimated cost of $10 billion.[7]

These trends have several concerning implications.[8] A very significant one is that research in the field is likely to be skewed to applications driven by commercial rather than public interests. The ability for universities—or anyone outside of the leading AI companies—to conduct independent analysis of the weaknesses, risks, and vulnerabilities of AI (especially LLMs recently in the news) will become more important and simultaneously more difficult. Further, the more that industry offers unparalleled talent concentrations, computing power, training data, and the most sophisticated models, the more likely it is that future generations of the best AI minds will continue to flock there (see figure F.1)—potentially eroding the nation's ability to conduct broad-ranging foundational research in the field.

3. The view from Stanford is unique, important—and needed now more than ever.

Stanford University has a unique vantage point when it comes to technological innovation. It is not an accident that Silicon Valley surrounds Stanford; technology developed at Stanford in the 1930s served as the foundation for the pioneering companies like Varian Associates and Hewlett-Packard that first shaped industry in the region. Since then, the university has continued to fuel that innovation ecosystem. Stanford faculty, researchers, and former students have founded Alphabet, Cisco Systems, Instagram, LinkedIn, Nvidia, Sun Microsystems, Yahoo!, and many other companies, together generating more annual revenues than most of the world's economies.

FIGURE F.1 Most new AI PhDs hired in North America are flocking to industry



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 CRA Taulbee Survey, 2023

Start-ups take flight in our dorm rooms, classrooms, laboratories, and kitchens. Technological innovation is lived every day and up close on our campus—with all its benefits and downsides. This ecosystem and its culture, ideas, and perspectives often seem a world apart from the needs and norms of Washington, DC. Bridging the divide between the locus of American policy and the heart of American technological innovation has never been more important.

Stanford has a rich history of policy engagement, with individuals who serve at the highest levels of government as well as institutional initiatives that bring together policymakers and researchers to tackle the world's toughest policy problems. And as Stanford's School of Engineering celebrates its one hundredth anniversary in 2025, we are reminded of the profound impact that generations of Stanford faculty, students, and staff have had through their discoveries—from the klystron, a microwave amplifier developed in the 1930s that enabled radar and early satellite communications; to the algorithms driving Google; to optogenetics, a technique pioneered in 2005 that uses light to control neurons, enabling precise studies of brain function. In this moment of rapid technological change, we must do even more to connect emerging technologies with policy. We are proud and excited to highlight this collaboration between Stanford's Hoover Institution and the School of Engineering to bring policy analysis, social science, science, medicine, and engineering together in new ways.

Today, technology policy and education efforts are often led by policy experts with limited technological expertise. The Stanford Emerging Technology Review flips the script, enlisting ten of the brightest scientific and engineering minds at the university to share their knowledge of their respective fields by working alongside social scientists to translate their work to nonexpert audiences. We start with science and technology, not policy. And we go from there to emphasize the important interaction between science and all aspects of policy.

How to Use This Report: One Primer, Ten Major Technology Areas

This report is intended to be a one-stop-shopping primer that covers developments and implications in ten major emerging technology areas: AI, biotechnology and synthetic biology, cryptography, lasers, materials science, neuroscience, robotics, semiconductors, space, and sustainable energy technologies. The list is broad by design, and it includes fields that are widely regarded as pivotal to shaping society, economics, and geopolitics today and into the future.

That said, the ten major technology areas covered in this report are nowhere near an exhaustive catalogue of technology research areas at Stanford. And the list may change year to year—not because a particular technology sputtered or because we got it wrong, but because categorizing technologies is inherently dynamic. Limiting this report to ten areas imposes discipline on what we cover and how deeply we go. We seek to highlight relationships among technologies in ways that may not be obvious: Quantum computing, for example, is an important field but does not have its own chapter. Instead, it is covered within the semiconductor chapter because we wanted to emphasize that even if quantum breakthroughs are realized, they will not address many important computing needs and challenges. Of note, nine of the ten technology chapters appearing in this edition are on the same subjects as in our previous report. In this report, we have combined nuclear energy and sustainable energy technologies into a single chapter and added a chapter on lasers.

Many of the most important issues cut across technological fields. We have expanded our previous report's crosscutting themes chapter to highlight fourteen of these themes and offer more examples and discussion. The themes include broad trends, like the tendency for technological breakthroughs to come in fits, starts, and lengthy plateaus that are extremely difficult even for leaders in those fields to predict. (AI leaders have experienced several so-called AI winters over decades as well as moments of profound and sudden progress like the 2022 release of ChatGPT.) They include enduring and widespread technological challenges like cybersecurity. And they include cognitive blind spots like frontier bias—the natural but mistaken assumption that the only transformational technologies sit on the frontiers of a field.

For each of the ten technology chapters, reviews of the field were led by world-renowned tenured Stanford faculty members who also delivered seminars with other faculty discussants within and outside their areas of expertise. (SETR contributors and their fields are listed at the end of each chapter.) The SETR team also involved more than a dozen postdoctoral scholars and undergraduate research assistants who interviewed faculty across Stanford and drafted background materials.

Each technology chapter begins with an overview of the basics—the major technical subfields, concepts, and terms needed to understand how a technology works and could affect society. Next, we outline important developments and advances in the field. Finally, each chapter concludes by offering an over-the-horizon outlook that covers the most crucial considerations for policymakers over the next few years—including technical as well as policy, legal, and regulatory issues. The report ends with a chapter that looks across the ten technologies, offering analysis of implications for economic growth, national security, environmental and energy sustainability, human health, and civil society.

Three points bear highlighting. First, **we offer no specific policy recommendations in this report**. That is by design. Washington is littered with reports offering policy recommendations that were long forgotten, overtaken by events, or both. Opinions are plentiful. Expert insights based on leading research are not.

We aim to provide a reference resource that is both timeless and timely, an annual state-of-the-art guide that can inform successive generations of policymakers about how to think about evolving technological fields and their implications. Individual SETR faculty may well have views about what should be done. Some of us engage in policy writing and advising. But the mission of this collective report is informing, not advocating. We encourage readers interested in learning more about specific fields and policy ideas to contact our team at SETReview2025@stanford.edu.

Second, **SETR offers *a* view from Stanford, not *the* view from Stanford**. There is no single view of anything in a university. Faculty involved in this report may not agree with everything in it. Their colleagues would probably offer a different lay of the technology landscape with varying assessments about important developments and over-the-horizon issues. The report is intended to reflect the best collective judgment about the state of these ten fields—guided by leading experts in them.

Third, this report is intended to be the **introductory product that translates a broad swatch of technological research for nontechnical readers**. Other SETR offerings provide deeper dives into specific technological areas that should be of interest for subject-matter experts.

Ensuring continued American leadership in science and technology is essential, and it's a team effort. We hope this edition of the *Stanford Emerging Technology Review* continues to spark meaningful dialogue, better policy, and lasting impact. The promise of emerging technology is boundless if we have the foresight to understand it and the fortitude to embrace the challenges.

Condoleezza Rice
John B. Taylor
Jennifer Widom
Amy Zegart

Co-chairs, Stanford Emerging Technology Review

NOTES

1. Simon Baker, "China Overtakes United States on Contribution to Research in Nature Index," *Nature*, May 19, 2023, https://doi.org/10.1038/d41586-023-01705-7.

2. Council on Foreign Relations, *Innovation and National Security: Keeping Our Edge*, Independent Task Force Report No. 77, James Manyika and William H. McRaven, chairs, 2019, 10, https://www.cfr.org/keeping-our-edge/pdf/TFR_Innovation_Strategy.pdf.

3. OECD, "OECD Data Explorer: Gross Domestic Expenditure on R&D by Sector of Performance and Type of R&D," OECD.org, 2024, https://data-viewer.oecd.org/?chartId=b08256f3-6000-42a5-ad81-b1ab5bd8ac6a. Analysis was conducted by comparing two methods of measuring annual increases of US and China spending on basic research and projecting rates into future.

4. Council on Foreign Relations, *Innovation*, 21.

5. 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.

6. Eric Schmidt and Robert Work, *NSCAI Interim Report*, National Security Commission on Artificial Intelligence, 2020, 30, https://apps.dtic.mil/sti/pdfs/AD1112059.pdf.

7. Poornima Apte, "Princeton Invests in New 300-GPU Cluster for Academic AI Research," AI at Princeton, March 15, 2024, https://ai.princeton.edu/news/2024/princeton-invests-new-300-gpu-cluster-academic-ai-research; Michael Kan, "Zuckerberg's Meta Is Spending Billions to Buy 350,000 Nvidia H100 GPUs," *PCMAG*, January 18, 2024, https://www.pcmag.com/news/zuckerbergs-meta-is-spending-billions-to-buy-350000-nvidia-h100-gpus; Kif Leswing, "Nvidia's Latest AI Chip Will Cost More than $30,000, CEO Says," CNBC, March 19, 2024, https://www.cnbc.com/2024/03/19/nvidias-blackwell-ai-chip-will-cost-more-than-30000-ceo-says.html.

8. Roman Jurowetzki, Daniel Hain, Juan Mateos-Garcia, and Konstantinos Stathoulopoulos, "The Privatization of AI Research (-ers): Causes and Potential Consequences—From University-Industry Interaction to Public Research Brain-Drain?," arXiv, 2021, https://arxiv.org/ftp/arxiv/papers/2102/2102.01648.pdf.

EXECUTIVE SUMMARY

Emerging technologies have never been more important or difficult to understand. Breakthrough advances seem to be everywhere, from ChatGPT to the COVID-19 mRNA vaccines to constellations of cheap commercial shoebox-size satellites that can track events on Earth in near–real time. This is a pivotal technological moment offering both tremendous promise and unprecedented challenges. Policymakers need better expert resources to help them understand the burgeoning and complex array of technological developments—more easily and more continuously.

The *Stanford Emerging Technology Review* is designed to meet this need, offering an easy-to-use reference tool that harnesses the expertise of Stanford University's leading science and engineering faculty in ten major technological areas.

***SETR 2025* FOCUS TECHNOLOGIES**

Artificial Intelligence
Biotechnology and Synthetic Biology
Cryptography
Lasers
Materials Science
Neuroscience
Robotics
Semiconductors
Space
Sustainable Energy Technologies

These particular fields were chosen for this report because they leverage areas of deep expertise at Stanford and cover many critical and emerging technologies identified by the Office of Science and Technology Policy in the White House and other US government departments. However, *SETR* focus technologies are likely to change over time, not because we were incorrect, but because science and technology never sleep, the borders between fields are porous, and different people categorize similar research in different ways.

Report Design

This report is organized principally by technology, with each area covered in a standalone chapter that gives an overview of the field, highlights key developments, and offers an over-the-horizon view of important technological and policy considerations. Although these chapters can be read individually, one of the most important and unusual hallmarks of this moment is convergence: Emerging technologies are intersecting and interacting in a host of ways, with important implications for policy. We examine these broader dynamics in chapters 11 and 12. In chapter 11, we describe a number of themes and commonalities that cut across many of the technologies we describe earlier in the report. In chapter 12, we consolidate technological developments across all ten areas and discuss how they apply to five policy domains: economic growth, national security, environmental and energy sustainability, health and medicine, and civil society.

Three tensions run throughout and are worth keeping in mind:

1. Timeliness and timelessness Each chapter seeks to strike a balance between covering recent developments in science and in the headlines and providing essential knowledge about how a field works, what is important within it, and what challenges lie ahead.

2. Technical depth and breadth This report intentionally skews toward breadth, offering a 30,000-foot view of a vast technological landscape in one compendium. Readers should consider it an introductory course. Other products and/or educational tools will be released in the months ahead that will offer additional insights into each field.

3. Technical and nontechnical aspects of innovation We start with the science but do not end with the science. Technological breakthroughs are necessary but not sufficient conditions for successful innovation. Economic, political, and societal factors play enormous and often hidden roles. Johannes Gutenberg invented the printing press in 1452, but it took more than 150 years before the Dutch invented the first successful newspapers—not because they perfected the mechanics of movable type, but because they decided to use less paper, making newspapers sustainably profitable for the first time.[1] Each chapter in this report was written with an eye toward highlighting important economic, political, policy, legal, and societal factors likely to impede, shape, or accelerate progress.

Technologies and Takeaways at a Glance

Artificial Intelligence

Artificial intelligence (AI) is a computer's ability to perform some of the functions associated with the human brain, including perceiving, reasoning, learning, interacting, problem-solving, and even exercising creativity. In the past year, the main AI-related headlines have been the rise of large language models (LLMs) like GPT-4, on which some versions of the chatbot ChatGPT are based, and the recognition of AI's significance through the awarding of two Nobel Prizes in physics and chemistry for AI-related work.

KEY CHAPTER 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.

Biotechnology and Synthetic Biology

Biotechnology is the use of cellular and biomolecular processes to develop products or services. Synthetic biology is a subset of biotechnology that involves using engineering tools to modify or create biological functions—like creating a bacterium that can glow in the presence of explosives. Synthetic biology is what created the COVID-19 mRNA vaccine in record time—although it relied on decades of earlier research. Just as rockets enabled humans to overcome the constraints of gravity to explore the universe, synthetic biology is enabling humans to overcome the constraints of lineage to develop new living organisms.

KEY CHAPTER TAKEAWAYS

- Biotechnology is poised to emerge as a general-purpose technology by which anything bioengineers learn to encode in DNA can be grown whenever and wherever needed—essentially enabling the production of a wide range of products through biological processes across multiple sectors.

- The US government is still working to grasp the scale of this bio-opportunity and has relied too heavily on private-sector investment to support the foundational technology innovation needed to unlock and sustain progress.

- Biotechnology is one of the most important areas of technological competition between the United States and China, and China is investing considerably more resources. Lacking equivalent efforts domestically, the United States runs the risk of Sputnik-like strategic surprises in biotechnology.

Cryptography

The word *cryptography* originates from Greek words that mean "secret writing." In ancient times, cryptography involved the use of ciphers and secret codes. Today, it relies on sophisticated mathematical models to protect data from being altered or accessed inappropriately. Cryptography is often invisible, but it is essential for most internet activities, such as messaging, e-commerce, and banking. In recent years, a type of cryptographic technology called blockchain—which records transactions in distributed ledgers in the computing cloud that cannot be altered retroactively without being detected—has been used for a variety of applications, including time-stamping and ensuring the provenance of information, identity management, supply chain management, and cryptocurrencies.

KEY CHAPTER TAKEAWAYS

- Cryptography is essential for protecting information, but alone it cannot secure cyberspace against all threats.

- Cryptography is the enabling technology of blockchain, which is the enabling technology of cryptocurrencies.

- Central bank digital currencies (CBDCs) are a particular type of cryptography-based digital currency supported by states and one that could enhance financial inclusion. Although the United States lags some countries in experimenting with a CBDC, it may benefit from a cautious, well-timed approach by learning from other nations' efforts.

Lasers

Improvements in laser technology since its invention in 1960 have enabled light to be manipulated and used in previously unimaginable ways. It is now applied so broadly that it can be considered an enabling technology—one whose existence and characteristics enable other applications that would not be feasible and/or affordable without it. New lasers are being developed by researchers and companies across a wider range of light wavelengths, which will make the devices even more useful.

KEY CHAPTER TAKEAWAYS

- Laser technology has become essential for a wide range of applications, including communications, high-end chip production, defense, manufacturing, and medicine.

- Because advances in laser technology tend to occur in the context of specific applications, laser technology research and development is widely dispersed among different types of laboratories and facilities.

- Broad investment in next-generation lasers holds the potential to improve progress in nuclear fusion energy technology, weapons development, and quantum communication.

Materials Science

Materials science studies the structure and properties of materials—from those visible to the naked eye to microscopic features—and how they can be engineered to change performance. Contributions to the field have led to better semiconductors, "smart bandages" with integrated sensors and simulators to accelerate healing, more easily recyclable plastics, and more energy-efficient solar cells. Materials science has also been key to the development of additive manufacturing, often known as 3-D printing.

KEY CHAPTER TAKEAWAYS

- Materials science is a foundational technology that underlies advances in many other fields, including robotics, space, energy, and synthetic biology.

- Materials science will exploit artificial intelligence as another promising tool to predict new materials with new properties and identify novel uses for known materials.

- Future progress in materials science requires new funding mechanisms to more effectively transition from innovation to implementation and access to more computational power.

Neuroscience

Neuroscience is the study of the human brain and the nervous system—its structure, function, healthy and diseased states, and life cycle from embryonic development to degeneration in later years. The brain is perhaps the least understood and yet most important organ in the human body. Three major research subfields of neuroscience are neuroengineering (e.g., brain-machine interfaces), neurohealth (e.g., brain degeneration and aging), and neurodiscovery (e.g., the science of addiction).

KEY CHAPTER TAKEAWAYS

- Popular interest in neuroscience vastly exceeds the actual current scientific understanding of the brain, giving rise to overhyped claims in the public domain that revolutionary advances are just around the corner.

- Advances in human genetics and experimental neuroscience, along with computing and neuroscience theory, have led to some progress in several areas, including understanding and treating addiction and neurodegenerative diseases and designing brain-machine interfaces for restoring vision.

- American leadership is essential for establishing and upholding global norms about ethics and human subjects research in neuroscience, but this leadership is slipping with decreased strategic planning and increased foreign investments in the field.

Robotics

Robotics is an integrative field that draws on advances in multiple technologies rather than a single discipline. The question "What is a robot?" is harder to answer than it appears. At a minimum, the emerging consensus among researchers is that a robot is a physical entity that has ways of sensing itself and the world around it and can create physical effects on that world. Robots are already used across a range of sectors in a variety of ways—including assembly-line manufacturing, space exploration, autonomous vehicles, tele-operated surgery, military reconnaissance, and disaster assistance.

KEY CHAPTER TAKEAWAYS

- Future robots may be useful for improving the US manufacturing base, reducing supply chain vulnerabilities, delivering eldercare, enhancing food production, tackling the housing shortage, improving energy sustainability, and performing almost any task involving physical presence.

- Progress in artificial intelligence holds the potential to advance robotics significantly but also raises ethical concerns that are essential to address, including the privacy of data used to train robots, data bias that could lead to physical harm by robots, and other safety issues.

- Achieving the full potential of robots will require a major push from the federal government and the private sector to improve robotics adoption and research across the nation.

Semiconductors

Semiconductors, or chips, are crucial and ubiquitous components used in everything from refrigerators and toys to smartphones, cars, computers, and fighter jets. Chip production involves two distinct steps: (1) *design*, which requires talented engineers to design complex integrated circuits involving millions of components; and (2) *fabrication*, which is the task of manufacturing chips in large, specially designed factories called "fabs." Because fabs involve highly specialized equipment and facilities, they can cost billions of dollars. US companies still play a leading role in semiconductor design, but US semiconductor-manufacturing capacity has plummeted, leaving the country heavily dependent on foreign chips, most notably from Taiwan. The Creating Helpful Incentives to Produce Semiconductors (CHIPS) and Science Act of 2022 was intended to help the US semiconductor industry regain a foothold in fabrication, but progress will take years, if not decades.

KEY CHAPTER TAKEAWAYS

- The growing demand for artificial intelligence and machine learning is driving innovations in chip fabrication that are essential for enhancing computational power and managing energy efficiency.

- Advances in memory technologies and high-bandwidth interconnects, including photonic links, are critical for meeting the increasing data needs of modern applications.

- Even if quantum computing advancements are realized, the United States will still need comprehensive innovation across the technology stack to continue to scale the power of information technology.

Space

Space technologies include any technology developed to conduct or support activities approximately sixty miles or more beyond Earth's atmosphere. A single space mission is a system of systems—including everything from the spacecraft itself to propulsion, data storage and processing, electrical power generation and distribution, thermal control to ensure that components are within their operational and survival limits, and ground stations. While in the past, space was the exclusive province of government spy satellites and discovery missions, the number and capabilities of commercial satellites have increased dramatically in recent years. Today, around ten thousand working satellites, many no larger than a loaf of bread, circle the planet. Some operate in constellations that can revisit the same location multiple times a day and offer image resolutions so sharp they can identify different car models driving on a road.

KEY CHAPTER TAKEAWAYS

- A burgeoning "NewSpace" economy driven by private innovation and investment is transforming space launch, vehicles, communications, and key space actors in a domain that has until now been dominated by superpower governments.

- Space is a finite planetary resource. Because of dramatic increases in satellites, debris, and geopolitical space competition, new technologies and new international policy frameworks will be needed to prevent and manage international conflict in space and ensure responsible stewardship of this global commons.

- A race to establish a permanent human presence on the Moon is underway, with serious concerns that, despite Outer Space Treaty prohibitions against it, the first nation to reach the Moon may be in a strong position to prevent others from establishing their own lunar presences.

Sustainable Energy Technologies

This vital strategic resource for nations typically involves generation, transmission, and storage. In recent years it has also come to include carbon capture and carbon's removal from the atmosphere. Energy mix and innovation are key to efforts to address climate change. Success will also depend on tackling challenges such as decentralizing and modernizing electricity grids and achieving greater national consensus about energy goals to enable strategic and effective R&D programs and funding.

KEY CHAPTER TAKEAWAYS

- Although many clean energy technologies are now available and increasingly affordable, scaling them to a meaningful degree and building the massive infrastructure needed to deploy them will take decades.

- The largest impact on reducing emissions in the near to medium term will come from building a no- to very-low-emission electricity grid, electrifying passenger cars and small commercial vehicles, and transitioning residential and commercial heating and industrial energy.

- In the long term, technologies for decarbonizing buses and long-haul trucks, decarbonizing carbon-intensive industries, and reducing greenhouse gases from refrigerants and agriculture will play key roles in a net-zero, emissions-free energy infrastructure.

Important Crosscutting Themes

Chapter 11 discusses fourteen themes that cut across the technological areas. We split these themes into two categories.

Category 1: Key Observations About How Technologies Evolve over Time

1. The Goldilocks challenge: moving too quickly, moving too slowly. Innovation that emerges too fast threatens to disrupt the status quo around which many national, organizational, and personal interests have coalesced. It is also more likely to lead to unintended consequences and give short shrift to security, safety, ethics, and geopolitics. Innovation that moves too slowly increases the likelihood that a nation will lose the technical, economic, and national security advantages that often accrue to first movers in a field.

2. There is a trend toward increasing access to new technologies worldwide. Even innovations that are US born are unlikely to remain in the exclusive control of American actors for long periods.

3. The synergies between different technologies are large and growing. Advances in one technology domain often support advances in other technologies.

4. The path from research to application is often not linear. Many believe that technological breakthroughs arise from a step-by-step linear progression where basic research leads to applied research, which then leads to development and prototyping and finally to a marketable product. Yet innovation often does not work this way. Many scientific developments enhance understanding but never advance to the marketplace. Many marketable products emerge in a nonlinear fashion, after many rounds of feedback *between* phases. Other products emerge only when several different technologies reach some level of maturity.

5. The speed of change is hard even for leading researchers to anticipate. Technology often progresses in fits and starts, with long periods of incremental results followed by sudden breakthroughs.

6. Nontechnical factors often determine whether new technologies succeed or fail. Adoption of novel technologies hinges on economic viability and societal acceptability, not just scientific proof of concept and engineering feasibility.

7. The US government is no longer the primary driver of technological innovation or funder of research and development. Historically, technological advances (including semiconductors, the internet, and jet engines) were funded and advocated for by the US government. Today, private-sector R&D investment is playing a much larger role, raising important concerns about how to ensure that US national interests are properly taken into account and that basic science—which is an important foundation for future innovation—remains strong.

8. Technological innovation occurs in both democracies and autocracies, but different regime types enjoy different advantages and challenges. Democracies provide greater freedom for exploration, while authoritarian regimes can direct sustained funding and focus toward the technologies they believe are most important.

Category 2: Common Innovation Enablers and Inhibitors

1. Ideas and human talent play a central role in scientific discovery and cannot be manufactured at will. They must be either domestically nurtured or imported from abroad. Today, both paths for generating ideas and human talent face serious and rising challenges.

2. A policy bias toward science or technology at the frontiers of knowledge tends to overestimate the benefits accruing from such advances, at least in the short term. Many technologies with transformational potential are not necessarily on the technical frontier, and frontier bias carries with it the risk of overlooking older technologies that can be used in novel and impactful ways.

3. Good public policy anticipates wide variations in perspectives on any given technology. When everyone in a decision-making organization shares similar perspectives on technology—creating analytical blind spots and potentially groupthink, or unwarranted conformity in beliefs—the risks associated with innovation can be underestimated.

4. US universities play a pivotal role in the innovation ecosystem that is increasingly at risk. Although the US government frequently talks about the importance of public-private partnerships in emerging technology, universities also play a pivotal and often underappreciated role. They are the only organizations with the mission of pursuing high-risk research that may not pay off commercially for a long time, if ever. That high-risk focus has yielded high-benefit payoffs in a wide range of fields.

5. Sustaining American innovation requires long-term government R&D. Investments with clear strategies and sustained priorities—not the increasingly common wild swings in research funding from year to year—are crucial.

6. Cybersecurity is an enduring concern for every aspect of emerging technology research. State and nonstate actors will continue to threaten the confidentiality, integrity, and availability of information that is crucial for emerging technology R&D.

Finally, each of the ten technology fields covered in this report bears on five policy areas that are of interest to policymakers: economic growth, national security, environmental and energy sustainability, health and medicine, and civil society. Chapter 12 identifies applications and consequences of each field as they apply to these policy areas.

NOTES

1. Andrew Pettegree and Arthur der Weduwen, *The Bookshop of the World: Making and Trading Books in the Dutch Golden Age* (New Haven, CT: Yale University Press, 2019), 70–72.

INTRODUCTION

The Role of Science and Technology in Advancing National Interests

Vannevar Bush, an engineer and policymaker who oversaw the development of the Manhattan Project, was the nation's first presidential science advisor. In 1945, he wrote, "Advances in science when put to practical use mean more jobs, higher wages, shorter hours, more abundant crops, more leisure for recreation, for study, for learning how to live without the deadening drudgery which has been the burden of the common man for ages past. . . . Advances in science will also bring higher standards of living, will lead to the prevention or cure of diseases, will promote conservation of our limited national resources, and will assure means of defense against aggression."[1]

Science and technology (S&T) remain essential to our national interests. Advances in S&T are closely tied to national needs in transportation, agriculture, communication, energy, education, environment, health, and defense—as well as to millions of American jobs. S&T also underpins and drives many strategic objectives in foreign policy, such as reducing the proliferation of weapons of mass destruction, strengthening relationships with allies and partners, improving humanitarian assistance, and promoting growth in developing and transitional economies.[2] Research and development in S&T fields such as information technology, biotechnology, materials science, and lasers will impact both "hard power" issues—defense, arms control, nonproliferation—and "soft power" concerns, such as climate change, infectious and chronic diseases, energy supply and demand, and sustainable development.[3]

S&T is one important battleground for seeking advantage in geopolitical competition, as advances in scientific and technical fields can contribute to national interests, including a stronger national security posture, greater national pride and self-confidence, economic influence, and diplomatic leverage. But four other points about S&T are equally important:

- Advances in S&T must be leveraged alongside strong public policy if those advances are to serve the national interest. Coupling advanced technology with poor policy to influence that technology rarely ends well.

- Advantages gained from S&T advances are transient in the long run. Attempting to restrict the transfer of scientific and technical knowledge to other nations may delay its spread, but the first successful demonstration of a technological advance on the part of the United States is often the impetus for other nations to launch their own efforts to catch up.

- Internationally, S&T is not always a zero-sum game, as advances originating in one nation often benefit others. For example, the internet and what most people know simply as GPS navigation are US-born innovations whose uses have spread around the world—and the United States itself has gained from that spread.

- International competition does not occur only with adversaries. Our allies and partners also compete in the S&T space, developing technology or deploying policy that can leave the United States at a disadvantage.

Policy for Science and Technology

Policymakers have a wide variety of tools to influence the conduct of S&T research and development. Many of these are obvious, such as research funding, tax incentives to firms, intellectual property rights, export controls, classification authority,[4] regulation, public procurement, funding and other aid to strategic sectors, as well as labor force training and education.

On the other hand, policy need not be directed at S&T to have a meaningful impact. For example, immigration policy is not primarily directed at the S&T workforce, but it can have profound effects on the talent available to academic and industry research. Policy oriented in one direction attracts talent to the United States, while policy oriented in another diminishes such talent. Or consider the national economic environment. Stable fiscal and monetary policies make it easier for private-sector decision makers to plan and invest for the long term—a critical consideration when many S&T advances must be nurtured along an extended path from conception to maturity.

Ten Science and Technology Fields

Chapters 1 through 10 describe in more detail ten S&T fields important to the national agenda. Our selection of these fields was driven by several factors: inclusion on common lists of key technologies developed by government, the private sector, and academia and think tanks, as well as discussions with science and engineering colleagues at Stanford University and other research universities. We do not claim that any one of these ten is more important than the others, and the discussion below addresses subjects in alphabetical order. Indeed, one of the unexpected aspects of this technological moment is convergence: New technologies are intersecting, overlapping, and driving each other in all sorts of ways—some obvious, some more hidden.

The description of each field is divided into three parts. The first part is an **overview** of the field. The second part addresses noteworthy **key developments** in the domain that are relevant to understanding the field from a policy perspective. The last part, providing an **over-the-horizon** perspective, is itself subdivided into three sections: the potential impact of the field in the future (i.e., the field's potential over-the-horizon impact); the likely challenges facing innovation and implementation; and relevant policy, legal, and regulatory issues.

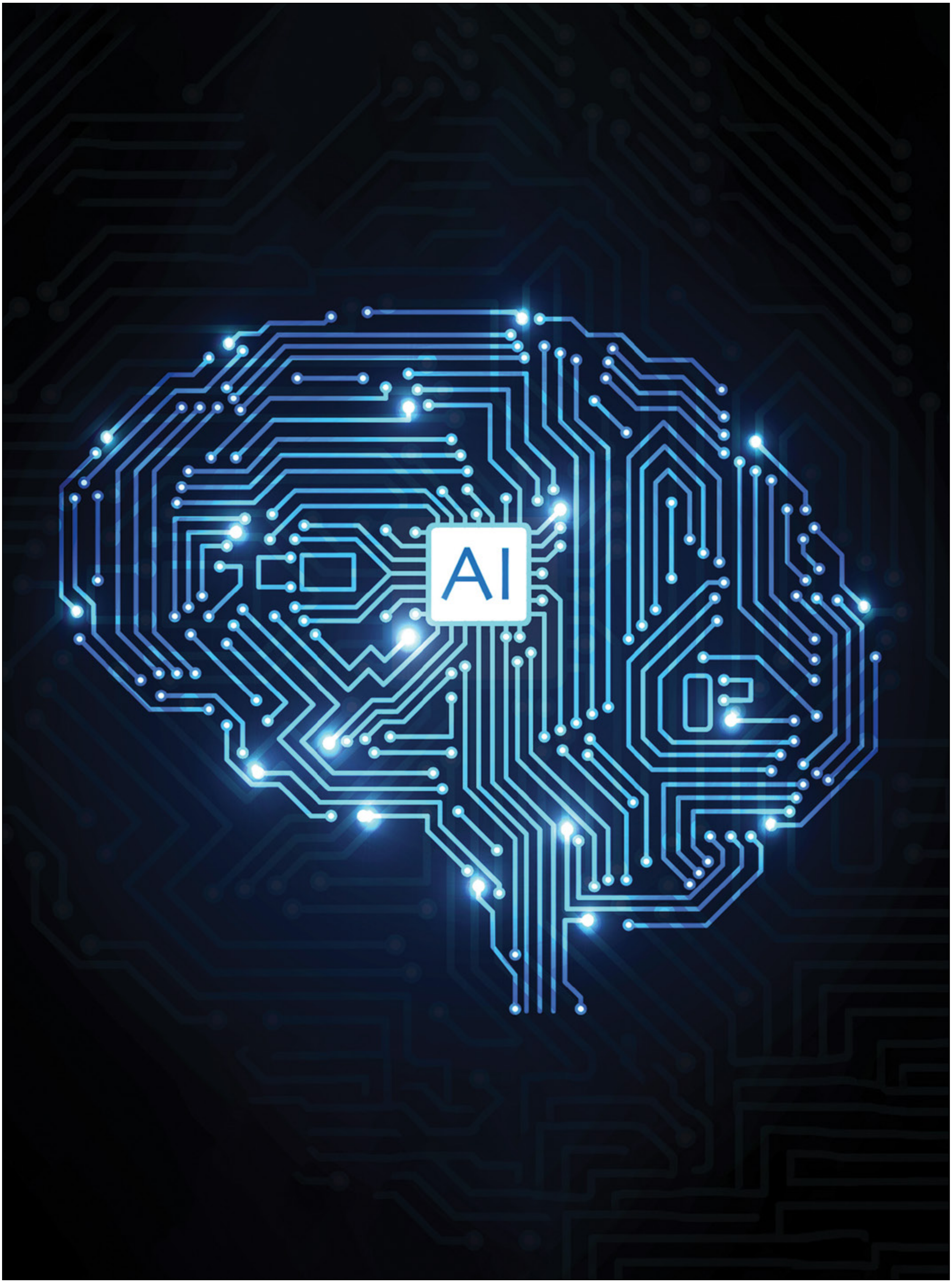
NOTES

1. Vannevar Bush, *Science: The Endless Frontier* (Washington, DC: US Government Printing Office, 1945), https://www.nsf.gov/od/lpa/nsf50/vbush1945.htm.

2. National Research Council, *The Pervasive Role of Science, Technology, and Health in Foreign Policy: Imperatives for the Department of State* (Washington, DC: National Academies Press, 1999), https://doi.org/10.17226/9688.

3. National Intelligence Council, *Global Trends 2015: A Dialogue About the Future with Nongovernment Officials*, December 2000, https://www.dni.gov/files/documents/Global%20Trends_2015%20Report.pdf; National Intelligence Council, *Mapping the Global Future: Report of the National Intelligence Council's 2020 Project*, December 2004, https://www.dni.gov/files/documents/Global%20Trends_Mapping%20the%20Global%20Future%202020%20Project.pdf.

4. Under some circumstances (such as when federal funding is involved), the US government may have the authority to classify research even if that research was performed without access to classified information.


AI

01

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 decrease[5] 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 compute[12]—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 capabilities[33] 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, BigScience'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

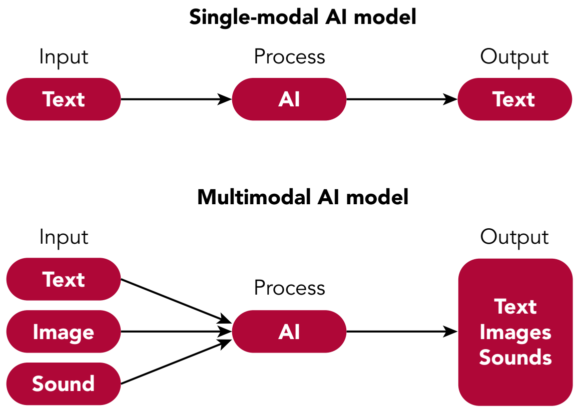
AI systems that incorporate multiple modalities—text, images, and sound—within single models are becoming increasingly popular. This multimodal approach, shown in figure 1.1, aims to create more humanlike experiences by leveraging various senses such as sight, speech, and hearing to mirror how humans 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 year[41]—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 NAIRR effort, investments from high-tech companies for AI exceeded $27 billion in 2023 alone.[44]

Over the Horizon

Impact of New AI Technologies

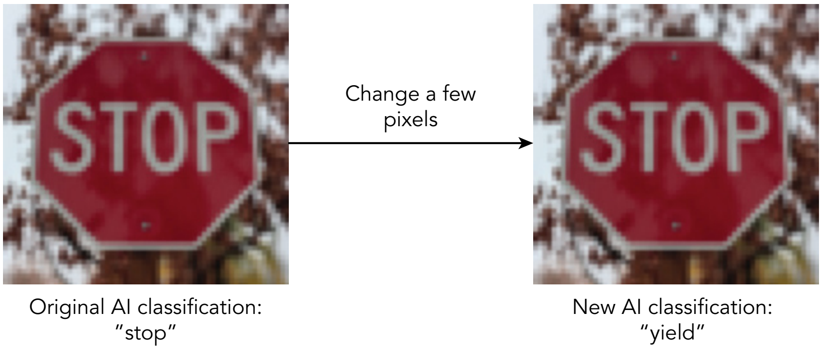
Potential positive impacts of new AI technologies are most likely to be seen in the applications they enable for societal use, as described in detail above. On the other hand, no technology is an unalloyed good. Potential negative impacts from AI will likely emerge from known problems with current state-of-the-art 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

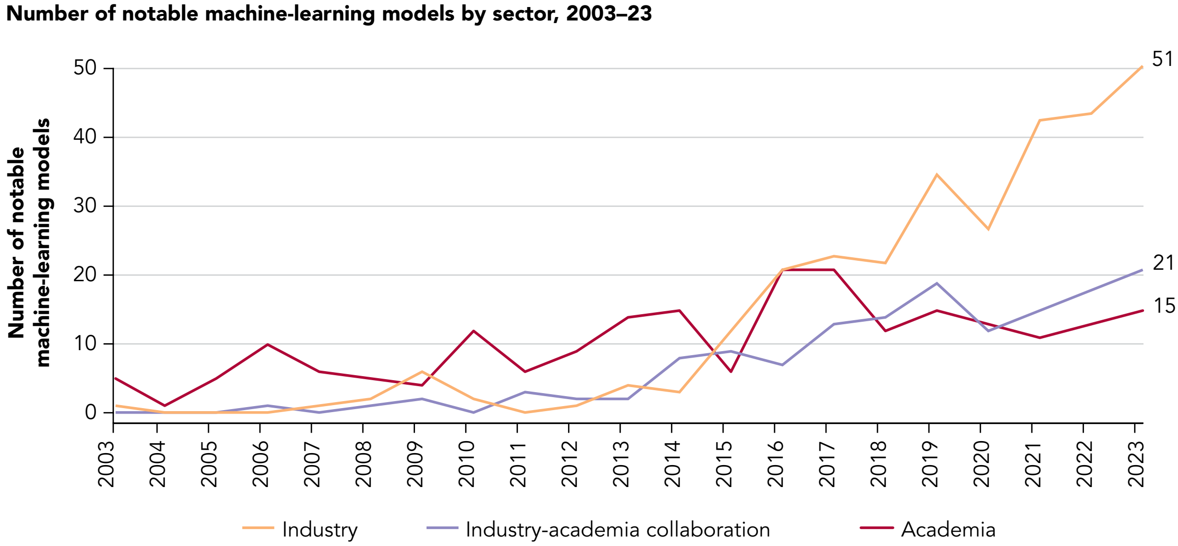
The United States is eating its seed corn with respect to the AI talent pool. As noted in the Foreword, faculty at Stanford and other universities report that the number of students studying in AI who are joining the industry, particularly start-ups, is increasing at the expense of those pursuing academic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



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-robert-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 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.

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

02

BIOTECHNOLOGY AND SYNTHETIC BIOLOGY

KEY TAKEAWAYS

- Biotechnology is poised to emerge as a general-purpose technology by which anything bioengineers learn to encode in DNA can be grown whenever and wherever needed—essentially enabling the production of a wide range of products through biological processes across multiple sectors.
- The US government is still working to grasp the scale of this bio-opportunity and has relied too heavily on private-sector investment to support the foundational technology innovation needed to unlock and sustain progress.
- Biotechnology is one of the most important areas of technological competition between the United States and China, and China is investing considerably more resources. Lacking equivalent efforts domestically, the United States runs the risk of Sputnik-like strategic surprises in biotechnology.

Overview

Biotechnology involves using living systems and organisms to develop or make products and solve problems. First-generation biotechnology arose over millennia and involved the domestication and selective breeding of plants and animals[1] for agriculture, food production, companionship, and other purposes.[2] Second-generation biotechnology was launched a half century ago with the invention of recombinant DNA[3] and has since encompassed techniques such as genetic engineering, polymerase chain reaction (PCR), high-throughput DNA sequencing, and CRISPR gene-editing technology.[4] Both breeding and editing approaches continue to advance, creating and using ever better tools for sculpting[5] and editing[6] living systems.

Biotechnology products and services realized through breeding and editing are already widely deployed. A 2020 National Academies of Sciences, Engineering, and Medicine report valued the US bioeconomy at around 5 percent of GDP, or more than $950 billion annually.[7] Existing applications involve primarily agriculture, medicines, and industrial materials.[8] A 2020 McKinsey & Company report noted that hundreds of biotechnology projects were under development and estimated that the resulting products could add $2 to $4 trillion in annual economic impact within the next two decades.[9] This projected doubling of the bioeconomy's contribution to worldwide GDP every seven years would match biotechnology's economic track record.[10] The McKinsey report concluded that, ultimately, biomanufacturing could account for around 60 percent of the global economy's physical inputs.[11]

Biology, as a natural manufacturing process, is remarkably distributed and localized. For example, leaves on trees do not come from factories or central facilities; rather, they grow on trees themselves—all over the place. Yet, outside of agriculture, biotechnology has until now been largely practiced and commercialized in a capital-intensive, industrialized, and centralized context.[12] This contrast between biology as a naturally distributed platform and industrialized biomanufacturing processes suggests that biotechnology may be ripe for new modes of practice and products.

Notably, synthetic biology continues to emerge as an important new approach within biotechnology. Synthetic biology combines principles from biology, engineering, and computer science to modify living systems and construct new ones by developing novel biological functions, such as custom metabolic or genetic networks, novel amino acids and proteins, and even entire cells. These new functions are performed through the construction of engineered biological parts that can be reused by humans when appropriate, thereby reducing the need for each project to start from scratch. Synthetic biology thus helps us to create more complex, biologically based systems, including those with functions that do not exist in nature.

In thinking about biotechnology's potential, it is instructive to consider the evolution of information technology over the past several decades. Fifty years ago, computers were mostly industrial, disconnected, and centralized.[13] The emergence of personal computers, packet-switching networks,[14] and programming languages that made computing accessible and fun[15] changed how information science and technologies developed and led to decentralized access to computing and information at unprecedented speed and scale.[16] Biology could experience the same transformation within the next two decades—manufacturing processes could move from being largely invisible to being obvious and apparent to people as they begin to manipulate some of these workflows for themselves.

Key Developments

Distributed Biomanufacturing

The significance of distributed biomanufacturing lies in its flexibility, both in location and timing. Because the apparatus for a fermentation process can be established wherever there is access to sugar and electricity, a production site can be set up almost anywhere. The timing aspect is equally transformative: By removing the need to grow feedstocks, biomanufacturers can swiftly respond to sudden demands, such as a rapid outbreak of disease requiring specific medications. This adaptability not only enhances efficiency but also revolutionizes how we approach manufacturing, making it far more responsive to urgent needs than traditional methods.

In an important demonstration illustrating that distributed biomanufacturing is not a mirage, the synthetic biology company Antheia reported in early 2024 that it had completed validation of a fermentation-based process for brewing thebaine, a key starting material used in treating opioid overdoses with Narcan.[17] The company partnered with Olon, an Italian contract manufacturing organization. Antheia's bioengineered yeast strain was sent to Olon's large-scale fermentation facility in

Italy. Working together, they repeatedly brewed 116,000-liter batches of bioengineered yeast, with each batch making broth containing a metric ton of thebaine—roughly enough for one hundred million Narcan doses.[18] This demonstration highlights the potential for on-demand production of critical pharmaceuticals, potentially revolutionizing drug supply chains and improving access to essential medicines.

In 2022, Chinese researcher Chenwang Tang and colleagues noted more generally how synthetic biology allows the rewiring of biological systems to support portable, on-site, and on-demand manufacturing of biomolecules.[19] In 2024, as one of many pioneering examples, Stanford researchers reported on-demand bioproduction of sensors enabling point-of-care health monitoring and detection of environmental hazards aboard the International Space Station.[20] They had already realized many similar demonstrations of distributed biomanufacturing on Earth, ranging from biotechnology educational kits to the production of conjugate vaccines used to stimulate stronger immune responses.[21]

These are just a few examples demonstrating how biotechnology can be used to make valuable products and services locally. Viewed from a traditional perspective, what's happening is a sort of molecular gardening: The energy and material inputs needed to make the biotechnology products are supplied locally, but the process differs from conventional gardening in that the genetic instructions for what the biology should do or make are being programmed by bioengineers. To fully unlock the power of distributed biomanufacturing, it must also become possible to make the physical DNA used to encode the genetic programs locally.

Distributed DNA Reading and Writing

DNA is physical material that encodes biological functions in natural living systems. It is often represented abstractly by its four constituent bases (A, C, T, and G), also known as nucleotides. Unique orderings of these bases encode different biomolecules, which in turn underlie different cellular behaviors and functions.

DNA sequencing (i.e., reading of DNA) and synthesis (i.e., writing of DNA) are two foundational technologies underlying synthetic biology.[22] Sequencers are machines that determine the precise order of nucleotides in a DNA molecule, effectively converting genetic information from a physical to a digital format. Synthesizers generate user-specified digital sequences of A's, C's, T's, and G's, creating physical genetic material from scratch that encodes the user-specified sequence, thus effectively transforming bits into atoms. If DNA reading and writing tools could themselves be distributed, anyone with an internet connection could upload and download application-specific DNA programs that direct distributed biomanufacturing processes powered by locally available energy and supplied by locally available materials.

In the 1990s, public funding for sequencing the human genome jump-started advances in DNA-sequencing tools by creating significant demand for reading DNA.[23] Private capital and entrepreneurs quickly responded.[24] The Human Genome Project (HGP) favored development of DNA sequencers that could read billions of bases of DNA as cheaply as possible, resulting in large-format DNA sequencers that were organized in centralized DNA-sequencing factories.[25] A complementary approach to DNA sequencing has since matured that allows for individual DNA molecules to be sequenced via tiny pores, or nanopores, in ultra-thin membranes.[26] UK-based Oxford Nanopore Technologies has exploited this approach to market small-format, portable DNA sequencers that can be used with laptop computers, allowing DNA sequencing to become a distributed technology (see figure 2.1).[27]

The market for DNA synthesis has developed organically over the past forty-five years.[28] So far, there has been no equivalent to the HGP that has resulted in significant public funding from Western governments for improving the technology of DNA synthesis.[29]

FIGURE 2.1 Portable DNA sequencers enable biotechnology to become more distributed

Source: Oxford Nanopore Technologies, 2024

Improvements in DNA synthesis in Western countries have been sporadic and dependent primarily on private capital.[30] Commercially available gene-length DNA-synthesis services in the United States have improved only modestly in the past six years.[31] Today, most DNA synthesis is carried out via centralized factories.[32] Customers order DNA online and receive it via express shipping; it typically takes these factories from days to weeks to make the DNA molecules themselves.

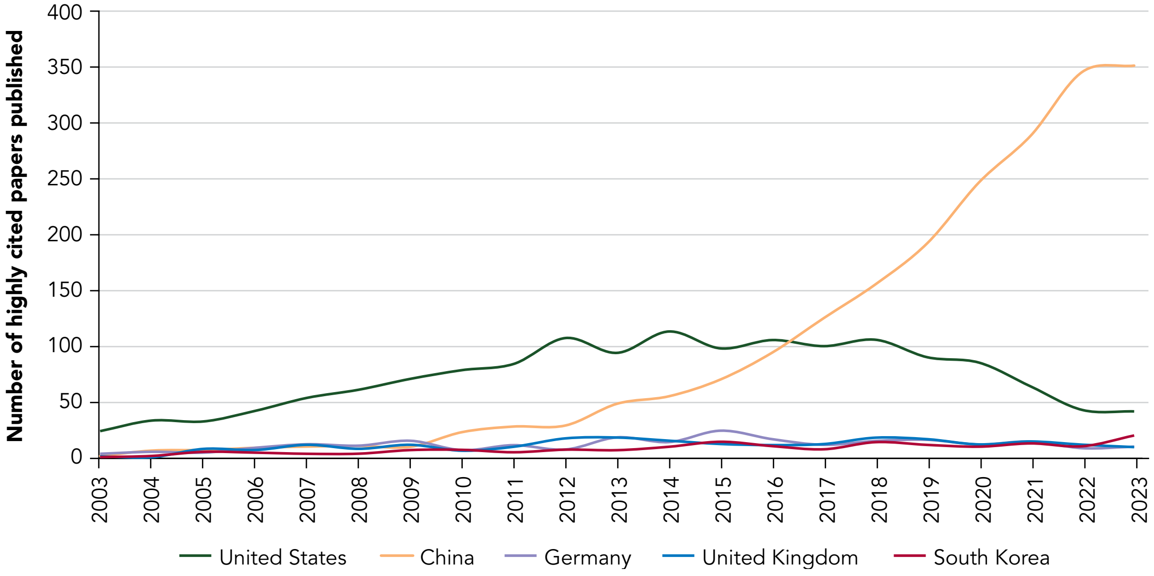
A new generation of companies is pursuing novel approaches to building DNA—most notably enzymatic DNA synthesis, which uses enzymes and simpler chemical inputs to build DNA.[33] These new approaches support hardware and reagent formats that could potentially enable fast, reliable, and distributed DNA synthesis. However, the creation of widely distributed DNA printers is not receiving significant public support, and existing private investments may not be sufficient to make the technology real in a practical sense.

Meanwhile, researchers in China have had the resources to advance gene and genome synthesis. For example, the first synthetic plant chromosome was reported by Chinese researcher Yuling Jiao and colleagues in January 2024.[34] More broadly, researchers in China published nearly 350 papers that ranked among the top 10 percent most-cited papers on synthetic biology in 2023, compared to 41 such papers in the United States (see figure 2.2).

Biology as a General-Purpose Technology

Biotechnology is currently used to make medicines, foods, and a relatively narrow range of sustainable materials. However, as noted earlier, anything whose biosynthesis engineers can learn to encode in DNA could be grown using biology.

FIGURE 2.2 China is outpacing the United States in publishing highly cited research papers on synthetic biology



Source: Adapted from Australian Strategic Policy Institute, Critical Technology Tracker, based on "Appendix 2: Detailed Methodology," in Jennifer Wong-Leung, Stephan Robin, and Danielle Cave, ASPI's Two-Decade Critical Technology Tracker, August 2024

Examples from nature highlight the potential here: Some bacteria are capable of growing arrays of tiny magnets,[35] while select sea sponges grow glass filaments with a refraction index—which determines the speed at which light travels through a medium—similar to that of human-made fiber-optic cables.[36] These bio-made magnets and filaments are created under ambient conditions through naturally sustainable processes and are often more robust than traditionally manufactured alternatives. These and other examples have inspired calls for biology to be recognized as a general-purpose technology that, with appropriate vision and leadership, could become the foundation of a much more resilient manufacturing base.[37]

As an example of the vision that's needed, in 2018, the Semiconductor Research Corporation (SRC) outlined an ambitious twenty-year synthetic biology road map.[38] SRC's first proposed step was to develop DNA as a data storage medium.[39] Such an approach was demonstrated in 2024 when the Hoover Institution Library & Archives partnered with Twist Bioscience to encode a digital copy of the telegram from President Hoover founding his namesake institution within synthetic DNA contained in a tiny ampule (see figure 2.3). Made in this way, the DNA serves as a data storage medium whose digital contents must be recovered via DNA sequencing.

The ultimate goal of SRC's road map is to enable bottom-up construction of microprocessors. To fully realize this goal might cost $100 billion in foundational research investment, smartly managed over twenty years, and as yet there is no such coordinated effort underway. However, the novel notion of growing computers already challenges many framing assumptions and realities underlying contemporary geopolitics.[40] Concerns about computer manufacturing and supply chains presume that making computers is hard. What if making them becomes as easy as growing zucchini?

FIGURE 2.3 DNA is used as a storage medium for a digital copy of Herbert Hoover's telegram founding his namesake institution

A more long-standing example of biology as an increasingly general-purpose technology with potential geopolitical impacts can be found in the 2011 US Navy program, Application of Synthetic Biological Techniques for Energetic Materials.[41] This program began exploring the ability to brew propellants and explosives through a process akin to how Antheia and Olon partnered to brew medicines—an ability that could enable any nation anywhere to create more resilient supply chains for key military materials. A distributed and resilient biomanufacturing network could, for example, help NATO members meet their Article 3 obligations related to supply chain resilience.[42] A bio-based approach to brewing fuels could also help meet climate and sustainability goals.[43]

Pervasive and Embedded Biotechnologies

Most modern biotechnology products and applications are presumed to be destined for deployments carefully contained within steel tanks or constrained by doctors' prescriptions. However, recent developments in consumer access to these products and applications suggest that this will not always remain the case. A US company called Light Bio, for example, now sells petunia plants bioengineered to emit light (see figure 2.4).[44] Light Bio's offering represents one of the first successful launches of a live consumer biologic, enabling anyone in the United States to source and keep a bioengineered organism for personal use.

In 2024, UK-based Norfolk Plant Sciences first made available to US consumers seeds for its purple tomato, a kind of tomato bioengineered to produce high levels of antioxidants thought to help prevent cancer (see figure 2.5).[45] Stanford faculty bought seeds, and soon bioengineered tomatoes were growing in gardens across campus. Indeed, these tomatoes are available for consumer purchase in a number of grocery stores in the American Southeast.[46]

Another category of pervasive and potentially consumer-facing biotechnology involves bioengineering the bacteria that live on our skin and inside our bodies, as well as within the environment around us. For example, in 2023, Stanford researchers pioneered the bioengineering of skin microbes to combat skin cancer.[47] They have since expanded such work to enable the eliciting of more broadly antigen-specific T cells, which target and eliminate cells infected with viruses and bacteria, as well as cancerous cells. T cells also play a role in providing long-term immunological memory.[48] In addition, researchers have identified specific odorants produced by human-skin microbes whose production could be modulated to reduce mosquito bites[49] and have also developed methods for bioengineering microbes to improve gut health.[50]

As these examples suggest, twenty-first-century biotechnologies may increasingly be deployed in, on, and around us—and be made available through established and far-reaching consumer channels.

Biological Large Language Models

In the 2023 edition of *The Stanford Emerging Technology Review*, we discussed how researchers had developed and deployed methods that are based on artificial intelligence (AI) to predict the three-dimensional structures of over 200 million natural proteins,[51] an accomplishment recently recognized via the 2024 Nobel Prize in Chemistry.[52] Anybody with a laptop can now take a DNA sequence encoding a protein and quickly estimate its expected shape. The shape of a protein helps determine its placement and function in a living system. The ability to rapidly generate predicted shapes helps bioengineers modify existing proteins and design new ones from scratch. However, the work of modifying an existing protein sequence and designing a new protein still requires direct human genius and labor.

FIGURE 2.4 Light Bio's petunias are bioengineered to emit light

Source: Light Bio Inc.

FIGURE 2.5 Norfolk Plant Sciences has bioengineered a purple tomato

Source: Norfolk Plant Sciences

Advances in AI may change that. In 2024, new large language models (LLMs) have emerged that are trained on natural DNA, RNA, and protein sequences. For context, ChatGPT and similar LLMs, when trained on sequences of letters and words from composing human languages like English, can generate meaningful new human-readable text. In similar fashion, biological LLMs (bioLLMs), trained on vast datasets of biological sequences, can generate novel sequences with potential biological functions, accelerating the design process in fields like protein engineering and synthetic biology. For example, in early 2024 Stanford researchers reported developing and using a general protein language model to quickly design better virus-neutralizing antibodies targeting Ebola and SARS-CoV-2.[53] Unlike widespread speculative concerns about the destabilizing potential of the use of AI in biotechnology,[54] actual known work in the field seems to instead have directly contributed to public health and biosecurity.

As a second example, researchers at Stanford released a genomic foundation model named Evo that performs prediction and generation tasks across DNA, RNA, and proteins.[55] (Foundation models are discussed in chapter 1 on artificial intelligence.) They then used Evo to help design synthetic gene-editing systems. DNA, RNA, protein, gene, and genome language models will continue to emerge and develop throughout the 2020s. The greatest bottleneck will likely be the limited capacity available to build and test the biological sequences generated by the models. Any adult English speaker can quickly read a passage of LLM-generated English text and evaluate its purpose and quality. For now, only living systems themselves can ultimately interpret and establish whether the function and performance of a bioLLM-generated design actually works as expected. The ability to operate platforms that scale high-throughput testing of bioLLM designs is a significant advantage in inventing, improving, and offering world-leading foundation models in biology and biotechnology.

Over the Horizon

Routinization of Cellular-Scale Engineering

There is no natural cell on Earth that is fully understood. Even for well-studied model organisms like *E. coli*, there remain genes with unknown or incompletely understood functions, highlighting the complexity of cellular systems. The microbes that have been subject to the most intense study still require more than seventy genes whose functions no researcher understands.[56] Each gene encodes some unknown life-essential mechanism. Our collective ignorance means that all bioengineering workflows remain Edisonian at the cellular scale—we are tinkering and testing. Bioengineering students are taught the mantra "design, build, test, learn,"[57] where the test portion implies a very large amount of empirical lab work to understand basic phenomenology. By contrast, the routinization of bioengineering workflows at the cellular scale sufficient to realize "design, build, work" workflows—a hallmark of all other modern technologies that implies doing a relatively small amount of empirical work primarily to validate the analysis underlying the construction of a biological artifact—remains

The ability to construct life for the first time, without being restricted to any terrestrial lineage, is akin to launching to orbit the first artificial satellite.

fringe foundational research. Consequently, such bioengineering workflows remain in their earliest stages.[58]

Nevertheless, because cells are the fundamental unit of life, researchers[59] and start-ups[60] across the United States, Europe, Japan, and China are scrambling to learn how to build fully understandable cells from scratch. The ability to construct life for the first time, without being restricted to any terrestrial lineage, is akin to launching to orbit the first artificial satellite. Just as rockets allow us to ascend Earth's gravity well, giving us access to the privilege, perspective, and power of space, the ability to transcend the constraints of Earth's existing life-forms[61]—organisms constrained by lineage and the requirements of reproduction and evolvability—will unlock the next level of biotechnologies, providing a powerful perch from which to access everything that biology can become.

A first organized and professional attempt to construct life from scratch will likely cost $100 million. The Institute of Synthetic Biology (ISB) at the Shenzhen Institute of Advanced Technology in China is one organization where such an effort could now be carried out rapidly. The ISB hosted a global summit on coordination of synthetic-cell building in October 2024.[62] Lacking equivalent efforts domestically, the United States is risking a Sputnik-like biotechnology surprise.[63]

Electrobiosynthesis

Carbon is central to life. Currently, we rely on photosynthesis for production of organic carbon molecules. Recent thinking, however, suggests that electricity could be used to fix carbon directly from the air to create organic molecules that could be fed to microbes—a process that may come to be known as electrobiosynthesis or, more simply, "eBio"—and that doing so could be an order of magnitude more efficient from a land-use perspective than traditional agriculture.[64]

In other words, the idea is to engineer a parallel carbon cycle that starts with air and electricity, perhaps generated via solar panels, to create organic molecules that can power bioproduction processes. For example, in August 2024, Stanford researchers reported the creation of a system that combines electrochemistry with biological processes that do not use cells to transform simple carbon compounds into a key organic molecule called acetyl-CoA, which is present in all living things and acts as a building block for other molecules within cells.[65]

Although eBio is still a very immature technology, its potential significance and impacts are hard to overstate. For example, surplus power from large-scale renewable energy generation could be used to directly produce biomolecules such as proteins and cellulose without requiring massive conventional battery banks to store energy that cannot be used immediately. The development of eBio could also enable bioproduction in places where soils are poor, water is scarce, or climate and weather are too uncertain. And it could raise the ceiling on how much humanity could make in partnership with biology. We would be constrained only by how much energy we can generate for such purposes. This approach could significantly reduce the land and water requirements for biomass production, potentially alleviating pressure on agricultural resources and offering a more sustainable path for biomanufacturing.

Challenges of Innovation and Implementation

Many first-generation synthetic biology companies continue to struggle.[66] Billions of dollars of private capital have been lost in biotechnology investments made with the best of intentions in the United States alone over the past two decades. One perspective is that these early big bets were simply too early.[67] The hope is that smaller and scrappy next-generation efforts will find their way to success. However, an immediate short-term issue is that many sources of private capital funding to support these next-generation commercial efforts are now shut off for synthetic biology, adding headwinds to the general challenges of obtaining capital that young, innovative businesses face.

Another perspective is that America has relied too heavily on the private sector to invent, advance, and deploy emerging biotechnologies. The biotech equivalent of the publicly funded tooling and infrastructure development in the early days of US strategic computing and networking programs is today pursued only via private investment and commercial platforms. Because private investors expect these foundational tools and platforms to quickly generate and sustain revenue growth to justify further funding, businesses developing them often fail repeatedly.

Breaking this cycle will require smart and sustained public investments in foundational bioengineering research, from tools for measuring, modeling, and making biology to public-benefit research platforms. The National Science Foundation's August 2024 investment in five academic biofoundries may be one small step forward in this respect.[68]

Policy, Legal, and Regulatory Issues

Safety and national security concerns New organisms not found in nature raise concerns about how they will interact with natural and human environments. For instance, bioengineered organisms that escape into the environment and possibly disrupt local food chains or natural species have long been a concern. Moreover, as the science and technology of synthetic biology becomes increasingly available to state and nonstate entities, there are legitimate concerns that malicious actors will create organisms harmful to people and the environment.[69]

Ethical considerations Different religious traditions may have different stances toward life and whether the engineering of new life-forms violates any of their basic precepts. Often classified as potential nonphysical impacts, the effects on biotechnology when considering these religious concerns are sometimes difficult to predict in advance. In the words of a Wilson Center report on this topic, such concerns involve "the possibility of harm to deeply held (if sometimes hard to articulate) views about what is right or good, including . . . the appropriate relationship of humans to themselves and the natural world."[70]

The United States and other nations are working hard to develop, advance, and refine strategies for biotechnology, biomanufacturing innovation, biosecurity, and the bioeconomy overall. For example, the United States' National Security Commission on Emerging Biotechnology continues its work.[71] The congressionally mandated Department of Defense Task Force on Emerging Biotechnologies and National Security is also underway.[72] Both efforts are expected to produce substantial reports and products throughout 2025, complementing activities ongoing within the executive office of the president, including work as ordered by Executive Order 14081 on biotechnology and by Executive Order 14110 on artificial intelligence. Internationally, the Organisation for Economic Co-operation and Development's Global Forum on Technology selected synthetic biology as one of three key initial technologies to focus on, with work now well underway.[73] The World Economic Forum has also renewed its Global Futures Council on Synthetic Biology, which continues its work.[74]

One overall challenge for policymakers—and the biotechnology community—is to preserve and advance the very significant public benefits of research into biosciences and biotechnology while minimizing the real and perceived risks associated with potential misuse of the resulting knowledge and capacities. For example, in response to the concern about the escape of harmful bioengineered organisms into the environment, synthetic biology itself offers the possibility of bioengineering organisms from scratch that are incapable of escaping or evolving.[75] But it is a matter of policy to ensure that necessary safeguards are included in projects intended to create new organisms.

In short, policymakers will have to be aware of—and able to navigate—issues and aspects of emerging biotechnologies, such as the ones included in this section, if they are to help guide the development of the field and the increasing diversity of the biotechnologies that emerge from it.

NOTES

1. Jared Diamond, "Evolution, Consequences, and Future of Plant and Animal Domestication," *Nature* 418 (August 2020): 700–707, https://doi.org/10.1038/nature01019.

2. Freeman Dyson, "Our Biotech Future," *New York Review of Books* 54, no. 12 (July 2007): 4–7, https://www.nybooks.com/articles/2007/07/19/our-biotech-future/.

3. Tim Beardsley, "Biotechnology: Cohen-Boyer Patent Finally Confirmed," *Nature* 311, no. 5981 (September 1984): 3, https://doi.org/10.1038/311003a0.

4. Irina Gostimskaya, "CRISPR-Cas9: A History of Its Discovery and Ethical Considerations of Its Use in Genome Editing," *Biochemistry* (Moscow) 87, no. 8 (August 2022): 777–78, https://doi.org/10.1134/S0006297922080090. CRISPR is an acronym standing for clustered regularly interspaced short palindromic repeats.

5. Frances H. Arnold, "Innovation by Evolution: Bringing New Chemistry to Life," December 8, 2018, Stockholm University, Stockholm, Sweden, PDF, https://www.nobelprize.org/prizes/chemistry/2018/arnold/lecture/.

6. "Genetic Scissors: A Tool for Rewriting the Code of Life," Press Release, The Nobel Prize, October 7, 2020, https://www.nobelprize.org/prizes/chemistry/2020/press-release/.

7. National Academies of Sciences, Engineering, and Medicine, *Safeguarding the Bioeconomy* (Washington, DC: National Academies Press, 2020), 73, https://doi.org/10.17226/25525.

8. Planetary Technologies, "Bioeconomy Dashboard," last modified 2023, https://www.planetarytech.earth/bioeconomy-dashboard-1.

9. Michael Chui, Matthias Evers, James Manyika, et al., "The Bio Revolution: Innovations Transforming Economies, Societies, and Our Lives," McKinsey & Company, May 13, 2020, https://www.mckinsey.com/industries/life-sciences/our-insights/the-bio-revolution-innovations-transforming-economies-societies-and-our-lives.

10. Robert Carlson, "Estimating the Biotech Sector's Contribution to the US Economy," *Nature Biotechnology* 34, no. 3 (March 2016): 247–55, https://doi.org/10.1038/nbt.3491.

11. Chui et al., "The Bio Revolution."

12. National Research Council, *Industrialization of Biology: A Roadmap to Accelerate the Advanced Manufacturing of Chemicals* (Washington, DC: National Academies Press, 2015), https://doi.org/10.17226/19001.

13. "IBM Mainframe," Wikipedia, last modified July 20, 2024, https://en.wikipedia.org/wiki/IBM_mainframe.

14. Vinton G. Cerf and Robert E. Kahn, "A Protocol for Packet Network Intercommunication," *IEEE Transactions on Communications* 22, no. 5 (May 1974): 637–48, https://doi.org/10.1109/TCOM.1974.1092259.

15. *Whole Earth Epilog* (Baltimore, MD: Penguin Books, 1974), https://archive.org/details/wholeearthepilog00unse.

16. "History of Personal Computers," Wikipedia, last modified August 19, 2024, https://en.wikipedia.org/wiki/History_of_personal_computers.

17. Antheia, "Antheia Completes Successful Product Validation," January 8, 2024, https://antheia.bio/antheia-completes-successful-product-validation/.

18. Antheia, "U.S. Secretary of State Antony Blinken Visits Antheia to Discuss Biotechnology Innovation," May 30, 2024, https://antheia.bio/u-s-secretary-of-state-antony-j-blinken-visits-antheia-to-discuss-biotechnology-innovation/.

19. Chenwang Tang, Lin Wang, Lei Zang, et al., "On-Demand Biomanufacturing Through Synthetic Biology Approach," *Materials Today Bio* 18, no. 100518 (February 2023), https://doi.org/10.1016/j.mtbio.2022.100518.

20. Selin Kocalar, Bess M. Miller, Ally Huang, et al., "Validation of Cell-Free Protein Synthesis Aboard the International Space Station," *ACS Synthetic Biology* 13, no. 3 (March 2024): 942–50, https://doi.org/10.1021/acssynbio.3c00733.

21. Jessica C. Stark, Thapakorn Jaroentomeechai, Tyler D. Moeller, et al., "On-Demand Biomanufacturing of Protective Conjugate Vaccines," *Science Advances* 7, no. 6 (February 2021): eabe9444, https://doi.org/10.1126/sciadv.abe9444; and Jessica C. Stark, Ally Huang, Peter Q. Nguyen, et al., "BioBits™ Bright: A Fluorescent Synthetic Biology Education Kit," *Science Advances* 4, no. 8 (August 2018): eaat5107, https://doi.org/10.1126/sciadv.aat5107.

22. National Security Commission on Emerging Biotechnology, "DNA: Reading, Writing, and Editing," February 2024, https://www.biotech.senate.gov/press-releases/dna-reading-writing-and-editing/.

23. National Institutes of Health, "The Human Genome Project," National Human Genome Research Institute, last modified May 14, 2024, https://www.genome.gov/human-genome-project.

24. Kristen Philipkoski, "Celera Wins Genome Race," *Wired*, April 6, 2000, https://www.wired.com/2000/04/celera-wins-genome-race/.

25. James M. Heather and Benjamin Chain, "The Sequence of Sequencers: The History of Sequencing DNA," *Genomics* 107, no. 1 (January 2016): 1–8, https://doi.org/10.1016/j.ygeno.2015.11.003.

26. Davied Deamer, Mark Akeson, and Daniel Branton, "Three Decades of Nanopore Sequencing," *Nature Biotechnology* 34 (May 2016): 518–24, https://doi.org/10.1038/nbt.3423.

27. Oxford Nanopore Technologies, "Company History," accessed September 3, 2024, https://nanoporetech.com/about/history.

28. Marvin H. Caruthers, "The Chemical Synthesis of DNA/RNA: Our Gift to Science," *Journal of Biological Chemistry* 288,

no. 2 (December 2012): 1420–24, https://doi.org/10.1074/jbc.X112.442855.

29. Ryan Cross, "At GP-write, Scientists Take First Steps on Way to Synthetic Human Genome," *Chemical & Engineering News*, May 14, 2018, https://cen.acs.org/biological-chemistry/synthetic-biology/GP-write-scientists-take-first/96/i20.

30. *Business Wire*, "Twist Bioscience Expands Gene Offering with Long Gene Fragments up to 5.0kb," August 8, 2024, https://www.businesswire.com/news/home/20240808863612/en/Twist-Bioscience-Expands-Gene-Offering-With-Long-Gene-Fragments-up-to-5.0kb.

31. Planetary Technologies, "Bioeconomy Dashboard."

32. *Business Wire*, "Integrated DNA Technologies Invests in New U.S. Synthetic Biology Manufacturing Facility," May 28, 2024, https://www.businesswire.com/news/home/20240528727951/en/Integrated-DNA-Technologies-Invests-in-New-U.S.-Synthetic-Biology-Manufacturing-Facility.

33. MaryAnn Labant, "Enzymatic DNA Synthesis: Shorter Waits, Longer Strands," *Genetic Engineering & Biotechnology News*, July 1, 2024, https://www.genengnews.com/topics/omics/enzymatic-dna-synthesis-shorter-waits-longer-strands.

34. Lian Ge-Chen, Tianlong Lan, Shuo Zhang, et al., "A Designer Synthetic Chromosome Fragment Functions in Moss," *Nature Plants* 10 (January 2024): 228–39, https://doi.org/10.1038/s41477-023-01595-7.

35. Pranami Goswami, Kuang He, Jinhua Li, et al., "Magnetotactic Bacteria and Magnetofossils: Ecology, Evolution, and Environmental Implications," *NPJ Biofilms and Microbiomes* 8, no. 43 (June 2022), https://doi.org/10.1038/s41522-022-00304-0.

36. Sarah Graham, "Sea Sponge Inspires Better Fiber-Optic Cables," *Scientific American*, August 21, 2003, https://www.scientificamerican.com/article/sea-sponge-inspires-bette/.

37. Abigail Kukura, PJ Maykish, David Lin, et al., *National Action Plan for U.S. Leadership in Biotechnology*, Special Competitive Studies Project, April 12, 2023, 1, https://www.scsp.ai/wp-content/uploads/2023/04/National-Action-Plan-for-U.S.-Leadership-in-Biotechnology.pdf.

38. Semiconductor Research Corporation, "SemiSynBio Consortium and Roadmap Development," accessed September 3, 2024, https://www.src.org/program/grc/semisynbio/semisynbio-consortium-roadmap.

39. DNA Data Storage Alliance, "DNA Data Storage Alliance," accessed September 3, 2024, https://dnastoragealliance.org.

40. Larry Diamond, James O. Ellis, Jr., and Orville Schell, eds., *Silicon Triangle: The United States, Taiwan, China, and Global Semiconductor Security* (Stanford, CA: Hoover Institution Press, 2023), https://www.hoover.org/research/silicon-triangle-book.

41. Strategic Environmental Research and Development Program, "Synthetic Biological Techniques for Energetic Materials," Environmental Security Technology Certification Program, accessed September 3, 2024, https://serdp-estcp.mil/newsitems/details/ac878993-2005-4182-948a-f63c95668499/synthetic-biological-techniques-for-energetic-materials.

42. NATO Allied Command Transformation, "Resilience in NATO," December 15, 2023, https://www.act.nato.int/article/resilience-in-nato/.

43. Tore Brinck, ed., *Green Energetic Materials* (Chichester, West Sussex, United Kingdom: John Wiley & Sons, 2014), https://doi.org/10.1002/9781118676448; North Atlantic Treaty Organization, "Environment, Climate Change and Security," last modified July 18, 2024, https://www.nato.int/cps/en/natohq/topics_91048.htm.

44. Light Bio, "Light Bio," accessed September 3, 2024, https://www.light.bio.

45. Sasa Woodruff, "Gardeners Can Now Grow a Genetically Modified Purple Tomato Made with Snapdragon DNA," NPR, February 6, 2024, https://www.npr.org/sections/health-shots/2024/02/06/1228868005/purple-tomato-gmo-gardeners.

46. Norfolk Healthy Produce, "Empress Limited Edition Tomato," accessed September 3, 2024, https://www.norfolkhealthyproduce.com.

47. Hadley Leggett, "Researchers Use Skin-Colonizing Bacteria to Create a Topical Cancer Therapy in Mice," *Stanford Medicine*, April 12, 2023, https://med.stanford.edu/news/all-news/2023/04/cancer-bacteria.html.

48. Michael A. Fischbach, Kazuki Nagashima, Yiyin E. Chen, et al., "Bacteria-engineered to Elicit Antigen-Specific T Cells," US Patent 2024/0024380 A1, filed December 22, 2021, and issued January 25, 2024.

49. Iliano V. Coutinho-Abreu, Omid Jamshidi, Robyn Raban, et al., "Identification of Human Skin Microbiome Odorants That Manipulate Mosquito Landing Behavior," *Scientific Reports* 14, no. 1631 (January 2024), https://doi.org/10.1038/s41598-023-50182-5.

50. Purna C Kashyap, Michael Fischbach, and Brianna B. Williams, "Methods and Materials for Using Ruminococcus gnavus or Clostridium sporogenes to Treat Gastrointestinal Disorders," US Patent 11,878,002 B2, filed April 8, 2021, and issued January 23, 2024.

51. Herbert S. Lin, ed., "Biotechnology and Synthetic Biology," in *The Stanford Emerging Technology Review 2023* (Stanford, CA: Hoover Institution Press, 2023): 33–43, https://setr.stanford.edu/; Ewen Callaway, "'The Entire Protein Universe': AI Predicts Shape of Nearly Every Known Protein," *Nature*, July 29, 2023, https://www.nature.com/articles/d41586-022-02083-2.

52. The Nobel Prize in Chemistry 2024, "Summary," The Nobel Prize, October 12, 2024, https://www.nobelprize.org/prizes/chemistry/2024/summary/.

53. Brian L. Hie, Varun R. Shanker, Duo Xu, et al., "Efficient Evolution of Human Antibodies from General Protein Language Models," *Nature Biotechnology* 42 (April 2024): 275–83, https://doi.org/10.1038/s41587-023-01763-2.

54. Doni Bloomfield, Jaspreet Pannu, Alex W. Zhu, et al., "AI and Biosecurity: The Need for Governance," *Science* 385, no. 6711 (August 2024): 831–33, https://doi.org/10.1126/science.adq1977.

55. Eric Nguyen, Michael Poli, Matthew G. Durrant, et al., "Sequence Modeling and Design from Molecular to Genome Scale with Evo," bioRxiv, March 6, 2024, https://doi.org/10.1101/2024.02.27.582234.

56. John I. Glass, Chuck Merryman, Kim S. Wise, et al., "Minimal Cells—Real and Imagined," *Cold Spring Harbor Perspectives in Biology* 9, no. 12 (December 2017): a023861, https://doi.org/10.1101/cshperspect.a023861.

57. Shohei Kitano, Ciai Lin, Jee Loon Foo, et al., "Synthetic Biology: Learning the Way Toward High-Precision Biological Design," *PLOS Biology* 21, no. 4 (April 2023): e3002116, https://doi.org/10.1371/journal.pbio.3002116.

58. Akshay J. Maheshwari, Jonathan Calles, Sean K. Waterton, et al., "Engineering tRNA Abundances for Synthetic Cellular Systems," *Nature Communications* 14, no. 4594 (July 2023), https://doi.org/10.1038/s41467-023-40199-9. A useful description of the contrast between "design-build-test-learn" and "design-build-work" can be found at https://centuryofbio.com/p/design-build-work.

59. Build-A-Cell, "Build-A-Cell," accessed September 3, 2024, https://www.buildacell.org; BaSyC, "BaSyC," accessed September 3, 2024, https://www.basyc.nl.

60. B.Next, "B.Next," accessed September 3, 2024, https://bnext.bio.

61. Drew Endy, "Upwelling," *Original Syn*, Blog, October 28, 2020, https://blog.originalsyn.bio/2020/10/upwelling.html.

62. Sznews, "Six Countries Sign MOU on Synthetic Biology," General Office of Guangming District People's Government, last modified April 7, 2024, https://www.szgm.gov.cn/english/news/latestnews/content/post_11232084.html.

63. Defense Advanced Research Projects Agency, "The Sputnik Surprise," US Department of Defense, accessed September 3, 2024, https://www.darpa.mil/about-us/timeline/creation-of-darpa.

64. Dorian Leger, "Photovoltaic-driven Microbial Protein Production Can Use Land and Sunlight More Efficiently than Conventional Crops," *Proceedings of the National Academy of Sciences USA* 118, no. 26 (June 2021): e2015025118, https://doi.org/10.1073/pnas.2015025118 and Emiliano Bellini, "Solar-Powered Large Scale Microbial Food Production," *pv Magazine*, August 3, 2021, https://www.pv-magazine.com/2021/08/03/solar-powered-large-scale-microbial-food-production.

65. Grant M. Landwehr, Bastian Vogeli, Cong Tian, et al., "A Synthetic Cell-Free Pathway for Biocatalytic Upgrading of One-Carbon Substrates," bioRxiv, August 8, 2024, https://doi.org/10.1101/2024.08.08.607227.

66. Reuters, "Biotech Firm Amyris Files for Bankruptcy in US," August 10, 2023, https://www.reuters.com/business/biotech-firm-amyris-files-bankruptcy-us-2023-08-10; Amy Feldman and Angel Au-Yeung, "The Inside Story of How SoftBank-Backed Zymergen Imploded Four Months after Its $3 Billion IPO," *Forbes*, October 13, 2021, https://www.forbes.com/sites/amyfeldman/2021/10/13/the-inside-story-of-how-softbank-backed-zymergen-imploded-four-months-after-its-3-billion-ipo.

67. Robert F. Service, "Synthetic Biology, Once Hailed as a Moneymaker, Meets Tough Times," *Science*, August 22, 2024, https://www.science.org/content/article/synthetic-biology-once-hailed-moneymaker-meets-tough-times.

68. US National Science Foundation, "NSF Invests in BioFoundries to Drive Advances Across Science and Engineering," August 28, 2024, https://new.nsf.gov/news/nsf-invests-biofoundries-drive-advances-across-science.

69. For example, polio, horsepox, SARS-CoV-2, and the Spanish flu virus have been synthesized from scratch in laboratories. See, respectively, Jeronimo Cello, Aniko V. Paul, and Eckard Wimmer, "Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template," *Science* 297,1016–18 (2002), https://www.science.org/doi/10.1126/science.1072266; Ryan S. Noyce, Seth Lederman, and David H. Evans, "Construction of an Infectious Horsepox Virus Vaccine from Chemically Synthesized DNA Fragments," *PLOS ONE* 13,1 e0188453, January 19, 2018, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0188453; Tran Thi Nhu Thao, Fabien Labroussaa, Nadine Ebert, et al., "Rapid Reconstruction of SARS-CoV-2 Using a Synthetic Genomics Platform," *Nature* 582, 561–65 (2020), https://www.nature.com/articles/s41586-020-2294-9; Terrence M. Tumpey, Christopher F. Basler, Patricia V. Aguilar, et al., "Characterization of the Reconstructed 1918 Spanish Influenza Pandemic Virus," *Science* 310, 77–80 (2005), https://www.science.org/doi/10.1126/science.1119392.

70. Erik Parens, Josephine Johnston, and Jacob Moses, "Ethical Issues in Synthetic Biology: An Overview of the Debates," Woodrow Wilson International Center for Scholars, June 2009, https://www.wilsoncenter.org/sites/default/files/media/documents/publication/synbio3.pdf.

71. National Security Commission on Emerging Biotechnology, "Home," accessed September 3, 2024, https://www.biotech.senate.gov.

72. The Under Secretary of Defense, "Terms of Reference—Defense Science Board Study on Emerging Biotechnologies and National Security," Memorandum, US Department of Defense, https://dsb.cto.mil/wp-content/uploads/TORs/TOR_DSB_EmergingBiotechnologiesandNationalSecurity.pdf.

73. Organisation for Economic Co-Operation and Development, "Global Forum on Technology," accessed September 3, 2024, https://www.oecd.org/en/networks/global-forum-on-technology.html; OECD Science, Technology and Innovation, "What Is Synthetic Biology? OECD Global Forum on Technology," YouTube, December 13, 2023, video, 1:36, https://youtu.be/-0OJU_l-pLA.

74. World Economic Forum, "Global Future Council on the Future of Synthetic Biology," accessed September 3, 2024, https://www.weforum.org/communities/gfc-on-synthetic-biology/.

75. Jonathan Calles, Isaac Justice, Detravious Brinkley, et al., "Fail-Safe Genetic Codes Designed to Intrinsically Contain Engineered Organisms," *Nucleic Acids Research* 47, no. 19 (November 2019): 10439–51, https://doi.org/10.1093/nar/gkz745; Akos Nyerges, Svenja Vinke, Regan Flynn, et al., "A Swapped Genetic Code Prevents Viral Infections and Gene Transfer," *Nature* 615 (2023): 720–27, https://doi.org/10.1038/s41586-023-05824-z.

STANFORD EXPERT CONTRIBUTORS

Dr. Drew Endy
SETR Faculty Council and Associate Professor of Bioengineering

Dr. Michael Jewett
Professor of Bioengineering

Dr. Jenn Brophy
Assistant Professor of Bioengineering

Dr. Brian Hie
Assistant Professor of Chemical Engineering

Dr. Artem Trotsyuk
SETR Fellow and Postdoctoral Scholar in Biomedical Ethics

03

CRYPTOGRAPHY

KEY TAKEAWAYS

- Cryptography is essential for protecting information, but alone it cannot secure cyberspace against all threats.
- Cryptography is the enabling technology of blockchain, which is the enabling technology of cryptocurrencies.
- Central bank digital currencies (CBDCs) are a particular type of cryptography-based digital currency supported by states and one that could enhance financial inclusion. Although the United States lags some countries in experimenting with a CBDC, it may benefit from a cautious, well-timed approach by learning from other nations' efforts.

Overview

The word *cryptography* originates from Greek words that mean "secret writing." In ancient times, cryptography involved the use of ciphers and secret codes. Today it relies on sophisticated mathematics to protect data from being altered or accessed inappropriately.[1] We are typically unaware that many of our day-to-day interactions with computers and the internet involve cryptography, from securing our online shopping to protecting our cell phone calls.

Cryptography is often invisible, but it is essential for most internet activities such as messaging, e-commerce, banking, or even simple internet browsing. Yet cryptography alone will never be enough to ensure the confidentiality, integrity, or availability of information. Inherent vulnerabilities in the software code that underpins all our internet-connected devices and the strong incentives for bad actors—from criminals to nation states—to engage

Cryptography is often invisible, but it is essential for most internet activities such as messaging, e-commerce, banking, or even simple internet browsing.

in cyberattacks that exploit human and technical vulnerabilities help to explain why cybersecurity will be an ongoing challenge.

Cryptography Basics: Public Keys, Private Keys, and Hashes

Here's an example: Drew has a private message intended only for Taylor. To keep it confidential, she scrambles (encrypts) the message using an encryption algorithm and transmits the scrambled message to Taylor as ciphertext. When Taylor receives the ciphertext, he unscrambles (decrypts) it to reveal what it originally said. This piece of decrypted text is known as the plaintext. Along comes Ellen, a third-party eavesdropper who wants to see the plaintext, so she must use any means at her disposal to break the cryptographically provided protection.

An example of an encryption algorithm is the shift cipher. Each letter in the plaintext is replaced by a letter that is some fixed number N of positions later in the alphabet. For example, if $N = 2$, Drew substitutes an A in the plaintext with a C in ciphertext, B in plaintext with D in ciphertext, and so on. If $N = 3$, then Drew substitutes A in plaintext with D in ciphertext. To decrypt the ciphertext, Taylor must know that Drew is using the shift cipher and must also know the value of N so that he can invert it. For example, knowing that $N = 2$, he knows to write down A when he sees C in the ciphertext. (Note that modern encryption algorithms are more sophisticated and secure than what has been presented here; they are also harder to explain.)

In this scenario, both Drew and Taylor must share a secret piece of information, namely N. N is the cryptographic key, which in general is a string of digits needed both to encrypt and to decrypt the message. Drew and Taylor must also know that the algorithm is the shift cipher. If Ellen somehow learns both of those facts, she can decrypt the message as well. This type of encryption algorithm—of which the shift cipher is an example—is known as symmetric cryptography, or secret-key cryptography. It requires a secure key distribution, which is a method of distributing secret keys to all parties who should have them—but preventing those who shouldn't from obtaining them.

Symmetric key cryptography proved to be cumbersome because parties wishing to communicate securely must connect physically to share the cryptographic key before such a communication can take place. Imagine how awkward phone communications would be if you had to meet every telephone partner in person before talking to that party.

In the 1970s, Stanford professor Martin Hellman and Whitfield Diffie codeveloped a technique known as asymmetric cryptography, or public-key cryptography. Public-key cryptography relies on a public key for encrypting messages that is freely available to everyone, which means it can be widely distributed even over insecure channels. However, decrypting a message requires a private key that is held only by the authorized party (see figure 3.1).[2] Although it is theoretically possible to derive a private key from a public key, that process (if well designed) would take much too long for practical purposes (it would take longer than the age of the universe). It is this essential property that is placed at risk by quantum computing, as discussed below.

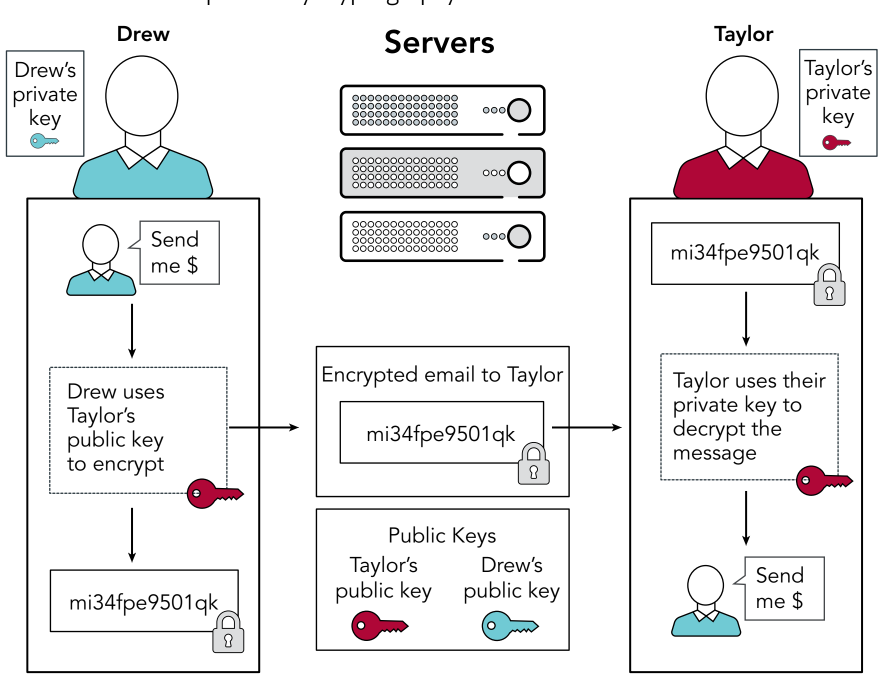
The mathematics of cryptography also underlie the creation of secure hashes. A hash is designed to accept a message of any length and compute a unique fixed-length string of numbers—called the hash value—corresponding to that message. Hashes have two key properties. First, it is extremely difficult to find another message that results in the same string of numbers. Second, if all you have is the string of numbers, it is infeasible to recover the original message.

Using a secure hash function, the sender can use public-key cryptography to provide assurances of integrity—information that cannot be tampered with or altered in any way—and identity, in that the originator of the message is who he or she claims to be.

To illustrate, Alice (the sender) first computes the hash value of her message. Next, she encrypts the hash value with her private key, a process analogous to signing a document, generating a digital signature of the message's hash.[3] Alice then sends the message and its digital signature to Bob (the receiver).

Upon receipt of the message, Bob can recover the hash value for the message that Alice purportedly sent and compare that value to his own computation of the hash value. If these match, Bob can be assured that the message has not been altered in transmission and also that Alice was the party who sent it, since only Alice could have used her private key to create a digital signature of the message's hash.

FIGURE 3.1 How public-key cryptography works



Messages can also be digitally time-stamped. A known authoritative time and date server—such as the Internet Time Service, operated by the National Institute of Standards and Technology—accepts a message, appends the current date and time, and then provides a digital signature for the stamped message.

Blockchain

Blockchain is a technology that enables multiple parties to coordinate when there is no central trusted party. This often comes up in financial settings. A blockchain records transactions so that they cannot be altered retroactively without detection. Because the entire blockchain can be distributed over thousands of computers, it is always accessible; anyone can deploy an application for it, and no one can prevent any such deployment. Moreover, anyone can interact with this application, and no one can prevent such an interaction. Finally, data cannot be erased. Later transactions may indicate that corrections are necessary, but the original data remain.

A blockchain can be visualized as a chain of blocks where each block contains a single transaction and a cryptographic hash of the previous block. This creates a chain in which every block except the first is linked to the previous block. As more transactions occur, the blockchain gets longer because more blocks are added to the chain.

The distributed nature of blockchain also increases security. A new transaction is broadcast to every party in the network, each of which has a replica of the entire blockchain (see figure 3.2). Each party tries to validate the new transaction. It could happen that these replicas may not be fully synchronized; some might have received the new transaction while others have not. To ensure that all replicas are identical, blockchains have mechanisms for coming to consensus on the correct information. Ethereum, for example, accepts transactions that have been validated by two-thirds of the participants. Blockchains are designed with economic incentives for replicas to behave honestly.

Applications that run on a blockchain are called smart contracts. These are computer programs that are always available and whose execution cannot be reversed—once a smart contract processes an incoming request, that processing cannot be rolled back. Smart contracts can be used to implement financial instruments, to record ownership of digital assets, and to create marketplaces where people can buy and sell assets. Smart contracts are composable—one smart contract can use another—thus creating a vibrant ecosystem of innovation where one project can make use of a service developed by another project. Once deployed, they are available forever, running whenever someone interacts with them. By contrast, cloud computing applications are inherently transient—as soon as the application developer stops paying the cloud fees, the cloud provider kills the application.

Key Developments

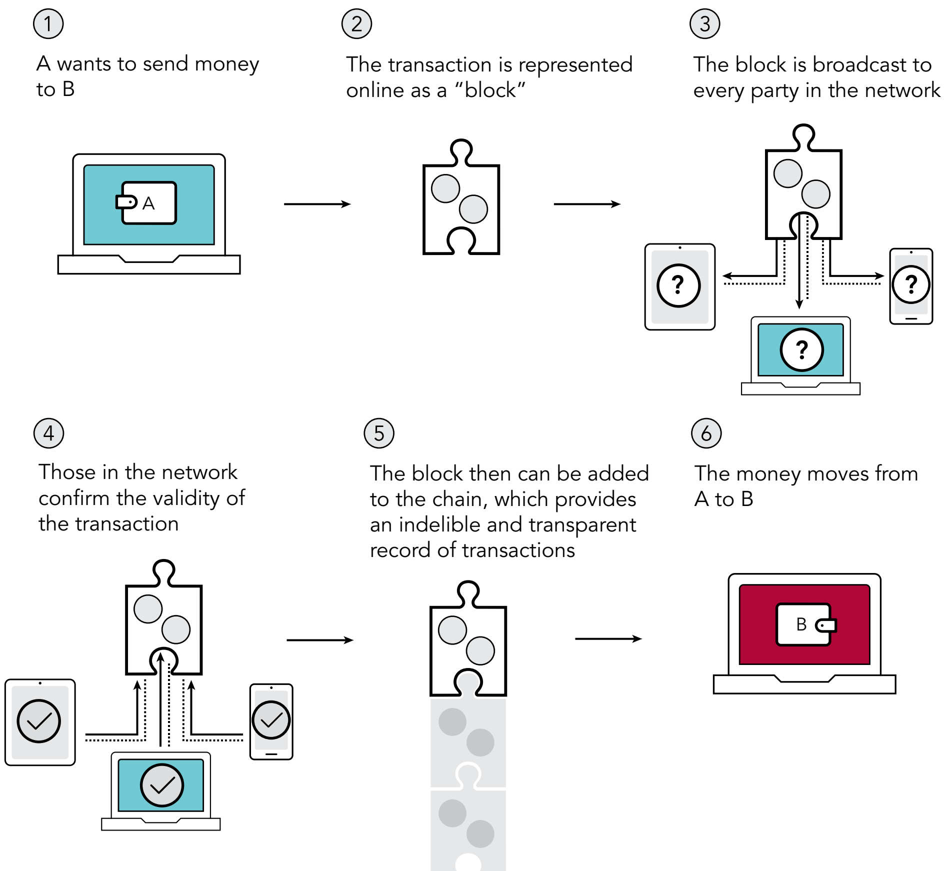
A Host of Blockchain Applications

Blockchain technology was developed decades ago but has recently been used for a variety of applications. All those listed below have been implemented in some form and are operational today, though perhaps not on particularly large scales.

Time-stamping and data provenance Because data written to a blockchain cannot be modified or removed, blockchains provide a good mechanism for data provenance and time-stamping. An artist or an author who creates a new work of art can post a hash of the work to the chain, thereby proving the time at which the object was created. If later someone else claims authorship of the creation, the artist can point to the chain to prove its provenance.

Identity management A blockchain stores all the data from a person's important documents—diplomas, healthcare and financial records, tax returns, birth certificate—in encrypted form. These

FIGURE 3.2 How a blockchain manages transactions



original records are saved digitally, signed by their original providers, and, when made available through the blockchain, provided with provenance and time-stamping. Blockchain also facilitates selective revelation: Upon request, the person can authorize release of data only to the minimal extent necessary to satisfy the request. For example, people can prove that their age is above some legal minimum, like twenty-one, but not have to reveal their date of birth. A woman can allow a healthcare researcher to look at her records for specific data—for example, whether she has ever had an abortion—without revealing her name. Applications of blockchain for identity management, such as SpruceID, are already being deployed.[4]

Supply chain management Blockchain can provide a transparent and secure way to track the movement of goods and their origin and quantity. This can be particularly valuable for high-value industries, such as the diamond industry; industries with significant counterfeit issues, such as luxury goods (see figure 3.3); or industries where the true source of goods is important, such as organic or vegan food. Blockchain can greatly simplify the job of forensic accountants trying to trace transactions.

Transactional records Many kinds of transactional records can be stored on a blockchain, thereby streamlining the process of buying and selling items by reducing fraud, increasing transparency, cutting

FIGURE 3.3 Blockchain helps tackle counterfeiting in the luxury goods industry

Source: Shutterstock / TY Lim

paperwork, and generally making the process more efficient.

Cryptocurrencies Cryptocurrencies are digital instruments that many people use as a medium of exchange. Well-known ones include Bitcoin, Ethereum, Avalanche, and Polygon, each of which has its own unique features and applications. Because they are not issued by any central authority, they are not subject to the same national regulatory regimes that govern traditional currencies (i.e., so-called fiat currencies). Cryptocurrencies use a blockchain structure to ensure the integrity and immutability of transaction data, making it resistant to fraud and counterfeiting and reducing its susceptibility to government interference or manipulation. Contrary to a common belief, cryptocurrencies can, but do not have to, support private or secret transactions—indeed, the most popular cryptocurrencies deliberately do not hide the details of their transactions. Those who transact in cryptocurrencies often wish to exchange their instruments for fiat currency (e.g., real dollars) and generally use a cryptocurrency exchange to do so. In the United States, such exchanges are regulated financial institutions and are presently under the jurisdiction of the Securities and Exchange Commission.

Secure Computation

The field of cryptography has also expanded in scope to include secure computation, a well-established subfield that enables multiple parties to contribute inputs to a function that they jointly compute in such a way that the specific inputs from each party are kept secret from the others. Secure computation enables data privacy during computation, ensuring that no party learns more information about the other parties' inputs than what can be inferred from the result alone. Secure computation also allows users to prove they possess knowledge of a statement without having to disclose the actual content of that statement.

To illustrate secure computation, consider the problem of determining the collective wealth of three people while keeping the individual wealth of each person secret. Alice chooses a large random number and in secret adds her wealth to that number. Alice then gives the sum to Bob privately, who adds his wealth secretly to the number received from Alice. Bob secretly passes the total to Charlie, who does the same computation and then passes the result to Alice. Alice then in secret subtracts her original random number from the number received from Charlie and reveals the result to everyone else. That revealed number is the sum of each party's wealth but at no time does anyone learn of anyone else's wealth.[5]

This example is oversimplified but is offered to suggest how computation on secret data might be accomplished. The example is not exactly how a real-world secure computation works (in fact, there is a subtle flaw in the procedure described); true secure computation protocols use more complex mathematics to defend against malicious behavior and to guarantee the privacy of each person's input during the computation process.

In addition, the example is somewhat artificial compared to more realistic examples (with more complex mathematics) such as tallying vote counts or bidding in an auction. For example, at an auction, three bidders each have a secret bid in mind, and the goal could be to determine which bid is the highest without publicly revealing information about the other bids.

Applications of secure computation allow data analytics to be performed on aggregated data without disclosing the data associated with any individual element of the dataset. Banks can detect fraud without violating the privacy of individual customers. A group of workers can calculate their average salary without revealing their colleagues' personal pay. A Stanford system called Prio allows for a network of connected computers to work together to compute statistics, with clients holding their individual data privately.[6] This was deployed, for example, on mobile phones during the COVID-19 pandemic to calculate how many people were exposed to COVID-19 in aggregate, without learning who was exposed.

Zero-Knowledge Proofs

A zero-knowledge proof is a cryptographic method that allows Paul (the prover) to prove to Vivian (the verifier) that Paul knows a specific piece of information without revealing to Vivian any details about that information. The term *zero knowledge* indicates that Vivian gains zero new knowledge about the information in question, apart from the fact that what Paul is saying is true.

Consider a simplified example that demonstrates the logic: two people dealing with a locked safe. Let's say Paul wants to prove to Vivian that he knows the combination to the safe, but he doesn't want to reveal the combination to Vivian. With a zero-knowledge proof, Paul can convince Vivian that he knows the combination without exposing the combination itself.

To do so, Paul has Vivian write something on a piece of paper without showing it to him. Together, they put the paper into the safe and spin the combination lock. Vivian now challenges Paul to say what is on the paper. Paul responds by asking Vivian to turn around (so that Vivian cannot see Paul) and then enters the combination of the safe, opens it, looks at the paper and returns it to the safe, and closes it. When Vivian turns around, Paul tells her what was on the paper. Paul has thus shown Vivian that he knows the combination without revealing to Vivian anything about the combination.

In practice, of course, zero-knowledge proofs are more complex, yet they already have seen real-world implementations:

Banking A buyer may wish to prove to a seller the possession of sufficient funds for a transaction without revealing the exact amount of those funds. This capability has been implemented in the Zcash cryptocurrency.[7]

Provenance for digital images Cameras can provide a digital signature for every photo, capturing an image and information about the time, date, and location. But such photos can then be digitally cropped, resized, or converted from color to black-and-white. Zero-knowledge proofs have been implemented in the standards of the Coalition for Content Provenance and Authenticity to ensure that the original photo was properly signed and that only permissible edits were made to the original without having to trust the editing software that was used.[8]

Cooperative tracking and verification of numbers of tactical nuclear warheads A zero-knowledge proof methodology has been developed to cooperatively provide updates on the movement and status changes of warheads in accordance with a political agreement to do so without revealing other sensitive information. This approach has not yet been implemented in any real arms control agreement, but its feasibility has been demonstrated in principle.[9]

Over the Horizon

Impact of Cryptography

The applications described above suggest a broad range of possibilities for cryptographically enabled data management services. Whether we will see their widespread deployment depends on complicated decisions about economic feasibility, costs, regulations, and ease of use.

Misaligned incentives can affect how fast innovations are deployed. Some of the applications described above provide significant benefits for the parties whose data can be better protected and kept more private. But existing companies, having built their business models on legacy systems that ingest all their customers' data, have no incentive to change their practices. They are the ones who would have to pay for these privacy-protecting capabilities, yet they would not benefit from their adoption.

A second point is that widespread deployment will require confidence that proposed innovations will work as advertised. That is, would-be users of these innovations must have confidence in them. But concepts such as secure computation and zero-knowledge proofs are math heavy and counterintuitive to most people. Expecting policymakers, consumers, and regulators to place their trust in these applications will be challenging.

Challenges of Innovation and Implementation

Although cryptography is fundamentally a mathematical discipline, it requires both human talent and substantial computing resources to examine the efficiency of new techniques, write software that is computationally expensive such as zero-knowledge provers, and conduct comprehensive scans of the internet. Progress also relies on interdisciplinary centers that bring together faculty from different fields to share problem sets and understand the potential benefits that cryptographically enabled techniques and approaches could provide.

Research is funded by both the US government and private industry, but funding from the US government is subject to many requirements that increase the difficulty of proposal submission manyfold (as much as by a factor of sixty). Thus, research faculty often prefer arrangements with the private sector, which tend to be much simpler. On the other hand, only the US government is able to fund research that may not pay off for many years (as in the case of quantum computing).

Policy, Legal, and Regulatory Issues

As a rule, public policy considerations are application specific; there has been no push to regulate basic research in cryptography for several decades.

EXCEPTIONAL ACCESS

Exceptional access regulations would require communications carriers and technology vendors to provide US law enforcement agencies access to encrypted information (both data storage and communications) under specific legal conditions. Opponents of exceptional access argue that implementing this capability inevitably weakens the security afforded by encryption to everyone. Supporters of exceptional access do not debate this technical assessment: It is true that exceptional access, by definition, weakens encryption. However, they argue that even if lower security is the result of implementing exceptional access, that price is worth the benefits to law enforcement.[10]

CRYPTOCURRENCY REGULATORY CONCERNS

Particularly considering the 2023 FTX trading scandal, in which the FTX cryptocurrency exchange went bankrupt and founder Sam Bankman-Fried was subsequently convicted of fraud, many have questioned the extent to which cryptocurrencies should

> **The lack of a regulatory framework for cryptocurrency affects many American users, consumers, and investors who are often confused about the basic workings of cryptocurrencies and their markets.**

be exchangeable for national currency and whether they are better regulated as investment instruments or as currency. The lack of a regulatory framework for cryptocurrency affects many American users, consumers, and investors who are often confused about the basic workings of cryptocurrencies and their markets. It may also prevent entrepreneurs from implementing their ideas in the United States or inadvertently incentivize them to move offshore.

ENERGY CONSUMPTION

Bitcoin, an older and today the dominant cryptocurrency, consumes an enormous amount of energy; Bitcoin mining uses more energy than the Netherlands.[11] For this reason, newer blockchains—notably Ethereum—are designed to use far less energy, and today Ethereum's annual energy use is less than a 10,000th of YouTube's annual consumption. But Ethereum's market capitalization is less than half that of Bitcoin, and it remains to be seen whether any less energy-intensive cryptocurrency will displace the latter.

QUANTUM COMPUTING AND CRYPTOGRAPHY

Current public-key cryptography is based on the extraordinarily long times (times comparable to the age of the universe) required with today's computers to derive a private key from its public-key counterpart. When realized, quantum computing (discussed more fully in chapter 8 on semiconductors) will pose a significant threat to today's public-key algorithms. Experts disagree on how long it will take to build quantum computers that are capable of this, but under the May 2022 National Security Memorandum 10, Promoting US Leadership in Quantum Computing While Mitigating Risks to Vulnerable Cryptographic Systems, the US government has initiated the transition to quantum-resistant public-key algorithms. Many experts in the field expect quantum-resistant algorithms will be widely available by the time quantum computing comes online.

At the intersection of quantum computing and cryptography are two important issues. The first is that support for the transition to a quantum-resistant encryption environment should continue with urgency and focus.

A second issue is that messages protected by pre-quantum cryptography will be vulnerable in a post-quantum world. If those messages had been saved by adversaries (likely in the case of parties like Russia), those bad actors will be able to read a host of old messages. Containing secrets from the past, they may reveal embarrassments and dangers with potentially detrimental policy implications.[12]

CENTRAL BANK DIGITAL CURRENCIES AND THE EROSION OF US FINANCIAL INFLUENCE

A central bank digital currency (CBDC) is a type of cryptography-based digital currency issued and regulated by a country's central bank, with legal tender status and value equivalent to the country's traditional currency—that is, digital assets backed by central banks. A CBDC can be designed with any number of the functional characteristics of cryptocurrencies and thus can be regarded as a national cryptocurrency. However, a CBDC could be implemented in a centralized manner to improve performance and efficiency instead of using distributed blockchain technology.

An important benefit of a CBDC is the marriage of convenience and lower costs of digital transactions—by cutting out intermediaries—and the regulatory oversight of traditional banking. In 2021, nearly six million Americans had no access to a bank account. Lower transaction costs would improve financial inclusion and enable many more people to have access to a well-regulated financial system. Those lower costs would also apply to cross-border transactions, therefore reducing the costs of international commerce.

The United States is considering issuing its own CBDC.[13] Although the dollar is the currency most used in cross-border transactions, the development of CBDCs by others could reduce global dependence on the US currency and on a financial infrastructure largely controlled today by the United States (e.g., the Society for Worldwide Interbank Financial Telecommunication, or SWIFT, which is used by banks and other institutions to send secure messages to each other about financial transactions). This could significantly undermine the effectiveness of US economic sanctions and other financial tools. Today, more than ninety nations are researching, piloting, or deploying CBDCs, with several already testing cross-border transactions. China is the first major country to deploy a CBDC, the digital yuan, widely within its own economy.[14] America may lag China and some other countries, but it could benefit from a cautious, well-timed approach by learning from earlier adopters' experiences.

NOTES

1. National Institute of Standards and Technology, "Cryptography," US Department of Commerce, accessed August 15, 2023, https://www.nist.gov/cryptography.

2. Whitfield Diffie and Martin Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory* IT-22, no. 6 (November 1976): 644–54.

3. In this context, encrypting the hash value simply means running the encryption algorithm using as the input key a string of numbers that just happen to be Alice's private key. In most cases involving public-key cryptography, the private key is used only for decryption purposes, but nothing stops a user from using it in other ways.

4. SpruceID, "SpruceID," Spruce Systems, accessed October 13, 2024, https://spruceid.com/.

5. This example is inspired by Keyless Technologies, "A Beginner's Guide to Secure Multiparty Computation," *Medium*, February 22, 2020, https://medium.com/@keylesstech/a-beginners-guide-to-secure-multiparty-computation-dc3fb9365458.

6. See Stanford University, "Prio," accessed September 25, 2023, https://crypto.stanford.edu/prio.

7. Zcash, "What Are Zero-Knowledge Proofs?," accessed August 30, 2023, https://z.cash/learn/what-are-zero-knowledge-proofs.

8. Trisha Datta and Dan Boneh, "Using ZK Proofs to Fight Disinformation," Medium, September 29, 2009, https://medium.com/@boneh/using-zk-proofs-to-fight-disinformation-17e7d57fe52f.

9. Miles A. Pomper, William Alberque, Marshall L. Brown Jr., et al., *OP55: Everything Counts: Building a Control Regime for Nonstrategic Nuclear Warheads in Europe*, CNS Occasional Paper Series, James Martin Center for Nonproliferation Studies, May 10, 2022, https://nonproliferation.org/op55-everything-counts-building-a-control-regime-for-nonstrategic-nuclear-warheads-in-europe.

10. Office of Public Affairs, "Attorney General William P. Barr Delivers Keynote Address at the International Conference on Cybersecurity," US Department of Justice, July 23, 2019, https://www.justice.gov/opa/speech/attorney-general-william-p-barr-delivers-keynote-address-international-conference-cyber.

11. Digiconomist, "Bitcoin Energy Consumption Index," accessed September 16, 2023, https://digiconomist.net/bitcoin-energy-consumption.

12. Herbert Lin, "A Retrospective Post-Quantum Policy Problem," *Lawfare*, September 14, 2022, https://www.lawfaremedia.org/article/retrospective-post-quantum-policy-problem.

13. Board of Governors of the Federal Reserve System, "Central Bank Digital Currency (CBDC): Frequently Asked Questions," accessed August 15, 2023, https://www.federalreserve.gov/cbdc-faqs.htm.

14. Darrell Duffie and Elizabeth Economy, eds., *Digital Currencies: The US, China, and the World at a Crossroads* (Stanford, CA: Hoover Institution, 2022), https://www.hoover.org/sites/default/files/research/docs/duffie-economy_digitalcurrencies_web_revised.pdf.

STANFORD EXPERT CONTRIBUTORS

Dr. Dan Boneh
SETR Faculty Council and Professor of Computer Science and of Electrical Engineering

Dr. David Tse
Thomas Kailath and Guanghan Xu Professor of Engineering

Neil Perry
SETR Fellow and PhD Student in Computer Science

04

LASERS

KEY TAKEAWAYS

- Laser technology has become essential for a wide range of applications, including communications, high-end chip production, defense, manufacturing, and medicine.
- Because advances in laser technology tend to occur in the context of specific applications, laser technology research and development is widely dispersed among different types of laboratories and facilities.
- Broad investment in next-generation lasers holds the potential to improve progress in nuclear fusion energy technology, weapons development, and quantum communication.

Overview

Improvements in laser technology since its invention in 1960 have allowed light to be manipulated and used in previously unimaginable ways. Lasers now underpin a huge range of scientific and industrial applications. Already, lasers with ever higher energy and power are being developed across a wider range of wavelengths and with pulse lengths that can illuminate many details of what is happening very rapidly at an atomic and molecular level.

A laser—an acronym derived from “light amplification by stimulated emission of radiation”—is a light source with three important characteristics. First, its light is monochromatic (i.e., single color), meaning the light is highly concentrated around a central wavelength, with very little emitted at other wavelengths. Monochromatic light enhances data transmissions by minimizing chromatic aberration, which occurs when a lens can’t focus different colors of light on a single point. Monochromatic lasers are also essential in scientific and medical applications that need specific wavelengths for controlled interactions with materials or tissues.

Second, a laser is directional, which means its energy can be concentrated into a small spot, significantly increasing intensity and making lasers useful for applications that require precision and high energy density, such as cutting, welding, and surgical procedures.

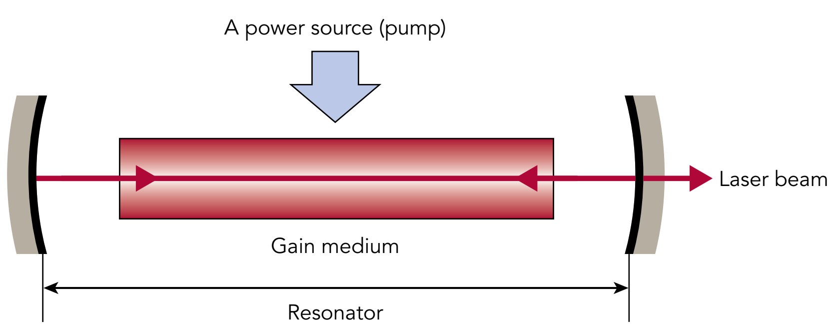
Third, laser light is coherent, which means that the light waves it uses are in phase with each other—that is, they repeatedly reach the same peak or trough at the same point in time and space. This property is important for holography, interferometry (the measurement of light sources), and optical sensing, where precise phase information is needed to create accurate and detailed images or measurements.

There are many ways to produce laser light. Lasers typically involve a power source (a pump), a gain medium (a material within which the energy supplied by the pump is turned into laser light), and a resonator that encloses the gain medium within which laser light is produced (see figure 4.1). Progress in laser technology depends on advancing one or more of these elements and is generally measured with respect to five technical characteristics (figures of merit) of the beam:

Peak power Generating the brightest possible laser pulses—which equates to the greatest possible power—for very short times. The 2018 Nobel Prize in Physics was awarded for the development of the chirped pulse amplification technique, a high-power, short-duration approach for producing laser pulses that significantly outperformed prior peak power achievements. Peak powers in state-of-the-art lasers can now reach levels that damage the laser itself. Because of this, to reach even higher power levels, scientists have used multiple beams from multiple lasers focused on a target. In 2024, one laser delivered a peak power of 10 petawatts (or 10^{16} watts) with a pulse time (duration) of 24 femtoseconds (a femtosecond is 10^{-15} seconds).[1] (For comparison, total global electrical generation capacity today is about 9 terawatts, or about 1/1,000th of the peak power of a 10-petawatt laser.)

Energy Delivering as much energy as possible in a beam. A high-energy laser beam generally delivers its energy on timescales of a few nanoseconds, or around a million times longer than the lasers

FIGURE 4.1 Typical components of a laser



discussed above. The highest energy lasers today are found at the National Ignition Facility of the Lawrence Livermore National Laboratory (LLNL). These deliver beam energies as high as 2.2 megajoules and have been used to drive controlled nuclear fusion reactions at the lab that produced a net energy gain.[2]

Average power Reliably delivering high power and energy at elevated repetition rates. Many laser applications need pulses whose quality is consistent and that are delivered frequently and reliably. An important technical challenge is managing heat buildup in the resonator, which can limit the number of pulses a laser can produce over a given time. Today, high-average-power lasers—ones rated in excess of 300 kilowatts—use active liquid or gas cooling in what is called a distributed gain laser architecture.[3] Some efforts to manage possible laser damage involve development of improved material production techniques that can, for instance, reduce erosion of the coating in lasers' optical systems. Other solutions involve the use of lasers based on gaseous media, which are inherently less prone to damage.

Pulse length Generating shorter laser pulses. Over the course of lasers' development, squeezing energy into ever shorter pulse lengths has been the primary way of generating beams with higher power. In addition, short pulse lengths can be used like a strobe light to observe rapid motion. For example, the generation of attosecond pulses (10^{-18} seconds) was recognized with the Nobel Prize in Physics in 2023.[4] These pulses are shorter than the timescale of electronic motion within atoms, enabling atomic processes to be observed with the electrons effectively frozen in place.

Wavelength Delivering laser-like pulses at more frequencies. Historically, the term *laser* is generally used to refer to devices operated near to optical wavelengths, distinguishing them from the microwave "masers" that preceded them. Today, however, pulses of radiation with the same properties that make lasers so useful can be generated and used across a far wider range of the electromagnetic spectrum. Being able to use a wider range of wavelengths enables the investigation and manipulation of matter under a wider variety of conditions.

The engineering characteristics of lasers are also an important aspect of how fast the technology advances. For example, different configurations of power sources, resonators, and gain mediums can result in lasers of different sizes, weight, reliability, cost, and other key features. Moreover, some applications require mechanisms that can steer beams in particular directions. Addressing these and other engineering issues helps take lasers from labs to the commercial world, where many non-research applications make important use of them.

For example, researchers have miniaturized a titanium-sapphire laser by polishing and etching a bulk titanium-sapphire crystal to a nanoscale-thick layer on a silicon dioxide support.[5] They then patterned a circular waveguide into the titanium-sapphire layer. The intensity of the generated light is increased over the length of the waveguide. The miniaturized laser is several orders of magnitude smaller and significantly less expensive than existing titanium-sapphire lasers, which are currently the best ones for a variety of applications including quantum optics, spectroscopy, and biomedical research.

Key Developments

The basic operating principles and physics of lasers are generally well understood. What stands out in reviewing key developments in laser technology is the wide variety of applications to which the technology is relevant. Below is a list of some examples of important applications.

Medicine

Lasers in medicine have historically been used to ablate, cut out, or vaporize tissue or to clot bodily fluids.[6] For example, a robot-guided laser has been used to perform bone surgery.[7] Traditional tools like saws, drills, and burs can cause mechanical and thermal damage to bone and tissue and are also limited to simple cuts. In contrast, lasers offer more precise, cleaner cuts with less damage to surrounding tissue, and they can handle complex trajectories, especially when guided by a computer-controlled robot arm for fully automated surgery. This technology not only enhances accuracy but also reduces recovery time for patients.

A well-known example is laser eye surgery, where ultrashort laser pulses are used to remove small amounts of corneal tissue with great precision, thus reshaping the cornea to improve how light is focused onto the retina (see figure 4.2). Interestingly, this technique, popularly known as LASIK, was inspired by a laser eye injury in a research lab.[8]

Lasers can also be used to destroy subsurface tumors with minimal thermal damage to surrounding healthy tissue. Researchers have demonstrated the use of a focused laser beam from an ultrashort-pulse diode laser source.[9] The beam is intense enough to destroy a tumor but focused enough and short enough that it causes only minimal damage to the surrounding tissue.

FIGURE 4.2 Laser eye surgery uses laser pulses to remove corneal tissue

Source: Shutterstock / Terelyuk

Lasers may come to play an important role in certain cancer treatments. Specifically, some recent cancer research has discovered that charged particle beams delivered at extremely high dose rates to cancerous tissue may have unique benefits. A very high dose of proton beam radiation delivered to a cancerous tumor over a very short time will kill it while significantly reducing collateral damage to surrounding tissue compared to current approaches. The production of such proton beams was driven by a laser whose operating characteristics could be very tightly controlled, leading to a beam precisely tailored for the tumor in question.[10]

Military Applications

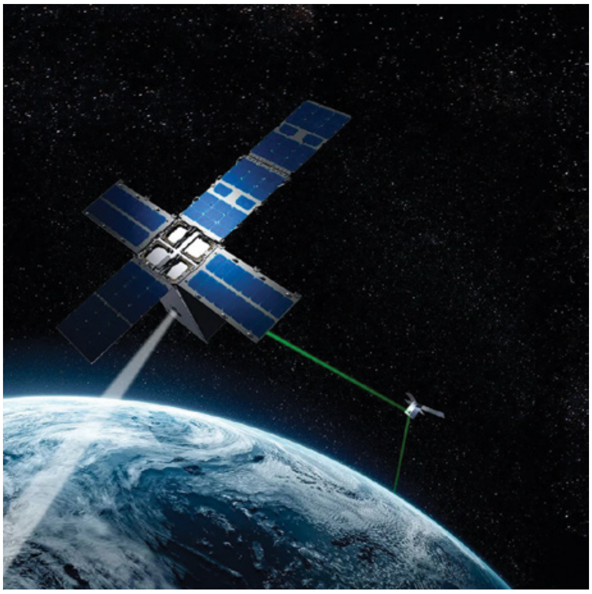
Lasers as weapons could serve a variety of ground-based missions,[11] including attacking satellites and providing short-range air defense to counter drones, rockets, artillery, and mortar rounds. In these roles, lasers have several advantages over conventional munitions—in particular, lower cost per shot and potentially more rounds in their magazines (assuming their power supplies are not exhausted). But they have certain disadvantages as well—most importantly, rain, fog, and some other atmospheric conditions potentially limit their range and beam quality.

Because a laser beam traveling through the atmosphere loses energy as the range to the target increases, laser weapons need high-power beams to damage distant targets. These two conflicting requirements can be resolved with a laser that delivers a beam with a very long pulse length. This means the beam must dwell on its target for the entire duration of the pulse—and if the target moves during that time, the weapon must have a pointing mechanism that keeps the beam on target for a few seconds. Another way to resolve these requirements is to select a wavelength for laser operation that is not strongly absorbed by the atmosphere. But since some degree of absorption will occur in any event, the tension between these requirements can only be reduced and not eliminated.

In general, laser weapons require a laser to supply the necessary beam, a power supply (typically an electrical battery source or chemicals that are mixed to produce energy), and a way of tracking a target and directing the beam to remain trained on it while it is in motion so enough energy can be delivered to destroy or disable the target. (When the mission is to disable things like sensors that a target may be carrying, such as cameras on a reconnaissance satellite, the power required is much lower than if the target's entire physical structure must be destroyed.)

Progress in laser weapons involves making them smaller and lighter, more rugged for an operational military environment, more powerful, and more efficient in their conversion of energy in the magazine to shots fired. Auxiliary technologies such as those for beam tracking and target sensing must also work in concert with the lasers themselves.

FIGURE 4.3 Lasers are well suited for space-to-space communications



Source: General Atomics Electromagnetic Systems

Communications

Lasers play a key role in communications by transmitting data through fiber-optic cables, which make up the bulk of the infrastructure behind the internet. As demand for information transfer grows, approaches that raise data transfer rates are increasingly important. Recent results have shown that data can be sent through fiber-optic cables using much shorter laser pulses without a loss of transfer fidelity, potentially lowering power requirements significantly.[12]

Lasers can transmit data over long ranges and are even being used to enable satellites in orbit to communicate with one another.[13] Compared to traditional radio transmission systems, laser communications allow for data-transfer rates that are 10 to 100 times faster than radio. They are also more secure than radio systems because they have directed narrower beam widths that make them harder to intercept. Laser communications systems are also more energy efficient, and the hardware required is smaller and lighter. The primary technical challenge they face is the issue of beam alignment: Because the beams need to be very narrow, aligning the sending laser and the laser receiver properly is hard if they are far apart.

Lasers are well suited for space-to-space communications, where there is very little to interfere with the beam (see figure 4.3). Starlink—the space-based internet service provider wholly owned and operated by SpaceX—uses laser communications to transfer data at high speeds directly between satellites in low Earth orbit (LEO) without going through ground stations.[14] In December 2023, NASA successfully demonstrated its first two-way laser communication link between the International Space Station in LEO and a geostationary satellite.[15] Efforts are also underway to adapt laser communications for ground/air/sea-to-space applications, which will mean overcoming challenges posed by atmospheric interference with data-carrying laser beams.

Additive Manufacturing

Lasers are useful for additive manufacturing (also known as 3-D printing), enabling precise and efficient creation of complex structures through various techniques. For example, in stereolithography an ultraviolet laser is used to cure a photosensitive resin layer by layer. The laser selectively turns on and off, curing the layer with the appropriate structure. The next layer is treated similarly until the artifact is fully formed. Another method, selective laser sintering, uses a laser to harden (sinter) a layer of powder, such as nylon or metal. These laser-based techniques can be adapted for various materials, making them suitable for rapid prototyping and other manufacturing applications.

Particle Traps / Quantum Computing

Lasers can be used to create the coldest temperatures achieved on Earth—significantly colder than the void of interstellar space. The record low temperature is around a few millionths of a degree kelvin from absolute zero (about minus 273 degrees Celsius) for small material samples, with parallel work focusing on cooling larger samples, such as the mirrors at the Laser Interferometer Gravitational-Wave Observatory, in order to reduce thermal noise in the system that interferes with the detection of gravitational waves from space.

Laser-cooled atoms demonstrate measurable quantum behavior and hence are one of the approaches being pursued to work with quantum bits, or qubits, in labs.[16] (Qubits are the building blocks for quantum computers, which are discussed in chapter 8 on semiconductors.) By focusing laser beams into a very small space, scientists can trap atoms and other particles and manipulate them into quantum states to produce qubits using yet more lasers.

Orbital Debris Removal

Chapter 9 on space describes the Kessler syndrome, a scenario in which the density of objects both large and small in LEO becomes so high that collisions between some of them create a cascade of debris, potentially rendering space activities and satellite operations in certain orbital ranges difficult or impossible for many future generations. Technologies for debris removal may become important in the future, and lasers could be used for this purpose.

Specifically, NASA is supporting a project to research a network of lasers mounted on space platforms.[17] These lasers are supposed to deflect debris of various sizes through ablation, which involves an intense laser pulse vaporizing surface material on an object. The material is ejected away from it, altering the object's momentum. If the impulse of that ejection is properly oriented, the object's speed can be reduced, and eventually it will deorbit and burn up in the atmosphere on reentry.

Imaging

At short wavelengths, pulses from an X-ray free-electron laser (XFEL) can penetrate through materials to image structures and measure a material's physical properties. The current Linac Coherent Light Source (LCLS)-II High-Energy upgrade to the XFEL at the SLAC National Accelerator Laboratory will push the maximum energy that it can reach even higher, allowing heavier and denser materials to be probed.[18]

XFELs are particularly useful for imaging where the shorter wavelengths of X-rays allow better spatial resolution compared to visible light—an example is the coherent X-ray imaging end station of SLAC's LCLS.[19] In addition, XFELs can emit very short pulses, which helps them excel at tracking changes over very short time periods. Previous results have allowed new proteins to be imaged and have enabled researchers to observe phase transitions of quantum materials in real time[20] or observe materials under extreme conditions of pressure, such as those in the center of the sun. The approach has also shown how biomolecules move in real time, and an extended research effort has followed the complex series of reactions that occur throughout the process of photosynthesis, with implications for future photovoltaic cells and other devices that seek to harness solar power.[21]

Materials Processing

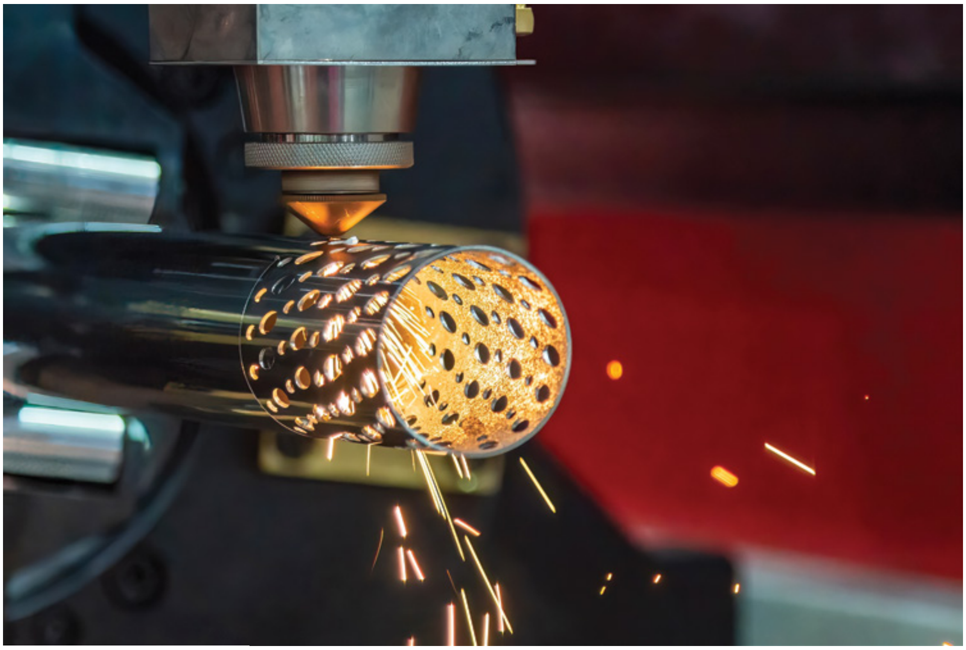
Lasers are now used for a wide range of applications in materials processing, including laser cutting of precise shapes (see figure 4.4), laser drilling of micron-scale holes, and laser peening—deliberately deforming surfaces—to add stress to materials. Ultrashort pulse lasers enable material to be ablated precisely with minimal damage to surrounding areas—a process useful both in manufacturing and in surgery. This process, sometimes called cold ablation, works by vaporizing material faster than heat can spread through it. However, to prevent overheating, each spot must be processed slowly, which limits overall throughput. To address this challenge, a beam from a powerful laser is split into smaller beams, which can work on multiple areas simultaneously.[22]

The limited collateral damage it causes means laser processing can be used even on biological samples. The benefits of short pulses also extend to surgery, as described above, and to dentistry, where the wavelength can be chosen to reduce the risk of damaging soft tissue. Both approaches are now being combined with robotics to automate treatments.[23]

Chip Fabrication

For a long time, the mass production of chips with structures smaller than 100 nanometers relied on the availability of high-average-power lasers that can produce light for lithography purposes, which involves transferring circuit patterns onto silicon wafers. In recent years, new processes have pushed average power requirements even higher, reflected in both peak power and the rate at which laser pulses can be generated. These new processes entail evaporating tin droplets with lasers to generate a plasma, which is then stimulated to produce extreme ultraviolet (EUV) light to project a mask that carries circuit

FIGURE 4.4 A laser is used to cut precise shapes in a metal form



Source: Shutterstock / Pixel B

patterns onto wafers. Since the process is not very efficient—a few hundred kilowatts to operate the laser generates only a few hundred watts of EUV light—power demands for state-of-the-art foundries have already grown dramatically and keep rising.

Producing structures smaller than 2 nanometers on very high-end chips relies completely on this technology. These chips are critical for applications that require high processing power, extremely energy-efficient operation, and miniaturization—requirements that characterize many systems of economic and national security importance. Although the capability originated from laser research programs in the United States, the chips are now being produced by a number of companies around the world, many of which are outside the United States.

Nuclear Fusion

Fusion occurs when two light atomic nuclei (usually deuterium and tritium, both isotopes of hydrogen) collide to form a heavier nucleus, releasing a large amount of energy in doing so. As an energy source, fusion energy is still in the research and development stage, as described in chapter 10 on sustainable energy technologies.

Today, the central issue in research on fusion for producing energy is the confinement problem—how to confine the fuel for long enough to ensure "ignition" of the fusion reaction. One approach to solving this problem is magnetic confinement fusion, which uses powerful magnets to contain and control a superheated plasma of deuterium and tritium. A second is inertial confinement fusion, which calls for rapidly compressing a deuterium-tritium fuel pellet using lasers to ignite the fusion reaction.

For inertial fusion energy to become commercially viable, high-energy, high-repetition-rate laser beams are needed to drive the samples to the extreme states required. The necessary lasers must deliver high energy beams at the relevant wavelength without the risk of damaging their components and with a much higher energy efficiency than is possible with current facilities.

In conjunction with operating the world's most powerful XFEL, SLAC is developing a major laser facility that will house a petawatt peak power laser and increase the beam energy of another of its lasers to hundreds of joules.[24] An important feature of this facility will be its ability to achieve highly symmetrical compressions of fusion targets, which will enable much more accurate measurements of implosion phenomena, both spatially and temporally, and support more precise modeling techniques.

Over the Horizon

Impact of Laser Technologies

As described earlier in this chapter, lasers are critical components across a wide range of applications, including communications, high-end chip production, defense, manufacturing, and medicine. This range is so broad that lasers could fairly be regarded as an enabling technology—that is, a technology whose existence and characteristics enable applications that would not otherwise be feasible or affordable.

Improving key laser figures of merit—peak power, energy, average power, pulse length, and wavelength—is a primary focus of extensive laser research. A recent Basic Research Needs report from the US Department of Energy's Office of Science emphasized that progress in all these areas is crucial for future scientific advances and new applications.[25] The report highlighted that progress requires novel approaches and techniques. It also noted the importance of additional engineering advances that address the limitations of current inefficient laser architectures and easily damaged optical components, calling for advancements in laser architectures, gain mediums, components, and control techniques.

Lasers could fairly be regarded as an enabling technology—that is, a technology whose existence and characteristics enable applications that would not otherwise be feasible or affordable.

Challenges of Innovation and Implementation

For lasers, the challenges of innovation and implementation are addressed in a highly distributed fashion—that is, across a multitude of laboratories and facilities. The reason is that progress in laser technology seems to be highly dependent on the specific application that requires a laser. An improvement in laser technology useful for application A may not be particularly useful for application B. For example, improvements in the average power of lasers used at the National Ignition Facility at LLNL will be of little value to lasers used in space-to-space communications. However, developments in beam pointing and alignment technology in a laser communications context may be helpful for laser weapons development, as some of the same problems arise with both the latter and the former.

Policy, Legal, and Regulatory Issues

Given that lasers are an enabling technology for many applications, public policy issues tend not to arise for lasers per se. Rather, they arise in the sectoral, societal, or policy context of a particular application. These issues could include the following:

Technological maturity Is laser technology at a state of maturity to support a given application? What are the alternatives to using lasers for that application? Is the growth path for a particular laser technology expansive and promising, or does it appear that it has plateaued?

Cost-effectiveness Are lasers really the best way to support a given application? Given total life-cycle costs, are there more cost-effective ways of performing the same missions?

Adequacy of the industrial base To what extent is the present industrial base capable of producing laser systems and components in necessary quantities? What resources are needed, if any, to develop its capacity for procuring a given laser-based system?

Dual-use considerations As laser technology advances, what are the implications, if any, for controlling dual-use laser technologies that have both military and civilian applications?

Environmental and safety concerns What, if any, are the environmental and safety concerns raised by the deployment of a given laser-based system? How should such concerns be addressed?

For illustrative purposes, consider how some of these questions might play out in two specific contexts.

Lasers as a defense against ballistic missiles The problem of using lasers for intercepting ballistic missiles is primarily characterized by the distance at which such intercepts must occur. Today, short-range rocket intercepts appear to be possible,[26] but longer-range intercepts are not, at least not with ground-based systems for most feasible laser technologies. Against short-range rockets, lasers have an economic advantage over missile interceptors, costing only a few dollars per laser shot as opposed to tens or hundreds of thousands of dollars per missile interceptor. Some technology usable for laser weapons has important civilian use—one example is deformable mirrors that can be used to enhance the quality of laser beams propagating through the atmosphere. When chemical lasers were contemplated for military use, environmental considerations were one negative aspect, as the lasers' exhaust was toxic.

Lasers for surgery Key concerns here include safety and cost-effectiveness. Safety guidelines for healthcare are constantly being updated and refined. For instance, in 2022 the American National Standards Institute released a new standard for the safe use of lasers that includes an updated section on maximum permissible medical-related exposures in terms of illuminance, or the amount of light allowed to fall on a given surface area.[27] In terms of cost, while some lasers for highly specific applications can be very expensive, others that can be used for multiple applications are much cheaper. For example, some excimer lasers, which emit short pulses of high-energy light and are used for medical procedures, such as LASIK and treating eczema, as well as in manufacturing, are available for under $100,000.

NOTES

1. Thales Group, "New World Record for Thales Lasers at ELI-NP," April 20, 2023, https://www.thalesgroup.com/en/worldwide-market-specific-solutions/lasers/news/new-world-record-thales-lasers-eli-np.

2. Lawrence Livermore National Laboratory, "LLNL's National Ignition Facility Delivers Record Laser Energy," US Department of Energy, November 16, 2023, https://www.llnl.gov/article/50616/llnls-national-ignition-facility-delivers-record-laser-energy.

3. Michael D. Perry, Paul S. Banks, Jason Zweiback, et al., "Laser Containing a Distributed Gain Medium," US Patent 6,937,629 B2, filed November 21, 2002, and issued August 30, 2005.

4. The Nobel Prize in Physics 2023, "Press Release," The Nobel Prize, October 3, 2023, https://www.nobelprize.org/prizes/physics/2023/press-release/.

5. Joshua Yang, Kasper Van Gasse, Daniil M. Lukin, et al., "Titanium:sapphire-on-insulator Integrated Lasers and Amplifiers," *Nature* 630 (June 2024): 853–59, https://doi.org/10.1038/s41586-024-07457-2; Nathan J. Szymanski, Bernardus Rendy, Yuxing Fei, et al., "An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials," *Nature* 624 (November 2023): 86–91, https://doi.org/10.1038/s41586-023-06734-w.

6. Bundesamt für Strahlenschutz, "Laser Applications," last modified March 14, 2024, https://www.bfs.de/EN/topics/opt/application-medicine-wellness/laser-applications/laser-applications_node.html.

7. Matthias Ureel, Marcello Augello, Daniel Holzinger, et al., "Cold Ablation Robot-Guided Laser Osteotome (CARLO®): From Bench to Bedside," *Journal of Clinical Medicine* 10, no. 3 (January 2021): 450, https://doi.org/10.3390/jcm10030450.

8. Jason Bates, "Invention to Impact: The Story of LASIK Eye Surgery," Science Matters, US National Science Foundation, March 15, 2024, https://new.nsf.gov/science-matters/invention-impact-story-lasik-eye-surgery.

9. Amir Yousef Sajjadi, Kunal Mitra, and Michael Grace, "Ablation of Subsurface Tumors Using an Ultra-Short Pulse Laser," *Optics and Lasers in Engineering* 49, no. 3 (March 2011): 451–56, https://doi.org/10.1016/j.optlaseng.2010.11.020.

10. Florian Kroll, Florian-Emanuel Brack, Constantin Bernert, et al., "Tumour Irradiation in Mice with a Laser-Accelerated Proton Beam," *Nature Physics* 18 (March 2022): 316–22, https://doi.org/10.1038/s41567-022-01520-3.

11. Congressional Research Service, *Department of Defense Directed Energy Weapons: Background and Issues for Congress* (R46925), July 11, 2024, https://crsreports.congress.gov/product/details?prodcode=R46925.

12. Hannah Weisman, "ASU Research Team Develops Breakthrough Ultrashort Laser Pulse Technology," ASU News, Arizona State University, June 16, 2023, https://news.asu.edu/20230616-asu-research-team-develops-breakthrough-ultrashort-laser-pulse-technology.

13. Owais Ali, "How is Laser Communication Used in Space?," AZo Optics, September 13, 2023, https://www.azooptics.com/Article.aspx?ArticleID=2474.

14. Evelyn Janeidy Arevalo, "SpaceX Reveals Operation of Over 8,000 'Space Lasers' Across Starlink Satellite Constellation—Enabling Faster Internet," Tesmanian, September 26, 2023, https://www.tesmanian.com/blogs/tesmanian-blog/starlink-video.

15. Andrew Jones, "Pew! Pew! Pew! NASA's 1st Successful Two-way Laser Experiment Is a Giant Leap for Moon and Mars Communications," Space.com, Future US, Inc., December 19, 2023, https://www.space.com/nasa-laser-communication-1st-two-way-link-iss.

16. Dmitry Solenov and Dmitry Mozyrsky, "Cold Atom Qubits," *Journal of Computational and Theoretical Nanoscience* 8, no. 3 (March 1, 2011): 481–89, https://doi.org/10.1166/jctn.2011.1713.

17. TechPort, "Rapid Response Debris Removal Using Reconfigurable Space-Based Laser Networks," National Aeronautics and Space Administration, accessed October 13, 2023, https://techport.nasa.gov/view/156377; Micaela Morrissette, "WVU Engineer Developing Laser System to Defend Space Assets from Debris in Earth's Orbit," WVU Today, West Virginia University, October 4, 2023, https://wvutoday.wvu.edu/stories/2023/10/04/wvu-engineer-developing-laser-system-to-defend-space-assets-from-debris-in-earth-s-orbit.

18. Linac Coherent Light Source, "LCLS-II-HE," SLAC National Accelerator Laboratory, accessed October 13, 2023, https://lcls.slac.stanford.edu/lcls-ii-he.

19. Mengning Liang, Garth J. Williams, Marc Messerschmidt, et al., "The Coherent X-Ray Imaging Instrument at the Linac Coherent Light Source," *Journal of Synchrotron Radiation* 22, no. 3 (May 2015): 514–19, https://doi.org/10.1107/S160057751500449X.

20. Jack Griffiths, Ana F. Suzana, Longlong Wu, et al., "Resolving Length-Scale-Dependent Transient Disorder Through an Ultrafast Phase Transition," *Nature Materials* 23 (June 2024): 1041–47, https://doi.org/10.1038/s41563-024-01927-8.

21. Asmit Bhowmick, Rana Hussein, Isabel Bogacz, et al., "Structural Evidence for Intermediates during O_2 Formation in Photosystem II," *Nature* 617 (May 2023): 629–36, https://doi.org/10.1038/s41586-023-06038-z.

22. Oskar Hoffman, Jochen Stollenwerk, and Peter Loosen, "Design of Multi-Beam Optics for High Throughput Parallel Processing," *Journal of Laser Applications* 32, no. 1 (February 2020): 012005, https://doi.org/10.2351/1.5125778.

23. Ureel et al., "Cold Ablation Robot-Guided Laser Osteotome."

24. Linac Coherent Light Source, "MEC-U," SLAC National Accelerator Laboratory, accessed October 13, 2023, https://lcls.slac.stanford.edu/mec-u.

25. Basic Research Needs Workshop on Laser Technology, *Report of the Basic Research Needs Workshop on Laser Technology: 2023 Report*, Office of Science, US Department of Energy, January 5, 2024, https://science.osti.gov/-/media/ardap/pdf/2024/Laser-Technology-Workshop-Report_20240105_final.pdf.

26. Mikayla Easley, "Israeli-Made High-Energy Laser Makes Debut," *National Defense*, March 21, 2023, https://www.nationaldefensemagazine.org/articles/2023/3/21/israeli-made-high-energy-laser-makes-debut.

27. ANSI Webstore, "ANSI Z136.1-2022: American National Standard for Safe Use of Lasers," American National Standards Institute, accessed October 13, 2024, https://webstore.ansi.org/standards/lia/ansiz1362022; Brad Kelechava, "ANSI Z136.1-2022: Safe Use of Lasers," *ANAB Blog* (blog), American National Standards Institute, March 13, 2023, https://blog.ansi.org/ansi-z136-1-2022-safe-use-of-lasers/.

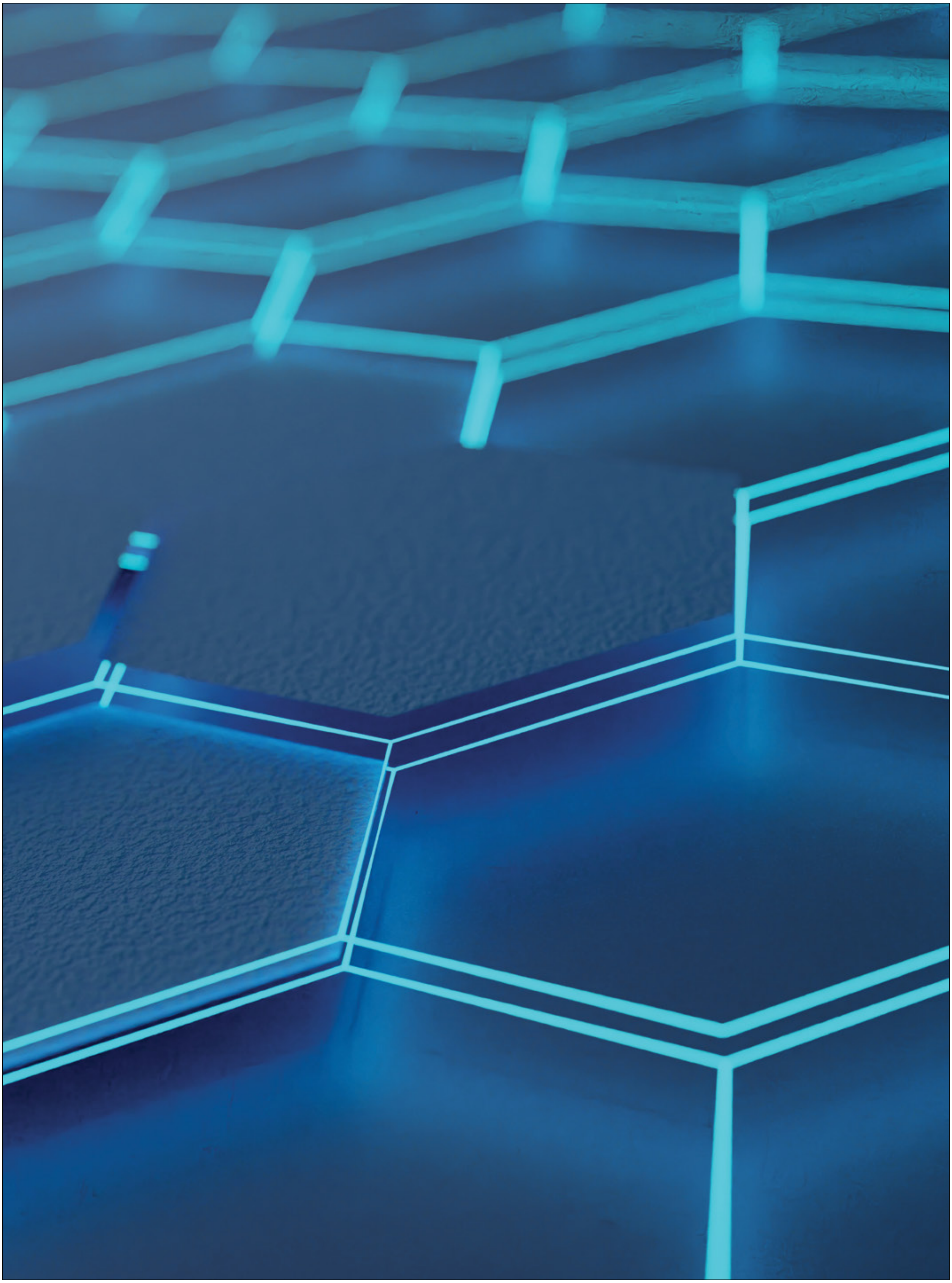
STANFORD EXPERT CONTRIBUTORS

Dr. Siegfried Glenzer
SETR Faculty Council, Professor of Photon Science, and Professor, by courtesy, of Mechanical Engineering

Dr. Norbert Holtkamp
SETR Advisory Board and Professor of Particle Physics and Astrophysics and of Photon Science

Dr. Eric Galtier
Lead Scientist at the SLAC National Accelerator Laboratory

Dr. Nicholas Hartley
SETR Fellow and Associate Scientist at the SLAC National Accelerator Laboratory



05

MATERIALS SCIENCE

KEY TAKEAWAYS

- Materials science is a foundational technology that underlies advances in many other fields, including robotics, space, energy, and synthetic biology.
- Materials science will exploit artificial intelligence as another promising tool to predict new materials with new properties and identify novel uses for known materials.
- Future progress in materials science requires new funding mechanisms to more effectively transition from innovation to implementation and access to more computational power.

Overview

From semiconductors in computer chips to plastics in everyday objects, materials are everywhere. Knowing how to synthesize and process them, as well as understanding their structure and properties, has helped to shape the world around us. Materials science contributes to the development of stronger, lighter, and more flexible materials that improve everything from battery electrodes to medical implants and from automobiles to spacecraft.

It is a broad field. At Stanford University, for example, faculty working on materials science research programs are found in many departments, including Materials Science and Engineering, Chemical Engineering, Electrical Engineering, Bioengineering, Chemistry, and Physics.

Researchers in materials science study the properties and behavior of materials to understand and predict their structure and performance under various conditions. Their goal is to develop new materials with desirable properties and improve existing ones by understanding how the structure of a material influences its properties and how processing it can change its structure and therefore its performance. This knowledge can then be used to design new materials with desirable properties for specific uses.

Broadly speaking, materials science and engineering research focuses on four major areas. The first is the study of the structure of materials to understand how they are composed and organized from atomic to macroscopic scales. The second involves verifying the properties of materials, such as their conductivity, strength, and elasticity. The third area covers analysis and benchmarking of how materials perform in specific situations. The final one involves assessing how materials can be fabricated and manufactured. Characterization of materials, or the general process by which their structure and properties are ascertained through spectroscopic, microscopic, and several other complementary methods, underlies these four areas.

Basics of Materials Science

All materials are composed of atoms. The periodic table of the elements (figure 5.1) lists all the known types of atoms. Certain atoms can be combined with others into molecules that have vastly different properties than the individual atoms involved. For example, table salt consists of sodium and chlorine, which are elements. Sodium burns on contact with water, and chlorine is a poisonous gas, yet the table salt we consume every day is a completely different substance.

There are two important points to note about the periodic table. First, there are a lot of elements—ninety-two naturally occurring ones and twenty-six that can be observed only in laboratory conditions. That's a lot of building blocks from which different

FIGURE 5.1 The periodic table of the elements

Group

Period	1	2		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H																		2 He
2	3 Li	4 Be												5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg												13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca		21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr		39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	*	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	* *	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Nh	114 Fl	115 Mc	116 Lv	117 Ts	118 Og
			*	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb		
			* *	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No		

Source: Adapted from Wikimedia Commons, CC BY-SA 4.0

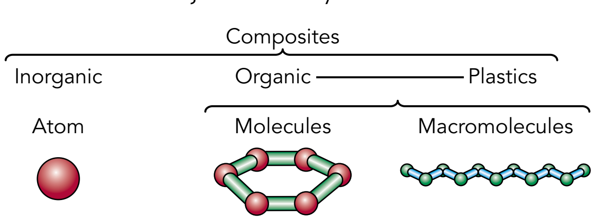
materials and molecules can be synthesized, and an astronomically large number of different compounds are possible. The challenge for materials science is to sift through this vast array of possibilities to find the ones that are useful. Machine-learning (ML) algorithms can accelerate this process by predicting the properties of materials and identifying promising candidates for synthesis, significantly reducing the time and resources required for experimental testing. (More detail on this subject is provided later in this chapter.)

The second important point is that the elements in the periodic table are lined up in a certain order. Those in the same column have properties that are often similar in key ways. This means insights developed through experimentation or calculation on one element also apply, with some modifications, to another element above or below it in the periodic table.

Atoms can be arranged spatially in various ways. A crystal, for example, is the result of arranging atoms in a periodically repeating lattice. The silicon wafer at the heart of the semiconductor industry is one such crystal; more precisely, it's a slice of a single silicon crystal.

Molecules, which are composed of atoms, can, in turn, be linked together into structures called macromolecules (see figure 5.2). These can occur naturally, as is the case for proteins, DNAs, and cellulose, or they can be synthesized artificially and used to create things such as polymers/plastics. Long chains

FIGURE 5.2 Objects of study in materials science



of macromolecules often mean the material is more flexible, making many plastics possible. Research on new macromolecular structures can be used to develop plastic materials that are easier to recycle or have advantageous mechanical properties while weighing less than metals.

Key Developments

Some interesting present-day applications of materials science are discussed below.

Flexible Electronics

Flexible or stretchable electronics involves the creation of electrical devices that can bend, stretch, and deform without compromising their performance. Such electronics can be used as wearable, skinlike devices. For example, "electronic skin," or e-skin, can conform to real skin and sense things such as temperature and pressure, as well as encode these into electrical signals.[1] A "smart bandage" with integrated sensors to monitor wound conditions and with electrical stimulation can accelerate the time needed to heal chronic wounds by 25 percent.[2]

Recent research has also shown the development of integrated circuits on soft and flexible substrates can drive a micro-LED (light-emitting diode) screen that can read out a braille array ten times more sensitively than human fingertips.[3] The latest version of this device is many times smaller than previous iterations and operates three orders of magnitude faster than them.

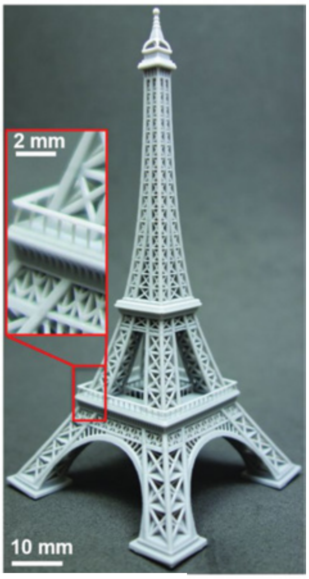
This performance is achieved using a combination of carbon nanotubes (tubular molecules made up of carbon atoms), a one-dimensional conductor, and a flexible polymer substrate. This substrate encases the carbon nanotubes and maintains a connected internal electrical circuit while bending and deforming. The transistors it uses are built onto this pliable substrate and enable the device to power the micro-LED display.

Commercialization will require more work to eliminate variations in electrical properties of the device's circuits when mounted on a moving host, as well as work to eliminate its susceptibility to moisture. Still, the research team responsible for this invention envisions using it to make biocompatible probes for the brain and gut that are more capable and energy efficient than current ones. This could pave the way to more complex and longer-lasting brain-machine interfaces.

Additive Manufacturing

One of the most promising advances in materials processing over the past fifteen years is additive manufacturing, colloquially known as 3-D printing. The technology comes in different forms. For instance, a method known as continuous liquid interface production (CLIP) uses directed ultraviolet (UV) light to form structures from a polymer resin (see figure 5.3).[4] A key aspect of CLIP is its use of an oxygen-permeable window placed above a UV light projector that prevents the resin from curing in unwanted places.

FIGURE 5.3 A CLIP-based 3-D printer created a miniature print of the Eiffel Tower



Source: Carbon Inc. / John Tumbleston

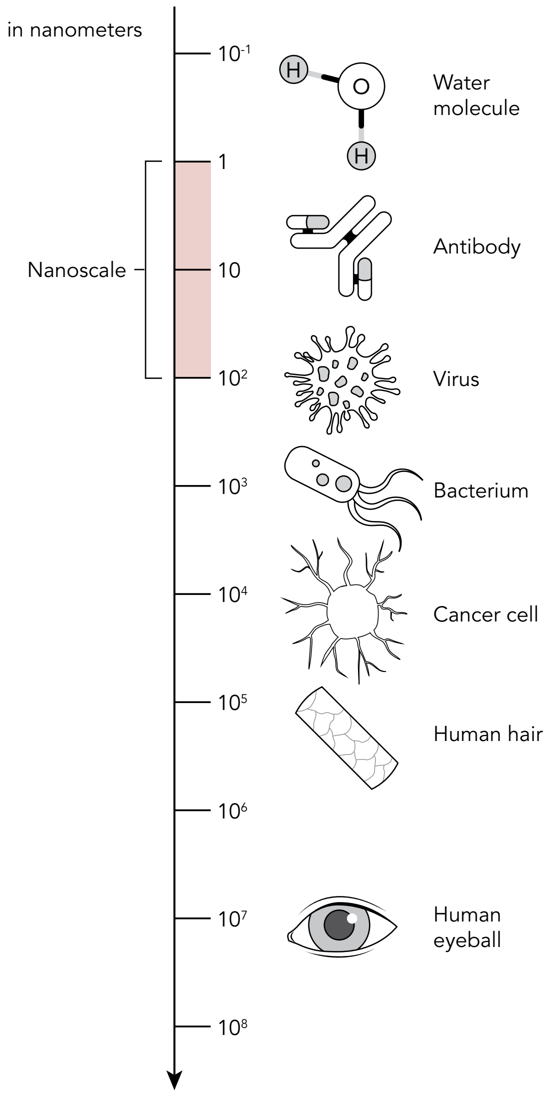
Especially at high speeds, 3-D printing struggles with producing small features. The 3-D printing process requires several components, including the material resin, the light source, and the build platform where an object is printed, to perform in concert. That is technically challenging, but by printing on a tensioned film made from polyethylene terephthalate that's fed through a CLIP printer, it's possible to 3-D print very small particles at a pace of one million a day from a single machine.[5]

Nanotechnology

Nanotechnology is a large and growing subfield of materials science. Size has a profound impact on the properties of a material. Figure 5.4 compares the length of a water molecule (below a nanometer), a human hair (roughly 10^5 nanometers), and a human eyeball (at 10^7 nanometers). A structure is typically referred to as nanoscale if at least one of its dimensions is in the 1-to-100-nanometer range.

In the past twenty years, nanoscience and nanotechnology have attracted enormous interest for two reasons. First, many significant biological organisms, such as viruses and proteins, are nanoscale in size. Second, it turns out that the properties of nanoscale materials—including their electronic, optical, magnetic, thermal, and mechanical properties—are often very different from the same material in bulk form.[6] Materials that are smaller than about 100 nanometers in one dimension, two dimensions, or all dimensions are called nanosheets, nanowires, and nanoparticles, respectively (see figure 5.5). Somewhat more complex shapes include nanotubes, which are flexible hollow tubes made of carbon atoms, and nanorods, which are slightly wider, rigid structures.

FIGURE 5.4 The size of nanoscale objects



Quantum dots—for which the Nobel Prize in Chemistry was awarded in 2023—have garnered public attention through their use in televisions. They are metallic, carbonaceous, or semiconductor spherical nanocrystals that emit bright monochromatic light in response to excitation by a light source with a higher energy, such as blue light from the back panel in a display.[7] Quantum dots are a model example of how size affects a material's properties because their optoelectronic properties differ from those of the same bulk material. The diameter of quantum dots determines the color of light that they produce, with larger ones emitting longer wavelengths. This enables tunable light emission based on the desired application.

Some current uses of the technology include the following:

Medical imaging Quantum dots can improve biomedical imaging by, for example, acting as fluorescent markers that make it possible to selectively label biological structures in vitro and in vivo.[8] Biocompatible nanomaterials can also be employed as optical probes that sense mechanical forces and electrical fields in biological organisms, removing the need for bulky, specialized equipment and making new experiments possible.[9]

Solar cells Quantum dots' ability to absorb different frequencies of light means they can potentially capture more of the solar spectrum, boosting the performance of solar panels.[10]

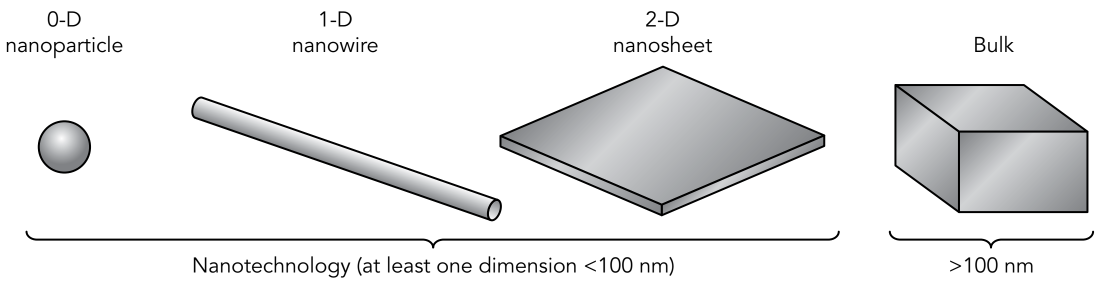
Sensors Quantum dots can be used in sensors for detecting chemicals and biological substances.[11]

Anticounterfeiting When they are embedded in product labels, quantum dots can help defend against counterfeiting.[12]

Apart from applications related to quantum dots, other examples of applications of nanomaterials include the following:

Pharmaceutical delivery An injectable polymer-nanoparticle hydrogel has been developed to precisely control the delivery of drugs, proteins, and cells.[13] The efficacy of insulin administration can also be improved through this research.[14] Among other innovative uses, nanoparticles can be engineered to permeate the blood–brain barrier, delivering drugs to treat neurodegenerative diseases.[15]

FIGURE 5.5 Dimensionality of nanomaterials



Vaccine stabilization Nanoassemblies can be used to stabilize certain types of vaccines—that is, protect their components from degradation—by encapsulating them.[16] In this form, it is easier to inject a vaccine into the human body and to ensure its release over time inside the body in a controlled manner. This is especially useful for mRNA vaccines such as the ones developed for COVID-19.

Smart windows Silver nanowires arranged into a thin film on a window become a transparent conductive surface. Running a current through the film can then change the opacity of the window electrically.[17]

Two-dimensional (2-D) semiconductors, graphene, carbon nanotubes, and nanoscale materials These are at the forefront of the next generation of high-tech electronic devices. Active research efforts are designing new methods to integrate 2-D or carbon nanotube semiconductors into electronics that are currently silicon-based to improve their energy efficiency and heat management.[18] (A 2-D semiconductor is a semiconductor with atomic-scale thickness.)

Higher-capacity batteries High-performance lithium battery anodes have been developed by integrating silicon nanowires as an anode material. When bulk silicon is used as an anode, it undergoes significant changes in volume as a battery charges and discharges, often leading to mechanical failure. Silicon nanowires solve this problem and deliver a tenfold increase in battery capacity.[19]

Catalysis Catalysts are important anywhere a chemical reaction must be speeded up to be useful. For example, catalytic converters in cars use platinum and palladium catalysts to rapidly break down carbon monoxide into carbon dioxide (CO_2). Nanomaterials are particularly well suited for this role.[20] This is because their high surface-to-volume ratio allows many more active catalytic sites to participate in a reaction than would be the case for the same material in bulk. Nanomaterials can also be chemically architected to catalyze various reactions. Advances have been made in converting CO_2 to value-added chemicals using electrified nanoparticle catalysts and in employing palladium catalysts for the combustion of methane, which could improve the efficiency of electricity generation from the gas.[21] Nanocatalysts have also been used to improve the rate at which hydrogen can be produced from water through electrolysis.[22] These approaches still face challenges, including developing catalysts that are sufficiently active and stable—as well as cheap enough—to inexpensively produce hydrogen in large quantities.[23]

Biosensing

Detecting pathogenic bacteria usually takes a long time, often involving culture-based methods or

One of the foremost challenges of materials science as a discipline is the vast number of possible materials and material combinations that can be used and the associated time and cost involved in synthesis and characterization.

polymerase chain reaction, a lab technique that reproduces DNA sequences for study. A new method has been developed that uses Raman spectroscopy—a means of analyzing chemical structures—to rapidly identify pathogenic bacteria in blood samples based on their unique optical signatures when exposed to laser light.[24] This system creates nanodroplets from blood samples, with each droplet containing only a few cells. Adding gold nanorods to these droplets makes it easier to detect dangerous bacteria because the rods adhere to suspect cells and amplify their signature in a spectrographic analysis. An ML algorithm can then use the results to determine if a pathogen is present.

The Application of Artificial Intelligence in Materials Science

One of the foremost challenges of materials science as a discipline is the vast number of possible materials and material combinations that can be used and the associated time and cost involved in synthesis and characterization. ML offers promising solutions by leveraging experimental and computational data on the properties of materials.[25] ML algorithms can recognize patterns in existing data and make generalized predictions about new materials. While this approach has been successful with relatively simple materials, much remains to be done when dealing with complicated ones.

ML provides a starting point for further exploration, but additional data is needed to make ML-informed solutions more accurate, especially in the case of complex materials. Databases such as the Materials Project led by Lawrence Berkeley National Laboratory represent a significant effort,[26] but they still have limitations in terms of the range of properties covered. To truly understand and predict the properties of materials, such as their thermal conductivity or optical characteristics, more accurate, comprehensive, and tailored databases are needed. Among their many applications, these could help accelerate the development of materials that enable researchers to overcome bottlenecks in chip assembly as semiconductors continue to be miniaturized.

Several major industry efforts are already underway to harness artificial intelligence (AI) for materials exploration. Google DeepMind's Graph Networks for Materials Exploration project aims to use neural networks—ML models that process data in ways similar to human brains—to predict new materials. Other companies like IBM, Citrine Informatics, and MaterialsZone are combining materials science expertise with data science and AI to accelerate materials development and optimize product design.

In addition to those mentioned above, some other current applications of ML in materials science include the following:

Knowledge discovery The technology has been used in materials science to examine the scientific literature for hidden relationships and predict new material properties. For example, researchers using an ML model were able to analyze patterns in scientific abstracts.[27] The model was trained to calculate

FIGURE 5.6 The A-Lab combines AI-guided synthesis with automated materials characterization

Source: © 2023 The Regents of the University of California, Lawrence Berkeley National Laboratory

the probability that a material's name would co-occur with the word *thermoelectric* in the analyzed text. The researchers then identified abstracts involving other materials that had words with close semantic relationships to *thermoelectric* but that had never been studied as such. Remarkably, their model's top predictions were eight times more likely to be studied as thermoelectrics in the succeeding five years compared to randomly chosen materials. This demonstrates the potential of AI to uncover latent knowledge in existing scientific literature and to guide future research directions.

Battery electrolyte design Improving electrolyte design in batteries can enhance their performance, but predicting and designing effective electrolytes is challenging because of their complexity. ML has been applied to a data-driven electrolyte design for lithium-metal batteries. Liquid electrolytes play an important role in determining how well these batteries can cycle, or charge and discharge. Researchers used models to analyze electrolyte compositions and identify key factors for better performance.[28] They discovered that lower oxygen content in the solvent leads to improved battery cycling and used this insight to develop new electrolytes that improved the process.

Automated labs To address the challenge of limited experimental data, there is growing interest in developing autonomous laboratories that can rapidly synthesize and characterize materials at scale. For example, the A-Lab developed by researchers at the University of California–Berkeley is a robotic platform that combines AI-guided synthesis with automated characterization to enable materials discovery (see figure 5.6).[29] The A-Lab system consists of several key components:

- A robotic arm that picks, weighs, and mixes dry chemical precursors, which are compounds needed for use in chemical reactions
- A sample preparation station where the precursors are turned into a slurry

- A conveyor belt system that transports the samples to different stations
- An oven to allow precursors to react, with the environment determined by an ML algorithm
- A grinding station where the product is turned into a fine powder
- Automated characterization equipment to validate morphology and elemental distribution

The A-Lab demonstration showcased the potential for autonomous materials synthesis and characterization. Out of 58 targeted materials, the authors claimed 36 successful syntheses and 7 partial successes. This high-throughput approach could significantly accelerate the process of validating computational predictions and generating new experimental data to train ML models.

Over the Horizon

Machine Learning in Materials Science

Over the horizon, there is the hope that ML-guided approaches will dramatically shorten the timescale for materials discovery and enable the design of materials optimized for specific applications. Continued development of both bottom-up computational approaches and top-down experimental data-driven methods will be needed to bridge the gap between fundamental material parameters and real-world device performance.

Computational approaches will need to be validated. For example, the A-Lab described above has faced some criticism from the scientific community. One critique centered on concerns about the accuracy of its characterization work and claims of new material synthesis,[30] pointing to errors in the A-Lab's analysis that included poor and incorrect fits of structural models to the experimental data, among other issues. Such criticisms highlight the need for continued validation and human oversight in AI-driven materials research.

Although much remains to be done, this field undoubtedly has great promise for optimizing complex material properties. For instance, in the future, automated laboratories may be able to accurately predict new materials, balancing the need for quick, high-volume calculations with the desire for precise, high-quality real-world results. They may also be able to predict and create actual material samples with sufficient accuracy to reduce the amount of human effort needed to confirm the automated analysis.

The integration of AI and ML into materials science presents immense opportunities for accelerating discovery and innovation. By combining advanced algorithms with expanded databases, automated experimentation, and increased computational resources, researchers aim to navigate the vast landscape of possible materials more efficiently than ever before.

Challenges of Innovation and Implementation

The materials science research infrastructure does not adequately support the transition from research to real-world applications at scale. Such transitions generally require construction of a small-scale pilot project to demonstrate the feasibility of potential large-scale manufacturing. At this point, the technology is too mature to qualify for most research funding—because basic science does not address issues related to scaling up—but not mature enough to be commercialized by actual companies. Neither government nor venture capital investors are particularly enthusiastic about financing pilot projects, so different forms of funding are required to bridge this gap between bench-scale research and company-level investment. Such support could also establish national rapid prototyping centers, where academic researchers can find the help and tools necessary to build prototypes and pilot plants for their technology.

Historically, the United States has led the world in nanotechnology, but the gap between it and China has narrowed.

Past research processes are also ill-suited to rapid transitions to real-world applications. Such processes emphasize sequential steps. The standard process has been to characterize a material and then proceed to a simple demonstration of how it might be used. Today, addressing big societal challenges calls for a more scalable, system-level approach that involves extensive rapid prototyping and fast, reliable demonstrations to provide feedback on the potential value of specific materials and to fill in knowledge gaps.

Current infrastructure makes this difficult. For example, in collaborations with a medical school, it is often necessary to bring almost-finished products to clinical tests to validate the true impact of a new medical device using innovative materials. With typically less than a thirty-minute window to place a device on a patient and gather data, any malfunction, such as a sudden equipment failure or a loose wire, can jeopardize an entire experiment and potentially halt future patient interactions. Lab-assembled devices may not meet this standard of reliability, even if they do demonstrate the value of the underlying science.

Policy, Legal, and Regulatory Issues

REGULATION OF PRODUCTS INCORPORATING NANOMATERIALS

As with regulation in other areas of technology, materials science faces concerns about the appropriate balance between the need to ensure public safety and the imperative to innovate quickly and leapfrog possible competitors. In the biomedical space, the US Food and Drug Administration (FDA) created a Nanotechnology Regulatory Science Research Plan in 2013.[31] Today, FDA regulation and review of nanotechnology is governed by Executive Order 13563.[32] Outside of biomedicine, regulation of and infrastructure for nanomaterials research from the government side is based largely in agencies of the National Nanotechnology Initiative, which include the Department of Energy, the National Cancer Institute, the National Institutes of Health, the National Institute of Standards and Technology in the Department of Commerce, and the National Science Foundation.

TOXICITY AND ENVIRONMENTAL ISSUES

Nanoparticles raise particular concerns because their small size may enable them to pass through various biological borders such as cell membranes or the blood–brain barrier, potentially affecting biological systems in harmful ways. Nanoscale particles inhaled into the lungs, for example, may lodge themselves there permanently, causing severe health outcomes, including pulmonary inflammation, lung cancer, and penetration into the brain and skin.[33]

Moreover, because engineered nanoparticles are, by definition, new to the natural environment, they pose unknown dangers to humans and the environment. These concerns include managing the risk of incorporating nanomaterials into products that enter the environment at the end of their life cycles. As nanomaterials are employed in, and considered for, electronic and energy products, it is paramount that they safely degrade or can be recycled at the end of a product's life. Policy will be particularly important in shaping responsible end-of-life solutions for products incorporating them.

Finally, end-of-life considerations that take into account environmental sustainability and resource conservation are inherently a part of developing and distributing new materials. This is especially important for plastics and materials containing per- and polyfluoroalkyl substances (PFAS), which pose significant environmental and health risks. Material developers can incorporate recyclability into their design processes. For PFAS and other persistent chemicals, the US Environmental Protection Agency strategic road map of October 2021 provides guidance for their use and disposal and calls for research into safe alternatives and effective degradation methods.[34]

FOREIGN COLLABORATION AND COMPETITION

Historically, the United States has led the world in nanotechnology, but the gap between it and China has narrowed. Notably, in 2016, the president of the Chinese Academy of Sciences openly announced Beijing's ambition to compete in the field of nanotechnology.[35]

As great power competition intensifies, many researchers are concerned that fundamental research could now be subject to export controls. Policy ambiguity can inadvertently hinder innovation by creating obstacles for non-US researchers wishing to contribute to work in America and by deterring international collaborations with allies and partners who are important for advancing the field. In nanomaterials, for example, researchers in South Korea are making significant strides with biomedical applications and ones for consumer electronics. There is an urgent need for clarification of these policies, particularly those delineating fundamental research and export-controlled research.

INFRASTRUCTURE FOR ML-ASSISTED MATERIALS SCIENCE

The United States benefits from having some of the world's largest supercomputing resources, which are essential not only for ML but also for developing extensive databases. However, better access to computing power is essential for researchers in materials science to generate and analyze databases effectively. Greater access to data, including to databases that might not always be openly available to academics, is also needed.

One additional area where policymakers could have a significant impact is in bridging the gap between the scientific community and makers of computational hardware. Frequent changes in computing architectures can lead to a loss of productivity for researchers because code must be constantly updated. Improved collaboration with hardware manufacturers and other providers of computing resources could ensure scientific needs are better aligned with advances in computing technology, enhancing overall research efficiency.

NOTES

1. Weichen Wang, Yuanwen Jiang, Donglai Zhong, et al., "Neuromorphic Sensorimotor Loop Embodied by Monolithically Integrated, Low-Voltage, Soft E-Skin," *Science* 380, no. 6646 (2023): 735–42, https://doi.org/10.1126/science.ade0086.

2. Yuanwen Jiang, Artem A. Trotsyuk, Simiao Niu, et al., "Wireless, Closed-Loop, Smart Bandage with Integrated Sensors and Stimulators for Advanced Wound Care and Accelerated Healing," *Nature Biotechnology* 41 (2023): 652–62, https://doi.org/10.1038/s41587-022-01528-3.

3. Donglai Zhong, Can Wu, Yuanwen Jiang, et al., "High-Speed and Large-Scale Intrinsically Stretchable Integrated Circuits," *Nature* 627 (March 2024): 313–20, https://doi.org/10.1038/s41586-024-07096-7.

4. John R. Tumbleston, David Shirvanyants, Nikita Ermoshkin, et al., "Continuous Liquid Interface Production of 3D Objects," *Science* 347, no. 6228 (March 2015): 1349–52, https://doi.org/10.1126/science.aaa2397; Kaiwen Hsiao, Brian J. Lee, Tim Samuelsen, et al., "Single-Digit-Micrometer-Resolution Continuous Liquid Interface Production," *Science Advances* 8, no. 46 (November 2022), https://doi.org/10.1126/sciadv.abq2846.

5. For more on polyethylene terephthalate, see Polymershapes, "PET Plastic Film," accessed October 13, 2023, https://www.polymershapes.com/pet-plastic-film/; Jason M. Kronenfeld, Lukas Rother, Max A. Saccone, et al., "Roll-to-Roll, High-Resolution 3D Printing of Shape-Specific Particles," *Nature* 627 (March 2024): 306–12, https://doi.org/10.1038/s41586-024-07061-4.

6. Hui Pan and Yuan Ping Feng, "Semiconductor Nanowires and Nanotubes: Effects of Size and Surface-to-Volume Ratio," *ACS Nano* 2, no. 11 (November 2008): 2410–14, https://doi.org/10.1021/nn8004872; Anna C. Balazs, Todd Emrick, and

Thomas P. Russell, "Nanoparticle Polymer Composites: Where Two Small Worlds Meet," *Science* 314, no. 5802 (November 2006): 1107–10, https://doi.org/10.1126/science.1130557.

7. A. P. Alivisatos, "Semiconductor Clusters, Nanocrystals, and Quantum Dots," *Science* 271, no. 5251 (February 1996): 933–37, https://doi.org/10.1126/science.271.5251.933.

8. X. Michalet, F. F. Pinaud, L. A. Bentolila, et al., "Quantum Dots for Live Cells, in Vivo Imaging, and Diagnostics," *Science* 307, no. 5709 (January 2005): 538–44, https://doi.org/10.1126/science.1104274.

9. Randy D. Mehlenbacher, Rea Kolbl, Alice Lay, et al., "Nanomaterials for In Vivo Imaging of Mechanical Forces and Electrical Fields," *Nature Reviews Materials* 3, no. 17080 (2018), https://doi.org/10.1038/natrevmats.2017.80.

10. Prashant V. Kamat, "Quantum Dot Solar Cells: Semiconductor Nanocrystals as Light Harvesters," *Journal of Physical Chemistry* C 112, no. 48 (October 2008): 18737–53, https://doi.org/10.1021/jp806791s; Ralph Nuzzo, Harry Atwater, Andrei Faraon, et al., "Light Material Interactions in Energy Conversion (Final Report)," Office of Scientific and Technical Information, US Department of Energy, April 1, 2019, https://doi.org/10.2172/1504275.

11. Babatunde Ogunlade, Loza F. Tadesse, Hongquan Li, et al., "Predicting Tuberculosis Drug Resistance with Machine Learning-Assisted Raman Spectroscopy," arXiv, Cornell University, June 9, 2023, https://doi.org/10.48550/arXiv.2306.05653; Fareeha Safir, Nhat Vu, Loza F. Tadesse, et al., "Combining Acoustic Bioprinting with AI-Assisted Raman Spectroscopy for High-Throughput Identification of Bacteria in Blood," *Nano Letters* 23, no. 6 (March 2023): 2065–73, https://doi.org/10.1021/acs.nanolett.2c03015.

12. Yang Liu, Fei Han, Fushan Li, et al., "Inkjet-Printed Unclonable Quantum Dot Fluorescent Anti-Counterfeiting Labels with Artificial Intelligence Authentication," *Nature Communications* 10, no. 2409 (June 2019), https://doi.org/10.1038/s41467-019-10406-7.

13. Hector Lopez Hernandez, Abigail K. Grosskopf, Lyndsay M. Stapleton, et al., "Non-Newtonian Polymer–Nanoparticle Hydrogels Enhance Cell Viability During Injection," *Macromolecular Bioscience* 19, no. 1 (January 2019), https://doi.org/10.1002/mabi.201800275.

14. Caitlin L. Maikawa, Leslee T. Nguyen, Joseph L. Mann, et al., "Formulation Excipients and Their Role in Insulin Stability and Association State in Formulation," *Pharmaceutical Research* 39 (2022): 2721–28, https://doi.org/10.1007/s11095-022-03367-y; Joseph L. Mann, Caitlin L. Maikawa, Anton A. A. Smith, et al., "An Ultrafast Insulin Formulation Enabled by High-Throughput Screening of Engineered Polymeric Excipients," *Science Translational Medicine* 12, no. 550 (July 2020), https://doi.org/10.1126/scitranslmed.aba6676.

15. Vladimir P. Torchilin, "Multifunctional, Stimuli-Sensitive Nanoparticulate Systems for Drug Delivery," *Nature Reviews Drug Discovery* 13 (November 2014): 813–27, https://doi.org/10.1038/nrd4333; Cláudia Saraiva, Catarina Praça, Raquel Ferreira, et al., "Nanoparticle-Mediated Brain Drug Delivery: Overcoming Blood-Brain Barrier to Treat Neurodegenerative Diseases," *Journal of Controlled Release* 235 (August 2016): 34–47, https://doi.org/10.1016/j.jconrel.2016.05.044.

16. Norbert Pardi, Michael J. Hogan, Frederick W. Porter, et al., "mRNA Vaccines—A New Era in Vaccinology," *Nature Reviews Drug Discovery* 17 (2018): 261–79, https://doi.org/10.1038/nrd.2017.243.

17. Zhiqiang Niu, Fan Cui, Elisabeth Kuttner, et al., "Synthesis of Silver Nanowires with Reduced Diameters Using Benzoin-Derived Radicals to Make Transparent Conductors with High Transparency and Low Haze," *Nano Letters* 18, no. 8 (2018): 5329–44, https://doi.org/10.1021/acs.nanolett.8b02479.

18. Weisheng Li, Xiaoshu Gong, Zhihao Yu, et al., "Approaching the Quantum Limit in Two-Dimensional Semiconductor Contacts," *Nature* 613 (2023): 274–79, https://doi.org/10.1038/s41586-022-05431-4; Eric Pop, "Energy Dissipation and Transport in Nanoscale Devices," *Nano Research* 3 (2010): 147–69, https://doi.org/10.1007/s12274-010-1019-z.

19. Candace K. Chan, Hailin Peng, Gao Liu, et al., "High-Performance Lithium Battery Anodes Using Silicon Nanowires," *Nature Nanotechnology* 3 (December 2008): 31–35, https://doi.org/10.1038/nnano.2007.411.

20. U. P. M. Ashik, Anchu Viswan, Shinji Kudo, et al., "Chapter 3. Nanomaterials as Catalysts," in *Applications of Nanomaterials: Advances and Key Technologies*, eds. Sneha Mohan Bhagyaraj, Oluwatobi Samuel Oluwafemi, Nandakumar Kalarikkal, et al. (Cambridge, MA: Elsevier, 2018), 4582, https://doi.org/10.1016/B978-0-08-101971-9.00003-X.

21. Weixin Huang, Aaron C. Johnston-Peck, Trenton Wolter, et al., "Steam-Created Grain Boundaries for Methane C–H Activation in Palladium Catalysts," *Science* 373, no. 6562 (September 2021): 1518–23, https://doi.org/10.1126/science.abj5291; Chengshuang Zhou, Arun S. Asundi, Emmett D. Goodman, et al., "Steering CO_2 Hydrogenation Toward C–C Coupling to Hydrocarbons Using Porous Organic Polymer/Metal Interfaces," *Proceedings of the National Academy of Sciences* 119, no. 7 (February 2022), https://doi.org/10.1073/pnas.2114768119.

22. Thomas F. Jaramillo, Kristina P. Jørgensen, Jacob Bonde, et al., "Identification of Active Edge Sites for Electrochemical H_2 Evolution from MoS_2 Nanocatalysts," *Science* 317, no. 5834 (July 2007): 100–102, https://doi.org/10.1126/science.1141483.

23. Zhi Wei Seh, Jakob Kibsgaard, Colin F. Dickens, et al., "Combining Theory and Experiment in Electrocatalysis: Insights into Materials Design," *Science* 355, no. 6321 (January 2017), https://doi.org/10.1126/science.aad4998.

24. Safir et al., "Combining Acoustic Bioprinting."

25. Steven G. Louie, Yang-Hao Chan, Felipe H. da Jornada, et al., "Discovering and Understanding Materials through Computation," *Nature Materials* 20 (2021): 728–35, https://doi.org/10.1038/s41563-021-01015-1.

26. The Materials Project, "The Materials Project," accessed September 18, 2024, https://next-gen.materialsproject.org.

27. Vahe Tshitoyan, John Dagdelen, Leigh Weston, et al., "Unsupervised Word Embeddings Capture Latent Knowledge from Materials Science Literature," *Nature* 571 (July 2019): 95–98, https://doi.org/10.1038/s41586-019-1335-8.

28. Sang Cheol Kim, Solomon T. Oyakhire, Constantine Athanitis, et al., "Data-Driven Electrolyte Design for Lithium Metal Anodes," *Proceedings of the National Academy of Sciences* 120, no. 10 (February 2023): e2214357120, https://doi.org/10.1073/pnas.2214357120.

29. Nathan J. Szymanski, Bernardus Rendy, Yuxing Fei, et al., "An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials," *Nature* 624 (November 2023): 86–91, https://doi.org/10.1038/s41586-023-06734-w.

30. Josh Leeman, Yuhan Liu, Joseph Stiles, et al., "Challenges in High-Throughput Inorganic Material Prediction and Autonomous Synthesis," ChemRxiv, Preprint, submitted January 7, 2024, https://doi.org/10.26434/chemrxiv-2024-5p9j4.

31. US Food and Drug Administration, "2013 Nanotechnology Regulatory Science Research Plan," last modified March 19, 2018, https://www.fda.gov/science-research/nanotechnology-programs-fda/2013-nanotechnology-regulatory-science-research-plan.

32. White House, "Executive Order 13563–Improving Regulation and Regulatory Review," Office of the Press Secretary, January 18, 2011, https://obamawhitehouse.archives.gov/the-press-office/2011/01/18/executive-order-13563-improving-regulation-and-regulatory-review.

33. Paresh Chandra Ray, Hongtao Yu, and Peter P. Fu, "Toxicity and Environmental Risks of Nanomaterials: Challenges and Future Needs," *Journal of Environmental Science and Health, Part C* 27 (February 2009): 1–35, https://doi.org/10.1080/10590500802708267.

34. EPA Council on PFAS, "PFAS Strategic Roadmap: EPA's Commitment to Action 2021–2024," US Environmental Protection Agency, October 18, 2021, epa.gov/system/files/documents/2021-10/pfas-roadmap_final-508.pdf.

35. Chunli Bai, "Ascent of Nanoscience in China," *Science* 309, no. 5731 (July 2005): 61–63, https://doi.org/10.1126.

STANFORD EXPERT CONTRIBUTORS

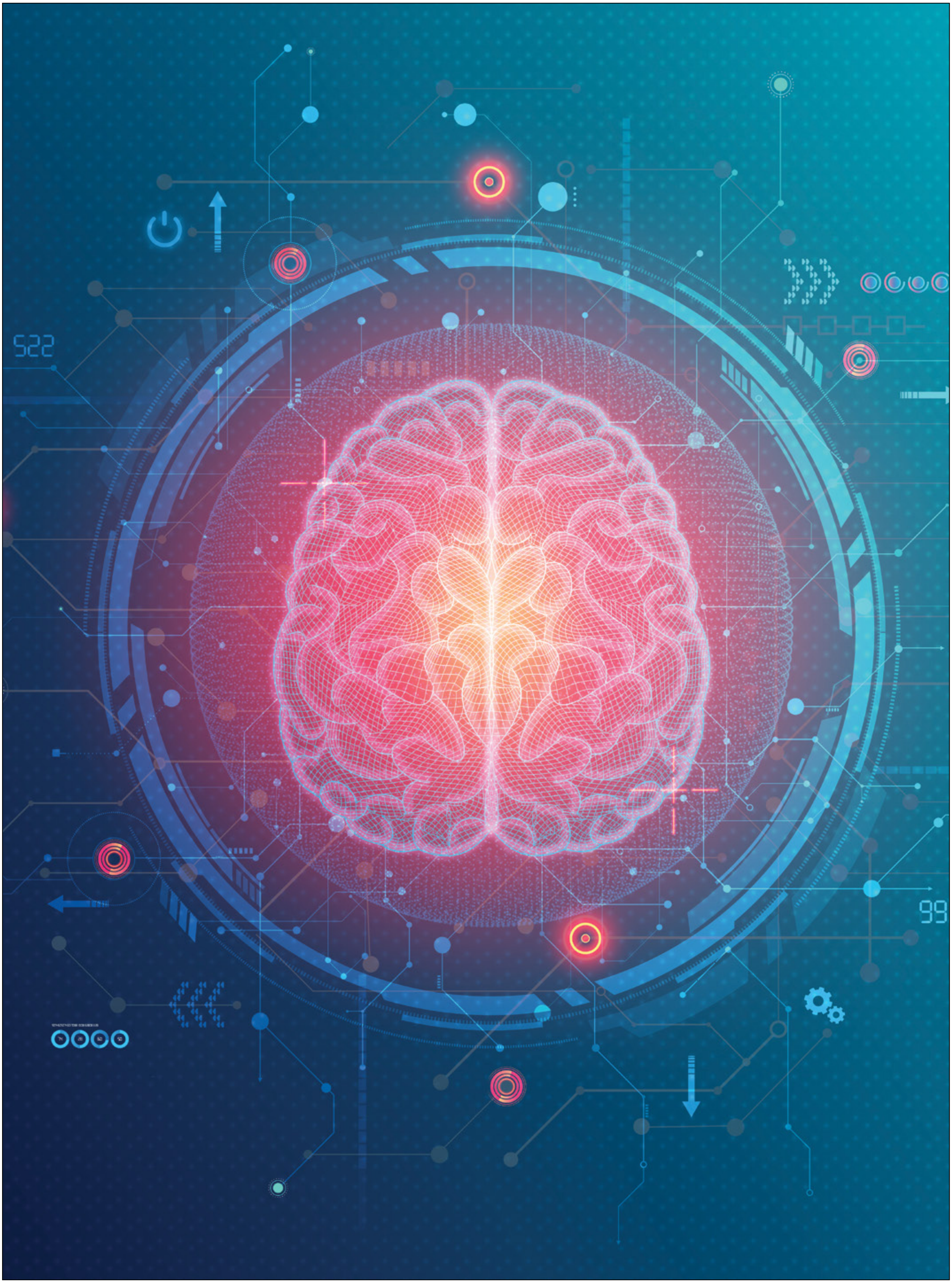
Dr. Zhenan Bao
SETR Faculty Council, K. K. Lee Professor of Chemical Engineering, and Professor, by courtesy, of Materials Science and Engineering and of Chemistry

Dr. Felipe Jornada
Assistant Professor of Materials Science and Engineering

Dr. Andrew Spakowitz
Tang Family Foundation Chair of the Department of Chemical Engineering and Professor of Chemical Engineering and of Materials Science and Engineering

Dr. Stefano Cestellos-Blanco
SETR Fellow and Postdoctoral Scholar in Chemical Engineering

Dr. Lukas Michalek
SETR Fellow and Postdoctoral Scholar in Chemical Engineering



06

NEUROSCIENCE

KEY TAKEAWAYS

- Popular interest in neuroscience vastly exceeds the actual current scientific understanding of the brain, giving rise to overhyped claims in the public domain that revolutionary advances are just around the corner.

- Advances in human genetics and experimental neuroscience, along with computing and neuroscience theory, have led to some progress in several areas, including understanding and treating addiction and neurodegenerative diseases and designing brain-machine interfaces for restoring vision.

- American leadership is essential for establishing and upholding global norms about ethics and human subjects research in neuroscience, but this leadership is slipping with decreased strategic planning and increased foreign investments in the field.

Overview

Neuroscience is a multidisciplinary field of study that focuses on the components, functions, and dysfunctions of the brain and our nervous system at every level. It reaches from the earliest stages of embryonic development to dysfunctions and degeneration later in life and from the individual molecules that shape the functions of a neuron to the study of the complex system dynamics that are our thoughts and dictate our behaviors.

The human brain consumes 20 to 25 percent of the body's energy even though it constitutes only a small percentage of a human's body weight, a fact that underscores its outsize importance.[1] The power of the human brain is what has allowed us to become the dominant species on Earth without being the fastest, strongest, or biggest.

The brain is unfathomably complex, containing approximately 86 billion neurons[2]—nerve cells that sense the physical world, transmit information to the brain, process information, and send information from the brain to other parts of the body. A single neuron can make thousands or tens of thousands of connections to other neurons. These connections are called synapses (see figure 6.1).

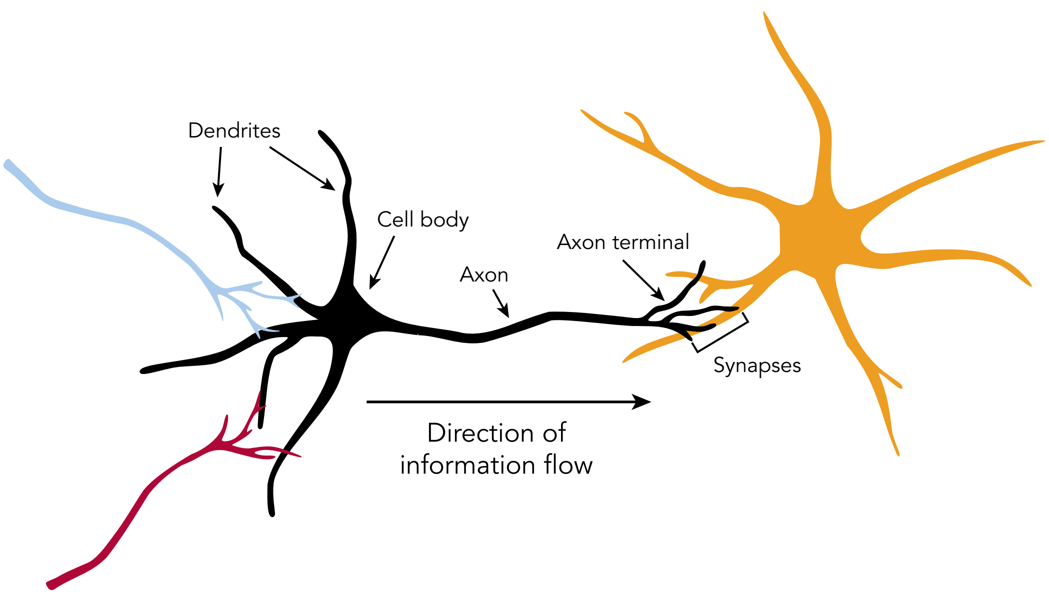
All of our consciousness and behavior, from the action of stabbing a potato with a fork to contemplating the mysteries of the universe, is underpinned by which neurons connect with one another, the neurotransmitter/receptor pairs involved, the strength of the connections, and the electrical properties of the neurons—as well as by how these various features change over time.

Neurons and synapses function in many ways that are similar to electrical circuits. Indeed, the exploration of the electrical properties of neurons came directly from the same technologies, theories, and equations developed for harnessing electricity. Many pioneering neuroscientists started as electrical engineers and physicists. Just as electrical connectors create a path for electricity to flow through a circuit, neural circuits can be defined by the parallel and recurrent connections between neurons that occur to compute a specific function, such as deciding to move a limb or identifying an object visually. Neurons can also communicate with each other using hormone-like signaling, which is relatively slow but longer lasting compared to fast-acting electric signals. These types of communications underlie mood and behavior states such as sleep/awake and hunger/satiety.

Complete understanding of what each neuron is doing at any given time is currently impossible. Even for a mouse brain, which is much simpler than a human brain, it is still a tremendous effort to characterize individual brain regions despite the availability of powerful techniques that allow us to identify activity in individual neurons or to noninvasively tag cells to respond to light signals.

Over the past several years, however, it has become clear that individual neurons are almost never

FIGURE 6.1 Structure of a neuron



responsible for any given behavior or computation; instead, they act in parallel, duplicating some functions and combining to determine thoughts and actions. This neural redundancy makes it easier to infer what is going on in the brain more broadly.

A particular brain region can be considered like a magnificent one-thousand-person choir. Just sampling 1 percent of the singers can provide a pretty good idea of the music the overall choir is producing at any given time. Researchers already have the ability to record from thousands of neurons at a time. This provides useful insight into how a brain functions, even if we don't understand in detail what the other 99 percent of its neurons are doing.

Key Developments

This chapter focuses on three research areas in neuroscience that show major promise for concrete applications: brain-machine interfaces (neuroengineering), degeneration and aging (neurohealth), and the science of addiction (neurodiscovery).

Neuroengineering and the Development of Brain-Machine Interfaces

A brain-machine interface is a device that maps neural impulses from the brain and translates these signals to computers. The potential applications for mature brain-machine interface technologies are wide-ranging: The augmentation of vision, other senses, and physical mobility; direct mind-to-computer interfacing; and computer-assisted memory recall and cognition are all within the theoretical realms of possibility. However, headlines about mind-reading chip implants are exaggerated and still more the realm of science fiction. Even with tremendous interest and rapid progress in neuroscience and engineering, the necessary theoretical understanding of how neurocircuits work is still limited to only a few areas of the brain. What's more, the technical problems of safely implanting electrodes have not been solved.

Perhaps the most encouraging example of a brain-machine interface is the recent development of an artificial retina. The retina is the part of the eye that converts light into corresponding electrical signals sent to the brain. People who have certain incurable retinal diseases are blind because the light-detecting cells in their retinas do not work. To restore sight, the Stanford artificial retina project aims to take video images and use electrodes implanted in the eye to simulate the electronic signals in a pattern that a functional retina would normally produce.[3]

The project involves recording spontaneous neural activity to identify cell types and their normal signals, understanding how electrodes activate cells, and stimulating retinal ganglion cells—which collect visual information from photoreceptors in retinas—to represent an image so that this information can be transmitted by the optic nerve to the brain. Solving these technical problems calls for deep knowledge of relevant surgical techniques as well as significant engineering know-how in multiple areas—including translating the scientific understanding of the stimulation algorithm used into practical applications, making experimental recordings, and fabricating and packaging the electrode into the device.

The artificial retina project is the most mature brain-machine interface to date in terms of its ability to "read" and "write" information. The retina, a part of the central nervous system, is well suited as an experimental environment, as its stimuli (light) is experimentally controllable and can be captured by a digital camera. It is the best-understood neural circuit and the theory of its function has developed to the point where much of retinal processing can be modeled. Compared to complex cognitive processes like learning and memory—where even the inputs aren't fully understood—the task of reconstructing vision is more achievable, albeit still challenging.

Other brain-machine interfaces are currently being developed, though they are less mature or less ambitious than the artificial retina project. Some of these decode brain activity without controlling a neural signal. For instance, one interface can translate brain activity in areas controlling motor functions into signals that can then be sent to an artificial prosthetic limb. Here, feeding high-dimensional patterns of recorded neural activity into an artificial intelligence (AI) algorithm can make it possible to control an artificial limb without requiring direct control of neural functions—a form of control that remains beyond our current scientific understanding.

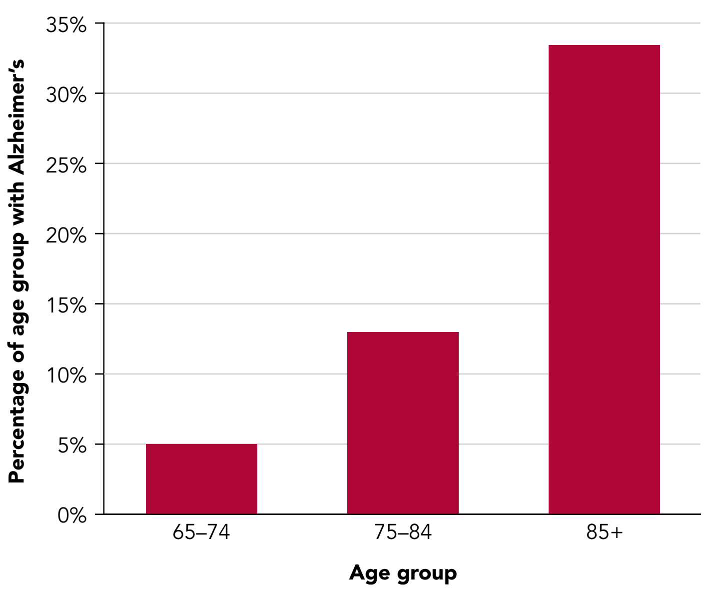
These demonstrations hint at the prospect of other brain-machine interfaces in the future, such as computer-assisted memory recall, even if the full suite of potential applications is still unclear. The scope and feasibility of these applications will be determined by advances in neuroscientific theory and in technical solutions to engineering problems such as how to safely and accurately insert probes into deep-layer tissues.

Neurohealth and Neurodegeneration

Neurodegeneration is a major challenge as humans live longer. Alzheimer's disease is of particular concern. In the United States alone, the annual cost of treating it is projected to grow from $305 billion today to $1 trillion by 2050.[4] Diseases like Alzheimer's and Parkinson's surge in frequency with age—while just 5 percent of 65- to 74-year-olds have Alzheimer's, this rises to 33 percent for those over 85 (see figure 6.2).[5] As modern medicine and society enable longer lifespans, the human body and brain remain maladapted to maintaining nervous system function for decades past childbearing age.

Alzheimer's disease is characterized by the accumulation of two different proteins—amyloid beta and tau—into toxic aggregates. Amyloid beta accumulates

FIGURE 6.2 Alzheimer's disease surges in frequency with age



Source: Data from "2023 Alzheimer's Disease Facts and Figures," *Alzheimer's & Dementia* 19, no. 4 (April 2023): 1598–1695 outside of neurons, induces cellular stress, and in turn may cause tau to build up inside the neurons. As the brain regions where tau accumulates are those most cognitively impacted, a reasonable consensus exists that tau is the more direct cause of the neural death responsible for dementia.

However, despite what is known about neurodegenerative diseases such as Alzheimer's, little progress has been made in producing effective treatments that slow disease progression. For example, tau remains harder to target therapeutically, and the recently approved drugs target amyloid beta. While these amyloid drugs are very effective in eliminating the amyloid plaques from patient brains, their effectiveness in improving patients' cognitive abilities remains questionable.

Another form of neurodegeneration results from traumatic brain injury (TBI), which can manifest itself in a range of complex symptoms and pathologies.[6] Traumatic impact to brain systems can affect cognitive and behavioral functions in ways that lead to long-term and severe psychiatric conditions requiring specialized care. This is particularly evident in the current surge of athletic and military brain injuries that exhibit predominantly psychiatric symptoms. A person's past medical and psychiatric records, as well as any coexisting conditions, play a vital role in diagnosis and treatment. TBI offers insights into other neuropsychiatric disorders and can pave the way for innovative concepts in neurodegenerative disease.

Neurodiscovery and the Science of Addiction

Researchers are working to understand the neural basis of addiction and of chronic pain while working with psychiatrists and policymakers to address the opioid epidemic.[7] Estimates of the economic costs of that epidemic range from $100 billion to $1 trillion a year when the loss of potential lifetime earnings of overdose victims is included.[8] Additional economic losses occur due to depletion of the labor force and the billions spent on the criminal justice system and healthcare related to addiction.[9] Beyond economics, there are the significant emotional costs that impact individuals experiencing addiction, as well as their families and friends. Death also takes its toll: The number of opioid deaths in the United States has risen from 21,000 in 2010 to 111,000 in 2022,[10] which places deaths from opioid overdoses on the same level as those caused by diabetes and Alzheimer's.[11] Overdose deaths fell by 3 percent in 2023 compared with the prior year,[12] but it is not clear yet whether the downtick is merely a pause in growth or a fundamental turning point.

Many of the most impactful changes for dealing with the societal problems arising from addiction come from public policy interventions and societal shifts, such as raising taxes on tobacco or changing physicians' prescribing practices for addictive substances such as opioids (see figure 6.3). Nevertheless, neuroscience has a potentially important role to play in addressing addiction. For example, a nonaddictive painkiller drug as effective as current-generation opioids could be transformative.[13]

FIGURE 6.3 Opioids prescribed by physicians

Source: iStock.com / Johnrob

Another approach is to leverage neuroscience to identify and target brain states that reinforce addiction or make it more likely. Consider the problem of relapse in tackling addiction. Scientists have found that the brain mechanisms leading to an initial opioid addiction differ significantly from those that trigger a relapse. It turns out that opioid receptors are found in neural circuits related to the desire for social interaction. Stanford neuroscientists have recently identified a circuit that is responsible for the onset of aversion to social interactions during recovery.[14] Such an aversion is a significant challenge to recovery because social interactions are often key to helping an individual cope with the vulnerabilities associated with the recovery process. The finding suggests it may be possible to develop drugs that inhibit social aversion during withdrawal, thereby assisting patients in seeking help or companionship from friends, families, recovery programs, and doctors.

Depression is also a major driver of addiction and a barrier to recovery. It involves a loss of the ability to feel good, which addictive drugs temporarily counteract by activating the brain's reward centers. However, over the long term, these drugs can also dull emotions, making normal experiences less rewarding and worsening patients' overall mood. Addiction also impairs executive control, makes normal life seem unsatisfactory, and creates a belief that drug use is essential for survival. Addressing any factor that contributes to depression-driven addiction can help facilitate the recovery process.

One nonpharmaceutical intervention for depression is Stanford neuromodulation therapy (SNT).[15] SNT employs transcranial magnetic stimulation (TMS)—the use of magnetic fields to stimulate specific brain regions—on regions involved in executive functioning and emotional regulation, particularly the left dorsolateral prefrontal cortex, which is responsible for functions such as problem-solving and self-control.[16] This approach aims to strengthen connections between brain areas to better regulate negative emotions. Initial trials have shown promising results, with nearly 80 percent of participants experiencing lasting remission.[17] SNT improves upon traditional TMS by using individualized brain scans and condensing treatment into five days. If these remission rates hold, it could represent a significant step forward in treating depression.

Over the Horizon

Progress and Prospects in Neuroscience

The pace of neuroscientific discovery is slow and limited by the biological nature and complexity of the nervous system. Year-over-year advances tend to be incremental. Researchers use simple model organisms like fruit flies with short generation times to study fundamental questions inexpensively. But the closer research gets toward human application, the more complex, time-consuming, and expensive it becomes. For instance, because neurodegeneration is a slow, progressive disease where day-to-day worsening is minimal, clinical trials often take many years.

Most of the economic impacts of neuroscience in some way connect to the healthcare industry and its search for treatments for neurodegenerative disorders (such as Alzheimer's and Parkinson's disease), neuropsychiatric disorders (addiction, depression, and schizophrenia), and neural prosthesis (brain-machine interfaces to restore limb function and speech).

It is important to keep in mind that the brain's complexity often prevents researchers from understanding fully why even effective treatments for neurological conditions actually work. For example, we know that drugs called selective serotonin reuptake inhibitors block the reabsorption of serotonin into neurons, but neuroscientists do not have a clear explanation for why this helps treat depression. New neurological therapies may work, but because we don't have a good understanding of exactly why they do so, fine-tuning and improving them often comes down to simple trial and error. Luckily for medical science, an in-depth understanding of how a particular treatment works may not always be necessary for therapeutic intervention.

ALZHEIMER'S DISEASE DETECTION AND TREATMENT

The potential for early detection prior to the onset of cognitive impairment is higher than it has ever been before. Current-generation diagnostic tools now include the ability to cheaply test for biomarkers from blood plasma paired with more accurate but expensive spinal taps and positron emission topography, or PET, scans for toxic tau and amyloid buildup. While anti-amyloid drugs are controversial for treating even mild cases of Alzheimer's because of their side effects (which include brain swelling and bleeding), a rollout of mass blood-plasma screening, along with confirmation using more expensive tests, might mean these drugs could be applied before clinical symptoms manifest, possibly increasing their effectiveness.

At this point, detection is more advanced than treatments. Antisense oligonucleotides (ASOs) are an up-and-coming class of drugs that may actively slow cognitive decline in patients already exhibiting disease symptoms. An ASO that disrupts the production of additional tau showed positive results in an early clinical trial for safety in early 2023.[18] The sample size was small, but the trial showed cognitive improvements from treatment.[19] Participants are currently being recruited for another clinical trial scheduled to stretch from 2022 to 2030.[20] While this approach suffers from the drawback of the treatment requiring spinal injections, extreme adverse events were mostly limited to side effects of the injections themselves, rather than the brain swelling or bleeding frequently observed with the anti-amyloid antibody drugs.

NEUROSCIENCE-BASED PROSTHESES

Neural redundancy has important ramifications for the development of brain-controlled prostheses. For example, if the goal were to develop an artificial limb controlled by the brain, it would be nearly impossible to monitor every neuron in the motor cortex. However, if about 80 percent of neural activity can be represented by a small group of neurons, then a single, minimally invasive probe might be sufficient to interpret movement intentions and control an artificial limb. Although the remaining 20 percent of neural activity, which the probe wouldn't capture, would likely still be important for fine-tuning limb movement, a computer could help manage these details once there's a clear understanding of how to interpret movement intentions.

Neural redundancy is also important in neural prostheses for seizure treatment. If a probe can be implanted into an area of the brain prone to seizures, then it might be possible, even without a complete sampling of the neural population, to predict the state of the relevant part of the brain and warn of an imminent seizure. Such a prediction could allow for intervening immediately to disrupt those network dynamics or informing the patient of imminent danger. It wouldn't be necessary to understand the complete set of neural computations to have a sufficiently clear signal for medical intervention.

NEUROSCIENCE AND AI

As understanding of the mathematics of our neural computations increases, these computational models may have direct relevance to AI. In particular, machine learning requires vast training datasets. By contrast, humans can learn languages with a small fraction of the training data that AI models require (for more discussion of this point, please refer to chapter 1 on artificial intelligence). Better understanding the mathematical principles that define how human brains compute may therefore improve AI. The melding of neuroscience theory and AI is a topic of increasing interest under the umbrella of Stanford's Wu Tsai Neurosciences Institute.[21]

Challenges of Innovation and Implementation

Contrasting the work on artificial retinas and the work on the science of neurodegeneration and addiction illustrates the dual-pronged nature of neuroscience applications. They have two primary components: a scientific one that focuses on identifying relevant brain circuits and understanding how these function and compute, and a technical engineering one that focuses on how to safely stimulate the relevant brain circuits to create the desired responses.

There is much about the brain's anatomy, physiology, and chemistry that is still not well understood, and addressing the theoretical issues in neuroscience is almost exclusively the purview of academia rather than industry. There are research programs in industry that solve basic biological questions in neuroscience, but these are tied to solving problems with a profit motive—usually the development of new drugs.

Once the basic science has been developed and a research area approaches an economically viable application, industry does a much better job of developing it. Consequently, helping to smooth the friction of moving a project from academia to industry is crucial to overcoming roadblocks in development. Incubators and accelerators can help transition the findings of basic research to application by aiding in high-throughput screening—the use of automated equipment to rapidly test samples—and prototyping. With viable prototypes, new companies can be created or licenses granted to existing companies to produce a final product. Such activities are critical in facilitating the integration of well-understood scientific theory, technical engineering, and final application.

Policy, Legal, and Regulatory Issues

DISCONNECT BETWEEN PUBLIC INTEREST AND CAPABILITY

The brain is perhaps the least understood, yet most important, organ in the human body. Demand for neuroscience research advances and applications—including understanding brain circuitry, developing new drugs, treating diseases and disorders, and creating brain-machine interfaces—is expected to continue to grow considerably over the coming years. The Society for Neuroscience's annual meeting draws close to thirty thousand attendees.[22]

Science fiction and fantastical headlines fuel beliefs that mind-reading technology, brains controlled by computers, and other dystopias are imminent. In reality, work to comprehend the brain's staggering complexity remains in its early stages. Most advances involve incremental progress, expanding our theoretical foundations rather than producing revolutionary leaps to futuristic applications. This vast gap between public expectations and scientific reality creates an environment ripe for exploitation. Impatience for solutions to pressing

Science fiction and fantastical headlines fuel beliefs that mind-reading technology, brains controlled by computers, and other dystopias are imminent. In reality, work to comprehend the brain's staggering complexity remains in its early stages.

medical problems like dementia and mental illness leave many open to dubious proclamations or pseudoscience.

DRUG POLICY AND NEUROSCIENCE RESEARCH

The Controlled Substances Act governs US policy regarding regulation of the manufacture, importation, possession, use, and distribution of certain substances. Substances on Schedule 1 are drugs or other substances with a high potential for abuse and not currently accepted for medical use in the United States. No research exceptions are provided for Schedule 1 substances such as cannabis or MDMA (often known as Ecstasy or Molly), which have potential for medical use that might be realized through research. In May 2024, the Biden administration proposed to reassign marijuana to Schedule 3, a schedule with fewer restrictions.[23] Placing drugs on Schedule 1 sharply constrains researchers because it becomes difficult to obtain these potentially helpful substances for study. This constraint also denies the public the benefits that might flow from such research—such as better medical treatments—and potentially harms the public if, for example, individual states choose to legalize certain drugs without adequate research into their safety, addictiveness, and public health impacts.

THE IMPACT OF COGNITIVE AND BEHAVIORAL NEUROSCIENCE ON LAW

Cognitive and behavioral neuroscience, which studies the biological basis of thoughts and actions, has broad implications for public policy. For example, a basic aspect of criminal law is the nature and extent of an individual's responsibility for a criminal act. Under a 2005 US Supreme Court ruling, minors under eighteen years of age cannot be subject to the death penalty for crimes they committed because adolescent brains are not fully developed, putting minors at higher risk of impulsive, irrational thoughts and behaviors.[24]

THOUGHT IMPLANTS

The possibility that information can be implanted directly into a person's consciousness is an interesting future research problem as the nature of brain-machine interfaces becomes more ambitious over the coming decades. As government is still figuring out how to regulate internet forums that influence what people believe and how they feel—a problem that has existed for three decades—regulation will likely not come fast enough to guide even the later-stage promises made about brain-machine interfaces. Establishing proper cultural norms at the outset and careful consideration of technologies is warranted.

FUNDING CUTS TO TRANSFORMATIVE NEUROSCIENCE

Over the past decade, much of the work outlined in this chapter was funded by the Brain Research Through Advancing Innovative Neurotechnologies (BRAIN) Initiative. Starting in 2014, this aimed to be the equivalent of the Human Genome Project for the human brain. Research from the BRAIN Initiative has helped neuroscience generate advances that specifically aid in translating science to medicine. In 2024, however, the initiative's budget was cut by 40 percent, from $680 million to $402 million. The decline was due to a combination of reduced funding from the National Institutes of Health (NIH) and through the 21st Century Cures Act. Funding through that act is expected to fall by an additional $81 million in 2025.[25] Without additional financial support through the NIH, neuroscience research in the United States and the country's ability to tackle some of the most societally impactful diseases will decline.

FOREIGN COLLABORATION

Human expertise will continue to be the primary driver of future advances in neuroscience, and success will continue to depend on the United States being the best place for international scientists to train, conduct research, and use their own expertise to teach the next generation of scientists. Against this backdrop, the apparent targeting of US scientists with personal and professional links to China raises concerns,[26] and the United States only loses if these scientists leave and move their labs to China.

ETHICAL FRAMEWORKS

Neuroscience research naturally raises many ethical concerns that merit careful, ongoing discussion and monitoring. Chief among these is research on human subjects, which is governed by several existing frameworks and regulations that guide neuroscience studies in American academia today. Ethical guidelines for scientific research are usually national, not international. Some countries might allow particular types of brain research and drugs, while others might not; for example, a nation might permit experimentation on prisoners or on ethnic minorities. Managing differences in state research regimes will be critical to harnessing the power of international collaboration.

NOTES

1. Marcus E. Raichle and Debra A. Gusnard, "Appraising the Brain's Energy Budget," *Proceedings of the National Academy of Sciences* 99, no. 16 (July 2002): 10237–39.

2. Frederico A. C. Azevedo, Ludmila R. B. Carvalho, Lea T. Grinberg, et al., "Equal Numbers of Neuronal and Nonneuronal Cells Make the Human Brain an Isometrically Scaled-up Primate Brain," *Journal of Comparative Neurology* 513, no. 5 (April 2009): 532–41, https://doi.org/10.1002/cne.21974.

3. Stanford Medicine, "The Stanford Artificial Retina Project," accessed August 30, 2023, https://med.stanford.edu/artificial-retina.html.

4. Winston Wong, "Economic Burden of Alzheimer Disease and Managed Care Considerations," *American Journal of Managed Care* 26, no. 8 (August 2020): S177–83, https://doi.org/10.37765/ajmc.2020.88482.

5. Alzheimer's Association, "2023 Alzheimer's Disease Facts and Figures," March 14, 2023, https://doi.org/10.1002/alz.13016; A. W. Willis, E. Roberts, J. C. Beck, et al., "Incidence of Parkinson's Disease in North America," *npj Parkinson's Disease* 8, no. 170 (December 2022), https://doi.org/10.1038/s41531-022-00410-y.

6. Vassilis E. Koliatsos and Vani Rao, "The Behavioral Neuroscience of Traumatic Brain Injury," *Psychiatric Clinics of North America* 43, no. 2 (June 2020): 305–30, https://doi.org/10.1016/j.psc.2020.02.009.

7. Wu Tsai Neurosciences Institute, "NeuroChoice Initiative (Phase 2)," Stanford University, accessed August 30, 2023, https://neuroscience.stanford.edu/research/funded-research/neurochoice.

8. Low end: Pew Charitable Trust, "The High Price of the Opioid Crisis, 2021," August 27, 2021, https://www.pewtrusts.org/en/research-and-analysis/data-visualizations/2021/the-high-price-of-the-opioid-crisis-2021; high end: Feijin Luo, Mengyao Li, and Curtis Florence, "State-Level Economic Costs of Opioid Use Disorder and Fatal Opioid Overdose—United States, 2017," *Morbidity and Mortality Weekly Report* 70, no. 15 (April 16, 2021): 541–46, http://dx.doi.org/10.15585/mmwr.mm7015a1.

9. Office of Justice Programs, "National Drug Threat Assessment, 2011," US Department of Justice, August 2011, https://www.ojp.gov/ncjrs/virtual-library/abstracts/national-drug-threat-assessment-2011.

10. National Institute on Drug Abuse, "Drug Overdose Deaths: Facts and Figures," National Institutes of Health, August 2024, https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates.

11. Overdose included in the accidental death statistic in Centers for Disease Control and Prevention, "Leading Causes of Death," last modified May 2, 2024, https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm.

12. National Center for Health Statistics, "U.S. Overdose Deaths Decrease in 2023, First Time Since 2018," US Centers for Disease Control and Prevention, May 15, 2024, https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2024/20240515.htm.

13. US Food and Drug Administration, "FDA Takes Steps Aimed at Fostering Development of Non-Addictive Alternatives to Opioids for Acute Pain Management," February 9, 2022, https://www.fda.gov/news-events/press-announcements/fda-takes-steps-aimed-fostering-development-non-addictive-alternatives-opioids-acute-pain-management.

14. Gordy Slack, "Social Aversion During Opioid Withdrawal Reflects Blocked Serotonin Cues, Mouse Study Finds," Wu Tsai Neurosciences Institute, Stanford University, November 2, 2022, https://neuroscience.stanford.edu/news/social-aversion-during-opioid-withdrawal-reflects-blocked-serotonin-cues-mouse-study-finds.

15 For more on this subject, see Nina Bai, "Researchers Treat Depression by Reversing Brain Signals Traveling the Wrong Way," News Center, Stanford Medicine, May 15, 2023, https://med.stanford.edu/news/all-news/2023/05/depression-reverse-brain-signals.html.

16. Wayne C. Drevets, Jonathan Savitz, and Michael Trimble, "The Subgenual Anterior Cingulate Cortex in Mood Disorders," *CNS Spectrums* 13, no. 8 (August 2008): 663–81, https://doi.org/10.1017/s1092852900013754.

17. Aditya Somani and Sujita Kumar Kar, "Efficacy of Repetitive Transcranial Magentic Stimulation in Treatment-Resistant Depression: The Evidence Thus Far," *General Psychiatry* 32, no. 4 (August 2019): e100074, https://doi.org/10.1136/gpsych-2019-100074.

18. Catherine J. Mummery, Anne Börjesson-Hanson, Daniel J. Blackburn, et al., "Tau-Targeting Antisense Oligonucleotide $MAPT_{RX}$ in Mild Alzheimer's Disease: A Phase 1b, Randomized, Placebo-Controlled Trial," *Nature Medicine* 29 (April 2023): 1437–47, https://doi.org/10.1038/s41591-023-02326-3.

19. Isabella Ciccone, "Biogen's BIIB080 Exhibits Favorable Outcomes in Phase 1b Trial for Early Alzheimer's Disease," Neurology Live, October 26, 2023, https://www.neurologylive.com/view/biogen-biib080-exhibits-favorable-outcomes-phase-1b-trial-early-ad.

20. Alzheimers.gov, "BIIB080 for Mild Cognitive Impairment Due to Ahlzeimer's Disease or Mild Alzheimer's," National Institute on Aging, National Institutes of Health, accessed March 14, 2023, https://www.alzheimers.gov/clinical-trials/biib080-mild-cognitive-impairment-due-alzheimers-disease-or-mild-alzheimers.

21. Wu Tsai Neurosciences Institute, "Center for Mind, Brain, Computation, and Technology," Stanford University, accessed October 14,

2024, https://neuroscience.stanford.edu/initiatives-centers/center-mind-brain-computation-and-technology.

22. Society for Neuroscience, "Attendance Statistics: Meeting Attendance," accessed August 30, 2023, https://www.sfn.org/meetings/attendance-statistics.

23. Federal Register, "Schedules of Controlled Substances: Rescheduling of Marijuana," May 21, 2024, https://www.federalregister.gov/documents/2024/05/21/2024-11137/schedules-of-controlled-substances-rescheduling-of-marijuana.

24. Roper v. Simmons, 543 U.S. 551 (2005).

25. Jocelyn Kaiser, "Major Budget Cuts to Two High-Profile NIH Efforts Leave Researchers Reeling," ScienceInsider, *Science*, April 30, 2024, https://www.science.org/content/article/major-budget-cuts-two-high-profile-nih-programs-leave-researchers-reeling.

26. Jeffrey Mervis, "Pall of Suspicion: The National Institutes of Health's 'China Initiative' Has Upended Hundreds of Lives and Destroyed Scores of Academic Careers," *Science*, March 23, 2023, https://www.science.org/content/article/pall-suspicion-nihs-secretive-china-initiative-destroyed-scores-academic-careers.

STANFORD EXPERT CONTRIBUTORS

Dr. Kang Shen
SETR Faculty Council, Frank Lee and Carol Hall Professor, and Professor of Biology and of Pathology

Dr. Keith Humphreys
Esther Ting Memorial Professor in the Department of Psychiatry and Behavioral Sciences

Dr. Paul Nuyujukian
Assistant Professor of Bioengineering, of Neurosurgery, and, by courtesy, of Electrical Engineering

Dr. Michael Greicius
Iqbal Farrukh and Asad Jamal Professor and Professor, by courtesy, of Psychiatry and Behavioral Sciences

Dr. Alec Condon
SETR Fellow and Postdoctoral Scholar in Biology and at the Howard Hughes Medical Institute

07

ROBOTICS

KEY TAKEAWAYS

- Future robots may be useful for improving the US manufacturing base, reducing supply chain vulnerabilities, delivering eldercare, enhancing food production, tackling the housing shortage, improving energy sustainability, and performing almost any task involving physical presence.

- Progress in artificial intelligence holds the potential to advance robotics significantly but also raises ethical concerns that are essential to address, including the privacy of data used to train robots, data bias that could lead to physical harm by robots, and other safety issues.

- Achieving the full potential of robots will require a major push from the federal government and the private sector to improve robotics adoption and research across the nation.

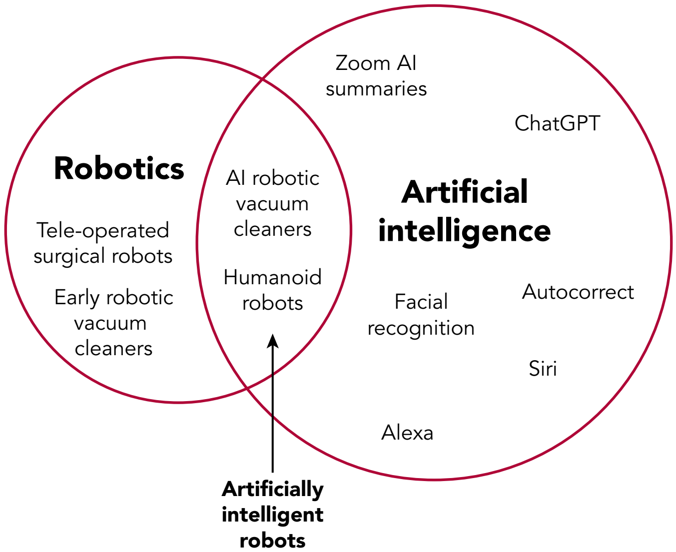
Overview

Researchers do not universally agree on the definition of a robot, but a consensus seems to have emerged that, at the very least, a robot is a human-made physical entity with ways of sensing itself or the world around it and ways of creating physical effects on that world.[1]

Robots, which include both static entities such as fixed robotic arms on production lines and mobile entities such as drones, must integrate many different component technologies to combine perception of the environment with action in it. Perception requires generating a representation of the robot's environment and its interaction with its surroundings. Action requires the robot to make physical changes to itself or the environment based on those perceptions.

The key engineering challenges in robotics involve the design of components and integration of them

FIGURE 7.1 Not all robots use artificial intelligence



within a robot's body so it can perform intended tasks in different settings in a given environment. Different types of robots operate in different environments—including factories, homes, and even space—and each environment poses distinct complexities beyond just a robot's technical performance. For example, working alongside humans raises critical issues of safety and liability.

Important component technologies include:

- Actuators that enable movement, such as motors and grasping appendages
- Sensors that receive real-time input about the immediate physical environment of the robot and the robot's own configuration
- Control systems that decide what the robot should do based on sensor readings
- Materials that robots are made of. Those built from rigid materials typically interact with their operating environments in highly prescribed and structured ways. "Soft" robots that are flexible and conform to the environment can offer better performance in more unstructured and chaotic environments.[2]
- Power sources that can be tethered to a robot or that are untethered. A robot tethered to a "mother ship" can be energized from a power source on that ship indefinitely, while untethered robots need self-contained power sources or sources that harvest energy from the environment.
- Real-time computing that determines the specific timeframes in which operations of robots take place. This ensures, for example, that a robotic arm in a workplace will stop very fast if the robot detects a human in its immediate proximity.

Finally, some robots use computer vision and other types of artificial intelligence (AI) for understanding their environments and decision-making, but robotics and AI do not always go together (see figure 7.1). Robots with varying degrees of autonomy have been used in everything from delicate surgical procedures to space exploration.

Key Developments

Academic and commercial robotics activities are influenced by global socioeconomic trends as well as by technology developments in other domains. Here we briefly review some of the most important current influences on the robotics field.

Manufacturing

The manufacturing sector is a major contributor to the US economy. In 2023, it accounted for just over a tenth of GDP, adding $2.87 trillion.[3] It is also a major contributor to employment, with an estimated thirteen million people working in manufacturing.[4] However, the sector is facing several challenges that robotics can help overcome.

One of the most important of these challenges is a shortage of skilled labor. Approximately two-thirds of the respondents in the National Association of Manufacturers' outlook survey, conducted in the first quarter of 2024, highlighted attracting and retaining skilled workers as their primary challenge.[5] Several reasons have led to this situation, including more people retiring and a declining population growth rate. Unless solutions are found, millions of manufacturing jobs could remain unfilled in the future, impacting America's prosperity and national security.

Another major challenge is the vulnerability of the US supply chain, a result of the trend toward outsourcing manufacturing to countries with lower labor costs. The globalization of manufacturing has helped reduce the price and increase the variety of goods available to US consumers, but it has left US supply chains more vulnerable to disruptions—as we saw during the COVID-19 pandemic with the sudden shortage of motor vehicles and consumer electronics, to name just two industries. These and other disruptions underline the importance of bringing more manufacturing back to the United States and making manufacturing supply chains more resilient across the world.

Robots have the potential to reduce vulnerabilities in these supply chains in many ways. New innovations such as robotic graspers that can handle even very fragile goods can make current manufacturing lines more adaptable and reconfigurable. The development and increased deployment of collaborative robots, or cobots, that can interact with human workers could help alleviate labor shortages—though human workers' reactions to these robots and safety issues associated with them are significant concerns that need to be addressed. Robots also hold promise for manufacturing in extreme environments, such as in space or underwater, and researchers are working on issues like precision control and visual perception to help them operate in these domains.

The "Now" Economy

Near-real-time delivery of goods and services, also known as the "now" economy, refers to the deployment of goods and services as close to customers as possible so that products are available very rapidly upon demand. Another dimension involves the remote delivery of services, including medical treatments.

These areas pose multiple challenges, including finding better ways to deliver goods to customers quickly and cost-effectively; applying expertise at the point of need through teleoperation, even if the experts are geographically distant; and finding ways to automate services to offset labor shortages.

Robotics is already providing solutions to address these needs. Various kinds of robots, including drones and multiwheeled delivery vehicles that can navigate sidewalks, are being trialed to conduct last-mile deliveries (the final step in the shipping process when goods are transported from the distribution hub to the customer). Remote robot-assisted surgery for gallbladder procedures and hernia repair is increasingly available, a trend covered in more detail later in this chapter. Robots can automate basic services around the home, such as floor cleaning. They can also serve as seasonal, on-demand labor in some industries such as agriculture, where autonomous robots can be used to pick fruit and perform other tasks in the harvesting season.

While the potential for further innovation is clear, there are also many challenges, both technical and nontechnical, to address. These include the following:

- Ensuring that the use of robots in last-mile delivery strategies does not lead to safety and privacy violations. For instance, increasing use of delivery drones has led to a number of reported shootings of drones, an activity with obvious safety implications.[6] A delivery drone could also inadvertently capture images of people in their homes and backyards.

- Developing adaptable and reliable manipulation capabilities that enable autonomous robots to handle a wide variety of goods and other objects for delivery to customers

- Creating more networking infrastructure that can support wider adoption of reliable remote robotic applications in healthcare and other areas

Food Production

The world's population increased from seven billion to eight billion between 2010 and 2022, and the United Nations estimates it will rise by a further two billion people by 2030.[7] To keep up with demand, food production is predicted to increase by 50 percent by 2050. With the increased frequency of dramatic weather events such as floods and droughts, agriculture is becoming increasingly challenging. Farmers will need to more rapidly adapt the types of crops they use—and their seeding and harvesting practices—to keep up with demand.

Robotics can support these efforts and help streamline the production and processing of food. Currently, robots are deployed mainly to reduce the cost of specific processes such as milking, seeding fields, and picking fruit. The main hurdle to expanding the use of robotics in agriculture and food production is the dexterity needed to accomplish certain complex tasks. Meat production, for example, is a heavily subsidized market with very little automation.[8] Although some companies such as Cargill and Tyson Foods have recently invested heavily in developing automated solutions,[9] much can still be done to improve the quality and efficiency of the meat-carving process.

Robotics alone cannot address the goal of significantly increasing food production. Integrating it with other technologies, such as AI and computer vision, is critical for agriculture and food processing. Combining these technologies can help increase food production yields in two main ways:

Direct interface with robots With the use of AI—and, specifically, the implementation of reinforcement learning, an approach that mimics the trial-and-error learning process of humans—robots can be trained to accomplish complex tasks in simulated environments before they are deployed and to keep learning from their mistakes once they are in place.

Data capture For example, by using computer vision, a seeding robot can also keep track of the health of a crop, check the level of ripeness of fruit, and generally provide farmers with a wealth of information they would otherwise not have access to or that would require a significant effort to collect.

Advancing automation in agriculture can increase the efficiency and productivity of the industry, helping it meet growing demand. The additional data acquired via robots will also help drive and improve land-management practices. Strategies such as well-timed crop rotation can improve soil fertility, and changing seed types can boost yields dramatically. Implementing precision agriculture, which uses sensors to collect data and algorithms to analyze the information, can help develop automated systems that track and report key metrics for farmers. This strategy can also improve the environment by decreasing the use of fertilizers and irrigation in agriculture.

Over the Horizon

Impact of Robotic Technologies

SUPPORTING AN AGING POPULATION

Robotic technologies can assist in the support and care of the elderly. Although the US population is not aging as fast as those of some other countries such as Japan and Italy, more than a fifth of people in America will be over sixty-five by 2030.[10] Americans are also living longer: Average life expectancy in the United States has risen to over seventy-seven years, according to the Centers for Disease Control and Prevention, up from less than seventy-four years in 1980.[11] Sadly, the prevalence of cognitive impairment increases with longer lifespans, and medical interventions for this issue are the focus of much neuroscience research (see chapter 6 on neuroscience).

There is a huge dearth of qualified personnel for eldercare. Long hours, low wages, and the intensity of the tasks involved have made the industry unattractive for many prospective employees. More restrictive immigration policies have exacerbated the issue, as a large fraction of eldercare workers are immigrants, and the demand for eldercare workers is growing.[12]

Against this backdrop, assistive and rehabilitative robots are being developed and deployed to support human caregivers. These robots can be electronic companions that help people with basic tasks associated with the activities of daily living both inside and outside their homes. Assistive robots can also take the form of exoskeletons. These are wearable robotic devices that provide support with movement by, for instance, working with calf muscles to give people extra propulsion with each step taken (see figure 7.2).[13] Smaller devices like trackers can also help monitor symptoms, recognize falls and alert health professionals, and detect early signs of cardiac issues.

FIGURE 7.2 A wearable exoskeleton

Source: Shutterstock / Unai Huizi Photography

Another application of robotics in healthcare is in surgery. Around 30 percent of the surgeries necessary worldwide every year go unperformed.[14] Using robots, like the one shown in figure 7.3, to automate all or parts of routine procedures can streamline surgical interventions, making them safer and more efficient. Developments in force sensor feedback and haptics—the science of simulating pressure, vibrations, and other sensations related to touch—are

FIGURE 7.3 Robotic surgical systems can make surgical interventions safer and more efficient

Source: © 2024 Intuitive Surgical Operations, Inc.

driving the development of surgical robots that can be controlled from remote locations, making it unnecessary for a doctor to be physically present to either diagnose or treat a patient. Telerobotics is particularly promising for rural areas, where it can be harder to get physical access to specialists. Robotic surgery focused on low- and middle-income countries may help to improve access to surgical care in those countries.[15]

The main challenge here is the complexity of the tasks involved. In a routine surgery, organ location, shape, and size can vary dramatically among patients, and robots must adjust for this. Even a seemingly easy task like feeding a patient can be difficult for a robot because small movements of the individual can be hard to adjust for. AI and machine learning (ML) are being talked about as potential solutions to such issues, but for every new task a robot has to learn, an immense amount of training data is required to ensure it will function safely.

As a parallel, before self-driving cars are allowed on the road, they must complete millions of miles of journeys to test their capabilities.[16] How can we achieve a comparable level of experience in healthcare? One solution is to use simulation so that human participants are not endangered in testing. However, there is still concern that even state-of-the-art simulation may not fully capture enough scenarios and behaviors to ensure a robot can interact safely with patients.

TACKLING THE HOUSING SHORTAGE

High prices and low supply have exacerbated a crisis in housing in the United States. Moreover, while demand for housing-construction workers continues to grow, the number of workers available is not rising enough to match it. For example, one study notes that even before the pandemic, construction firms ranked labor shortages as the biggest hurdle for their businesses, and 78 percent of them were experiencing difficulty in filling their job positions.[17] Robots could potentially address this shortage, increase overall construction productivity, and also reduce worker injuries and deaths over time.

Currently, there are commercial robots that are capable of bricklaying, house framing, wall finishing, and moving heavy items on construction sites. The SAM100 robot (SAM stands for semi-automated mason), developed by Construction Robotics in the United States, can assist with brick retrieval, mortar placement, and brick placement in a wall. It can help complete this process three to five times faster than a bricklayer working alone,[18] but it still needs human involvement for certain tasks such as smoothing out excess mortar. Another construction robot, the Hadrian X, developed by FBR in Australia, still requires manual loading of the bricks it handles. Future innovation will drive toward increased automation of such processes.

In addition to construction, robotics could contribute to other aspects of building and maintaining infrastructure. Paving, grading, striping and marking, and repairing roads could be either fully or partially automated. Robots equipped with cameras and sensors can conduct these tasks with greater accuracy and more precise control than humans, and they can be fully automated or operated via remote control. Currently, road inspection relies on surveyors to be on-site, but robots could be used to inspect and repair roads and other transportation infrastructure, especially in areas not easily accessible to humans.

There are still various challenges to address when integrating robotics with housing construction. Unpredictable changes, such as moving machinery or heavy equipment on work sites, can pose safety hazards for both personnel and robots that the latter must be better equipped to deal with. Training robots to respond appropriately to such scenarios using AI and other methods will depend on sufficient availability of high-quality data with which to train them. Human workers also need to be trained to interact safely and efficiently with robotic systems on construction sites and to maintain these systems.

ADVANCING SUSTAINABILITY

Robotics can contribute significantly to sustainable practices, addressing pressing global challenges while opening new avenues for industrial applications. These applications span renewable energy development and harvesting, sustainable agricultural practices, and infrastructure maintenance.

In the renewable energy sector, robots are indispensable in the construction, maintenance, and management of solar and wind farms. They assist in tasks ranging from the design and installation of infrastructure to ongoing upkeep, such as cleaning solar panels and maintaining wind turbines. For instance, BladeBUG has developed six-legged robots that scale turbines, conduct detailed inspections, and make minor repairs, reducing downtime and maintenance costs for wind farms.[19] Robots can facilitate the development of other renewable energy sources like geothermal and wave energy by automating labor-intensive and hazardous tasks.

They are also helping to gather resources in a more sustainable way. Robots can harvest materials, such as lumber, with minimal ecological impact and collect metallic nodules from the ocean floor without damaging marine ecosystems. In recycling and waste management, advanced robotic systems can handle hazardous materials and complex products, protecting workers and streamlining processing. For example, robotics company AMP has created AI-powered robots that pick out recyclable materials from complex waste streams with high precision.[20]

In agriculture, robots help to optimize water usage, reduce chemical inputs, and minimize runoff. John Deere's See & Spray uses advanced camera detection and ML to apply herbicide to crops only where needed with high precision, thereby minimizing chemical usage. Such technologies can enhance harvesting efficiency, detect the spread of diseases in crops, and make it easier to adapt agricultural practices quickly to the changing climate.

In infrastructure maintenance, robots can monitor and repair aging facilities, preventing failures that could harm people and business operations. For example, Boston Dynamics' Spot robot has been used to inspect industrial sites, monitor construction progress, and detect gas leaks.

This and the other examples cited above highlight the potential for robots to play an important role in advancing sustainability goals, but significant investment in technology development and infrastructure is still needed. For instance, ensuring the reliability and efficiency of robotic systems in diverse and often harsh and unpredictable environments is crucial. Creating safe and efficient human–machine interactions in work such as waste management and infrastructure maintenance will also be vital.

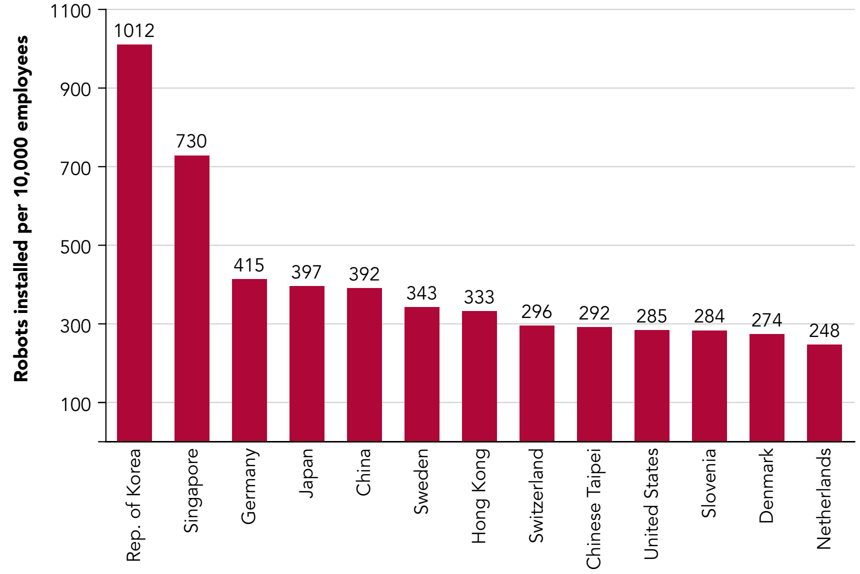
Challenges of Innovation and Implementation

ADOPTION AND FUNDING

Achieving the full potential of robots to help drive economic growth and address pressing supply chain and labor constraints will require a major push from the federal government and the private sector to improve robotics adoption and research across the nation.

In terms of adoption, manufacturing-robot density in the United States in 2022 was 285 robots per 10,000 employees, ranking the country tenth in the world behind nations such as Singapore, Germany, South Korea, Japan, and China (see figure 7.4).[21] This reflects the fact that US businesses—especially small and midsize ones—are still adopting robots more slowly than those in some other large economies. Addressing regulatory and standard-setting issues more comprehensively could help. For instance, developing comprehensive guidelines and policies

FIGURE 7.4 The United States lags behind many other countries in manufacturing-robot adoption



Source: Adapted from International Federation of Robotics, "Global Robotics Race: Korea, Singapore and Germany in the Lead," January 10, 2024

Achieving the full potential of robots to help drive economic growth and address pressing supply chain and labor constraints will require a major push from the federal government and the private sector to improve robotics adoption and research across the nation.

to ensure the safe and effective use of robots in construction is critical.

Furthermore, the adoption of robotics in this and other areas has implications for workforce dynamics. As robots become more capable of taking on dull, dirty, and dangerous tasks, enabling human workers to focus on more complex and rewarding activities, there is a clear need to manage the transition carefully to prevent significant job displacement. Investing in education and training programs is essential to prepare the workforce for new roles and ensure that the benefits of robotics are widely shared.

Creating clear standards for—or at least achieving a rough consensus on—use of robotic technologies across industries and regions could help accelerate more widespread use of robots. This is important in all of the areas touched on in this chapter but especially in relatively new domains such as the use of robots for advancing sustainable practices.

Finally, funding of robotics research and development remains an issue. Despite efforts such as the National Robotics Initiative, which was launched during the Obama administration, more support will be needed if the United States is to make the most of the exciting and transformative opportunities that robotics offers.

Policy, Legal, and Regulatory Issues

Robots can improve manufacturing efficiency, bolster sustainability, and provide treatment to the elderly, but their adoption raises legal and ethical questions that are essential to address. There is much talk about the benefits of AI, for instance, but the implications of the ongoing data collection required to train robots for complex tasks and ensure their safe interaction with the world are significant. This section categorizes the legal and ethical considerations associated with robotics that warrant further policy development into three broad areas: privacy and consent, inclusion and integrity, and safety.

PRIVACY AND CONSENT

Anyone using social media or browsing the internet has encountered warnings and consent requests for things like advertisements and tracking cookies. News reports are filled with cybersecurity breaches that involve unauthorized access to huge amounts of personal data. How, then, should we be thinking about handling the vast amount of personal data accumulated by robots in our hospitals, homes, and elsewhere?

Access to health-related information is heavily monitored and regulated to protect patient privacy. It is important that policies for safeguarding personal data be developed in tandem with current efforts to collect very large amounts of data for training robotic systems, as well as during the deployment of these systems. With the exponential increase of data production, harvesting, and use, it is also crucial to ensure data is held securely. In the wrong hands, health and other personal information can be used to coerce and control individuals, disrupt the smooth and secure conduct of daily activities, and undermine trust in essential systems.

INCLUSION AND INTEGRITY

In addition to concerns about consent and privacy in data acquisition, inclusion and integrity are also crucial issues—and ones that concern AI more broadly. The launch of products that make use of ML, like facial recognition systems, has revealed that bias in datasets can lead to unintended and sometimes harmful outcomes. This issue is going to be vitally important for robotics. For example, what if bias embedded in algorithms leads a surgical robot to be less well trained to operate on women than men? Or what if a robotic safety system scanning people infers that someone might be carrying a gun because of their ethnicity? The consequences could be grave. Promoting and enforcing norms and standards when it comes to robot-training datasets is essential to ensure that the diversity of America's population is adequately reflected and that robots are used safely.

SAFETY

The deployment of robots in society raises ethical questions regarding both physical safety and cybersecurity. As noted earlier, physical safety pertains not only to how well a robot is trained to perform a specific task, but also to how it manages unforeseen disruptions.

From the standpoint of public acceptability, one critical question is the nature of the standards that are appropriate for society to use for judging the safety of a robot in any given application. A threshold question is whether its safety performance must simply be comparable to—or slightly better than—that of an average human or whether it should be judged against a higher standard, perhaps one of near perfection. From a practical standpoint, the former is a reasonable standard, and widespread deployment of robots that met such a standard would improve safety on average. But whether the public is willing to accept this standard is unproven.

Setting standards for the performance of robotic control systems is crucial for ensuring their successful and continued adoption. Similarly, cybersecurity standards for robotics need to be on a par with those in the domains that robots are used in, including healthcare and national security.

NOTES

1. Ralph Lässig, Markus Lorenz, Emmanuel Sissimatos, et al., "Robotics Outlook 2030: How Intelligence and Mobility Will Shape the Future," Boston Consulting Group, June 28, 2021, https://www.bcg.com/publications/2021/how-intelligence-and-mobility-will-shape-the-future-of-the-robotics-industry.

2. Some examples of soft robots can be found at Alberto Paitoni Faustinoni, "Soft Robotics: Examples, Research and Applications," Robotics24, February 3, 2023, https://robotics24.net/blog/soft-robotics-examples-research-and-applications/.

3. National Association of Manufacturers, "Manufacturing in the United States," accessed September 15, 2024, https://nam.org/manufacturing-in-the-united-states/.

4. National Association of Manufacturers, "Manufacturing in the United States."

5. National Association of Manufacturers, "2024 First Quarter Manufacturers' Outlook Survey," March 5, 2024, https://nam.org/2024-first-quarter-manufacturers-outlook-survey.

6. Sasha Rogelberg, "Walmart's Latest Challenge with Its Drone Delivery System Is Gun Owners Shooting Packages Out of the Air," *Fortune*, July 3, 2024, https://fortune.com/2024/07/03/walmart-drones-gun-owners-delivery-florida-droneup/.

7. United Nations, "Global Issues: Population," accessed September 16, 2024, https://www.un.org/en/global-issues/population#.

8. David Gillette and Warren Barge, "The True Cost of a Hamburger," American Institute for Economic Research, April 20, 2022, https://www.aier.org/article/the-true-cost-of-a-hamburger/.

9. Simon Creasey, "How the Meat Industry Is Embracing Automation," Just Food, October 26, 2022, https://www.just-food.com/features/how-the-meat-industry-is-embracing-automation/.

10. America Counts Staff, "By 2030, All Baby Boomers Will Be Age 65 or Older," US Census Bureau, December 10, 2019, https://www.census.gov/library/stories/2019/12/by-2030-all-baby-boomers-will-be-age-65-or-older.html.

11. Shameek Rakshit, Matthew McGough, and Krutika Amin, "How Does U.S. Life Expectancy Compare to Other Countries?," Health

System Tracker, January 30, 2024, https://www.healthsystemtracker.org/chart-collection/u-s-life-expectancy-compare-countries/.

12. Robert Espinoza, Lina Stepick, Jessica King, et al., "Bridging the Gap: Enhancing Support for Immigrant Direct Care Workers and Meeting Long-Term Care Needs," PHI, September 26, 2023, https://www.phinational.org/resource/bridging-the-gap-enhancing-support-for-immigrant-direct-care-workers/.

13. Patrick Slade, Mykel J. Kochenderfer, Scott L. Delp, et al., "Personalizing Exoskeleton Assistance While Walking in the Real World," *Nature* 610 (October 2022): 277–82, https://doi.org/10.1038/s41586-022-05191-1.

14. One study estimates that about 143 million additional surgical procedures are needed in low- and middle-income countries each year to save lives and prevent disability, while 313 million such procedures are performed each year. See John G. Meara, Andrew J. M. Leather, Lars Hagander, et al., "Global Surgery 2030: Evidence and Solutions for Achieving Health, Welfare, and Economic Development," *The Lancet* 386, no. 9993 (August 2015): 569–624, https://doi.org/10.1016/S0140-6736(15)60160-X.

15. Aashna Mehta, Jyi Cheng Ng, Wireko Andrew Awuah, et al., "Embracing Robot Surgery in Low- and Middle-Income Countries: Potential Benefits, Challenges, and Scope in the Future," *Annals of Medicine and Surgery* 84 (December 2022): 104803, https://doi.org/10.1016/j.amsu.2022.104803.

16. Waymo was authorized to begin offering fully driverless rides in San Francisco in June 2024. Waymo had logged over 7.14 million fully autonomous miles driven 24/7 through the end of October 2023. See Waymo, "Waymo Significantly Outperforms Comparable Human Benchmarks Over 7+ Million Miles of Rider-Only Driving," *Waypoint* (blog), December 23, 2023, https://waymo.com/blog/2023/12/waymo-significantly-outperforms-comparable-human-benchmarks-over-7-million/.

17. Troup Howard, Mengqi Wang, and Dayin Zhang, "How Do Labor Shortages Affect Residential Construction and Housing Affordability?," Haas School of Business, University of California Berkeley, April 2023, https://www.haas.berkeley.edu/wp-content/uploads/Howard_Wang_Zhang_Housing-Supply-and-Construction-Labor-Dayin-Zhang.pdf.

18. Construction Robotics, "SAM," accessed September 16, 2024, https://www.construction-robotics.com/sam-2/.

19. Bladebug, "Home," accessed September 16, 2024, https://bladebug.co.uk/.

20. Contrary Research, "AMP Robotics," last modified October 5, 2023, https://research.contrary.com/company/amp-robotics.

21. Assembly, "US Ranks 10th in Robot Density," January 17, 2024, https://www.assemblymag.com/articles/98270-us-ranks-10th-in-robot-density.

STANFORD EXPERT CONTRIBUTORS

Dr. Allison Okamura
SETR Faculty Council and Richard W. Weiland Professor in the School of Engineering

Dr. Steve Collins
Associate Professor of Mechanical Engineering and, by courtesy, of Bioengineering

Dr. Mykel Kochenderfer
Associate Professor of Aeronautics and Astronautics and, by courtesy, of Computer Science

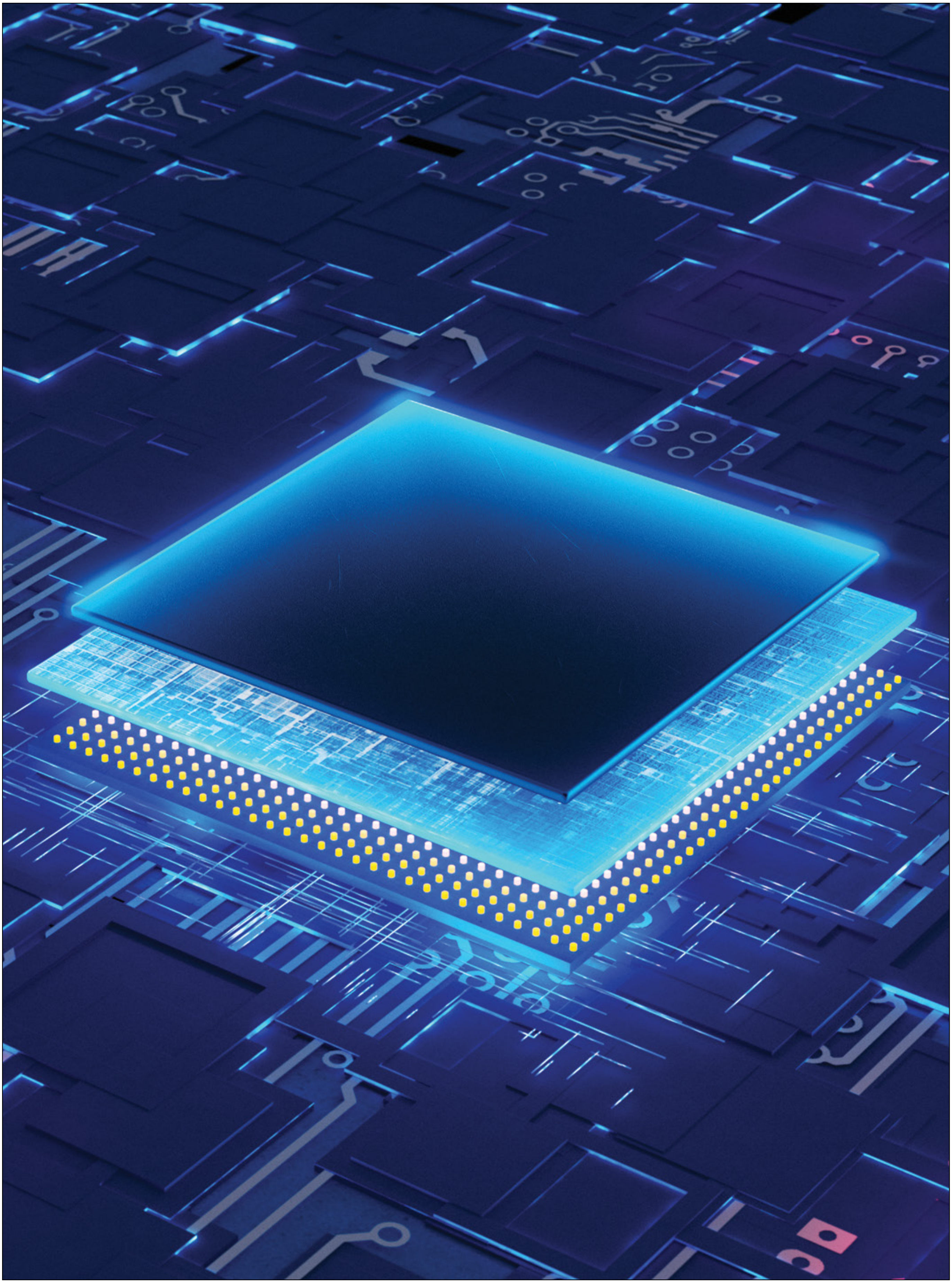
Dr. Jiajun Wu
Assistant Professor of Computer Science and, by courtesy, of Psychology

Dr. Karen Liu
Professor of Computer Science

Dr. Cosima du Pasquier
SETR Fellow and Postdoctoral Scholar in Mechanical Engineering

Dr. Alaa Eldin Abdelaal
SETR Fellow and Postdoctoral Scholar in Mechanical Engineering

Brian Vuong
SETR Fellow and PhD Student in Mechanical Engineering



08

SEMICONDUCTORS

KEY TAKEAWAYS

- The growing demand for artificial intelligence and machine learning is driving innovations in chip fabrication that are essential for enhancing computational power and managing energy efficiency.
- Advances in memory technologies and high-bandwidth interconnects, including photonic links, are critical for meeting the increasing data needs of modern applications.
- Even if quantum computing advancements are realized, the United States will still need comprehensive innovation across the technology stack to continue to scale the power of information technology.

Overview

Semiconductors, often in the form of microchips, are crucial components used in everyday physical devices, from smartphones and toasters to cars and lawn mowers. Chips also control heating and cooling systems, elevators, and fire alarms in modern buildings. Traffic lights are controlled by chips. On farms, tractors and irrigation systems are controlled by chips. Modern militaries could not function without chips in their weapons, navigation devices, and cockpit life-support systems in fighter jets. The list goes on and on—in every aspect of modern life, chips are essential.

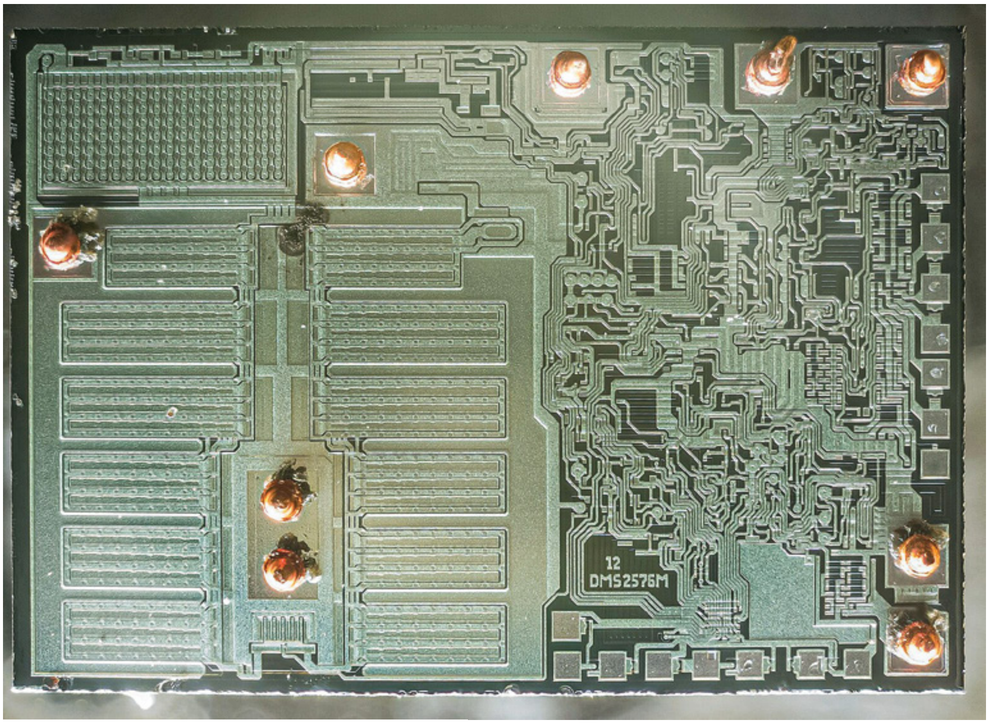
All chips are involved in the handling of information. Different types of them are specialized for different tasks. Some are processor chips that ingest data, perform computations on the data, and output the results of those computations. Memory chips store information and are used with processors. Still other chips act as interfaces between digital computations and the physical world. In all these cases, some amount of energy is needed to represent each bit of information inside a chip. The magic of chips is that it takes several orders of magnitude less energy to represent information inside one than it takes to do so outside it (e.g., in wires leading to and from the chip). This means that, in a multi-chip system, much more energy and chip space are required for data moving between chips than for data that remains on a chip; this is one of the driving forces to integrate more functions on single chips.

As chip fabrication technologies improve, it takes less energy and chip space to represent a given bit of information; hence, processing those bits becomes more energy efficient. This phenomenon is what has enabled the semiconductor industry to pack more processing power on chips over time—it enables designers to create chips that do more complex processing (see figure 8.1), although the cost of designing them also increases with their complexity.

Recently, however, the energy costs associated with the hardware that holds information on a chip have been falling more slowly, and the cost of manufacturing per unit area has increased. This means that the cost and energy advantages of scaling have nearly stopped. As a result, researchers have been investigating other ways to improve computer technology and to deal with the problem of high design costs.

Since the best technologies for performing different chip functions are themselves different, systems still need to use different chips for those functions. Finding new ways to manage the inefficiency of

FIGURE 8.1 Increased energy efficiency has allowed designers to create more complex chips



Source: Wikimedia Commons. Used under CC BY-SA 4

FIGURE 8.2 Chip fabrication requires large factories that can produce chips at scale

Source: SkyWater Technology

information movement in and among chips, along with the issue of high design costs, is a central focus of research on semiconductors. Further improvement will take the form of innovation in design, materials, and integration methods.

Two aspects of chips are important for the purposes of this report. They must be designed and then fabricated (i.e., manufactured), and each function calls for different skill sets. Chip design is primarily an intellectual task that requires tools and teams able to create and test systems containing billions of components. Fabrication is primarily a physical effort that requires large factories, or fab facilities, that can produce chips by the millions and billions—and can cost billions of dollars and take several years to build from the ground up (see figure 8.2).

Fabrication integrates many complex processes to produce chips. Each one requires substantial expertise to master and operate, and the integration of all of them requires still further expertise. For these reasons, modern fabrication plants are operated by workforces with a substantial number of people trained in engineering.

Fabrication also entails a significant degree of process engineering to continue to improve process technology and to achieve stringent manufacturing standards. For example, the "clean rooms" in which chips are made require air that is one thousand times more particle-free than the air in a hospital operating room.[1]

Because chip design and chip fabrication are so different in character, only a few companies, such as Intel, do both. Many businesses specialize in design, including Qualcomm, Broadcom, Apple, and Nvidia. Such companies are called "fabless" in recognition of the fact that they do the design work and outsource fabrication to others—a strategy based on the theory that the former activity has higher profit margins compared to the latter.

Today, "others" usually means one company: Taiwan Semiconductor Manufacturing Company (TSMC), which is by far the world's largest contract chip-manufacturing company. In 2024, TSMC controlled over 60 percent of the world's contract semiconductor manufacturing and 90 percent of the world's advanced chip manufacturing.[2] Samsung, in South

Korea, is a distant second, with around 13 percent of the world's chip manufacturing.[3] United Microelectronics Corporation, also based in Taiwan, ranks third at about 6 percent.[4]

By contrast, US chip-manufacturing capacity has lost significant ground. Fabrication plants in America accounted for 37 percent of global production in 1990, but their share dropped to just 12 percent by 2021.[5] Industry concentration, low US production capacity, and geopolitical concerns about China's intentions toward Taiwan mean the global supply chain for chips will remain fragile for the foreseeable future, despite the passage of the Creating Helpful Incentives to Produce Semiconductors (CHIPS) and Science Act of 2022 (discussed in more detail later in this chapter).

Key Developments

Moore's Law, Past and Future

For over half a century, information technology has been driven by improvements in the chip fabrication process. In 1965, Intel cofounder Gordon Moore observed that the cost of fabricating a transistor was dropping exponentially with time—an observation that has come to be known as Moore's law. It's not a law of physics but rather a statement about the optimal rate at which economic value can be extracted from improvements in the chip fabrication process.

Although Moore's law is often stated as the number of transistors on a chip doubling every few years, historically the cost of making a chip was mostly independent of the components on it. This has meant that every few years, a chip whose size and cost remains approximately the same will have twice the number of transistors on it.

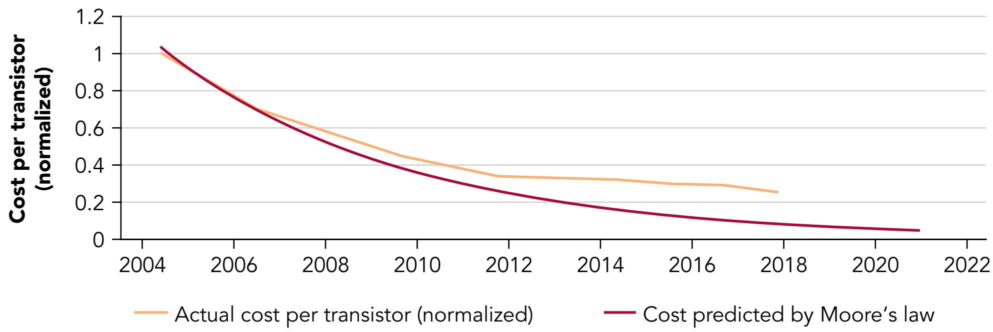
Moore's law scaling (i.e., the exponential increasing of the number of transistors on a chip) meant that each year one could build last year's devices for less money than before or build a more powerful system for the same cost. This scaling has been so consistent that it is widely believed that the cost of computing will always decrease with time. This expectation is so pervasive that in almost all fields of work, people are developing more complex algorithms to achieve better results while relying on Moore's law to rescue them from the consequences of that additional complexity.

But the future will not look like the past. As the complexity of chips increases, the traditional cost-benefit relationships associated with Moore's law scaling are diminishing, leading to rising costs in chip manufacturing. As figure 8.3 shows, the actual cost of a chip per transistor (represented by the solid red line) was tracking the cost predicted by Moore's law (the dashed blue line) relatively well from 2004 to 2012.[6] However, the actual cost per transistor started to level off around 2012—and it has not kept up with Moore's law predictions since then.[7]

Against this backdrop, the past year has witnessed significant advancements and challenges in the semiconductor industry. For example, increasing demand for computing power driven by artificial intelligence (AI) and machine learning (ML) applications (as discussed in chapter 1 on artificial intelligence) has led to a surge in the development of, and demand for, advanced graphical processing units (GPUs), creating a strain on both production and energy resources. Advanced GPU systems, such as Nvidia's GB200 NVL72 system, are pushing the boundaries of what is possible in terms of performance and power consumption. This surge in demand underscores the importance of finding innovative solutions to meet computational needs without compromising on energy efficiency or physical space.

The rapid expansion of AI applications is also driving demand for more specialized hardware. Traditional processors are no longer sufficient for the intensive computational tasks required by modern AI algorithms. As a result, there has been a significant investment in developing GPUs and other specialized

FIGURE 8.3 Cost per transistor over time



Source: Adapted from Steve Mollenkopf, "Our Future Is Mobile: Accelerating Innovation After Moore's Law," presentation at the Electronics Resurgence Initiative Summit, Detroit, MI, July 15–17, 2019

processors designed specifically for AI workloads. This shift is reshaping the semiconductor industry, emphasizing the need for high-performance, energy-efficient computing solutions.

Chiplets and 2.5-D Integration

Chiplets and 2.5-D (2.5-dimensional) integration represent a significant shift from traditional monolithic chip design. Chiplets are smaller functional blocks of silicon that can be combined to create complete systems, offering greater flexibility and customization. By enabling the use of different manufacturing technologies for various components, they facilitate the integration of high-density memory with high-performance processing units, resulting in improved system bandwidth and performance.

Central to 2.5-D integration, which puts chiplets next to one another in an integrated circuit, is the interposer—a specialized substrate facilitating communication between the chiplets. This technology enables more energy-efficient data transfer compared to traditional circuit board wiring. By allowing optimized memory, compute, and communication chips to reside side by side, it enhances system performance while reducing the need for full integration on a single chip.

From an economic perspective, the chiplet approach offers significant advantages. Semiconductor companies can create a set of building-block chiplets that can be combined in various ways, allowing for a wide range of products with different performance characteristics. This strategy allows companies to better tailor their products to specific application domains and more effectively monetize their silicon investments, ultimately leading to increased product diversity and market responsiveness.

Given the changing economics of scaling, the use of chiplets reduces costs overall and allows for more customized solutions. Semiconductor company AMD's approach of keeping components that transfer data to and from devices in older technology nodes while advancing core computing resources with the latest processes exemplifies this strategy. Moreover, this modular strategy facilitates the integration of emerging technologies such as photonics (discussed in greater detail later in this chapter), which can significantly enhance communication speeds and bandwidth within and between chips.

Three-Dimensional Heterogeneous Integration

The research community is working on an even more advanced technique: three-dimensional (3-D) heterogeneous integration. Unlike 2.5-D integration, where different chiplets are placed on a common substrate, 3-D heterogeneous integration is a semiconductor-manufacturing technique that involves the vertical stacking of different electronic components, such as processors and memory. The heterogeneous aspect means that these stacked components can be made from different materials and technologies optimized for their specific functions. For example, a processor made from one type of material can be stacked with memory made from another, each using the most suitable technology for its purpose. This approach has the potential to improve performance and efficiency by reducing the distance data travels between components, making devices faster and more compact—albeit at the cost of a more complex fabrication process and a harder heat-dissipation task for the resulting chips.

Need for High Bandwidth

A second approach to enhancing performance focuses on improving bandwidth and interconnect technologies. AI-training models must handle vast amounts of data, and high-speed interconnects such as those employed in Nvidia's H100 systems play a critical role in facilitating the rapid transfer of data between compute units and memory.

The physical limitations of traditional electrical interconnects are one of the primary barriers to improving bandwidth. As data rates increase, so do power consumption and heat generation, which can limit the performance and reliability of semiconductor devices. For example, Nvidia's GPU modules achieve bandwidths on the order of several terabytes per second. The company's NVLink enables the connection of up to 256 GPUs, forming a supercomputer pod with immense computational power and bandwidth. Each GPU in these setups is approximately ten times more powerful than a consumer one, dissipating a kilowatt of power each. This results in a pod that dissipates 10 kilowatts (kW), and a full structure of 256 GPUs dissipates over a third of a megawatt. (For comparison, a 2,500-square-foot house might require a furnace that generates about 20 kW of heat.) Thus, thermal management stands out as a critical issue. Effective thermal-management solutions, such as advanced cooling techniques and materials with high thermal conductivity, will be essential to maintaining performance and reliability in high-performance computing systems.

Photonic Links and Components

It is becoming difficult to scale electrical data-transmission links and their associated bandwidth. Innovations such as silicon photonics are emerging as potential solutions, offering the promise of higher bandwidth and lower-power consumption by using light to transmit data. Photonics are the optical analogue of electronics—the latter use electrons for signaling and carrying information while photonics use photons (light) for the same purposes.

Compared to traditional electrical interconnects, photonic links have the potential to reduce energy consumption and increase bandwidth in data centers and long-distance data transmissions.[8] Furthermore, they can handle different wavelengths on a single optical fiber simultaneously, thereby enhancing data-carrying capacity and making photonics an attractive solution for high-performance computing and data center applications. By replacing electrical interconnects with optical ones, data centers can reduce the amount of energy required for transmitting data, leading to lower operational costs and a smaller environmental footprint. These advantages of photonics have always been drivers of research in this area, but the recent rise in the demand for power-hungry AI-enabled applications has created even more impetus for such research.

Integrating photonic components with silicon-based technologies is challenging as a result of material incompatibilities; for example, efficient light-emitting materials like III–V semiconductors do not integrate well with silicon. (A III–V semiconductor is made by combining boron, aluminum, gallium, or indium with nitrogen, phosphorus, arsenic, or antimony.) While useful for light detectors, silicon is ineffective for light emission, complicating the scalable integration of these technologies at the chip or circuit board level. Overcoming these challenges is crucial for realizing the full potential of photonic links in large-scale, low-energy applications.

Memory Technology Developments

Memory technology continues to evolve, with innovations in both stacking and new materials. Techniques such as stacking multiple layers of flash memory are pushing the boundaries of what is possible, enabling higher density and better performance. These advancements are crucial for supporting the growing data needs of modern applications, from AI to big data analytics.

Dynamic random-access memory (DRAM) and flash memory technologies have both seen significant advancements in recent years, but the associated increase in manufacturing cost means there has been only modest improvement in cost per bit. The development of 3-D structures, such as vertical DRAM transistors, has allowed for continued scaling of memory density by overcoming the physical limitations of traditional planar structures and has enabled the production of memory devices with higher capacity and improved performance.

As memory technologies scale, maintaining performance and reliability becomes increasingly challenging. For DRAM, issues such as leakage currents and quantum effects limit the scalability of capacitors and transistors. To address these challenges, researchers have developed advanced manufacturing techniques for the creation of complex 3-D structures that enable increased storage density while maintaining the required electrical characteristics.

Similarly, NAND flash memory—the most common type of such memory—has faced scaling challenges due to the limitations of traditional planar cell architectures. The development of 3-D NAND, which involves stacking multiple layers of memory cells vertically, has enabled continued increases in storage density. However, this approach requires sophisticated manufacturing processes to ensure the reliability and performance of the resulting memory devices.

Emerging memory technologies, such as magnetoresistive random-access memory[9] and phase-change memory,[10] are also gaining traction in some applications. These technologies offer unique advantages in terms of speed, endurance, and energy efficiency, making them attractive alternatives to traditional embedded nonvolatile memory solutions. (Nonvolatile memory retains its contents even when power is turned off.)

Further innovations in memory technology are critical for enabling the continued growth of data-intensive applications. From AI-training models to cloud computing and big data analytics, modern applications require vast amounts of memory to store and process data efficiently.

Quantum Computing Advancements

Quantum computing remains a field of intense research and development, with significant progress made in both the number and quality of quantum bits, or qubits. Recent innovations in error correction and the potential for practical quantum computing could revolutionize specific applications,[11] although commercial viability remains years away. The promise of quantum computing lies in its potential to perform certain complex calculations at unprecedented speeds, with possible relevance for applications such as cryptography, materials science, and complex system simulations.

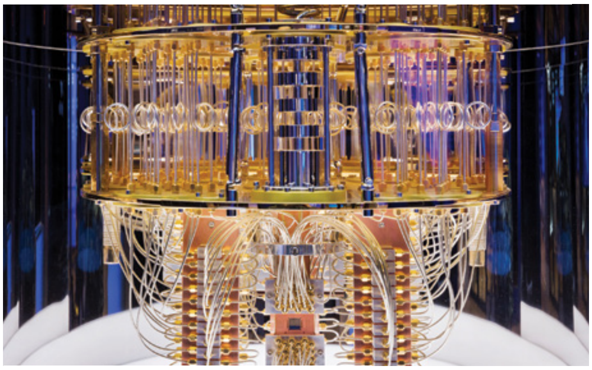
Quantum computing offers a fundamentally different approach to computation compared to classical computing. Classical computers use individual bits as the smallest unit of data, with each being 0 or 1. In contrast, the qubits in quantum computers, like the one shown in figure 8.4, can be in multiple states simultaneously due to the principles of quantum mechanics. This property, known as superposition, allows quantum computers to process a vast number of possibilities at once, a phenomenon called quantum parallelism.

However, quantum computing also presents significant challenges. Using it to implement an application is hard because reading the result of a quantum computation generally yields a result corresponding to only one of the many results possible when such a computation is performed on a superposition state. Realizing the advantage of a quantum computation means reading the correct result, which in turn requires designing computational algorithms so that the correct result is the one most likely to show up. Moreover, quantum computing operations are analog rather than digital in nature and thus can be disrupted by "noise" in the environment, such as vibrations and fluctuations in temperature.

Recent advancements in algorithms and error-correcting codes that can compensate for such errors have reduced the overhead associated with error correction. Still, many further advancements in error correction will be necessary before quantum computing can be more widely applied. Creating a useful quantum computer necessitates innovation across the entire quantum software and hardware stack: algorithms, compilers, control electronics, error correction, and quantum hardware.

Various technologies are under consideration for the physical construction of qubits, including trapped ion, superconducting, cold atom, photonic, crystal defect, quantum dot, and topological technologies.[12] The most advanced quantum computing machines currently use trapped ion or superconducting qubits. However, neither technology has a clear path to scaling up to larger machines, prompting ongoing research into other approaches.

FIGURE 8.4 A close-up view of a quantum computer



Source: IBM. Used under CC BY-ND 2.0

Recent work has improved the fidelity of a modest number (around thirty) of qubits and has reduced the overhead of quantum error correction. But far larger numbers of high-quality qubits (two or three orders of magnitude more) will be needed for quantum computers to become more broadly useful. A large number of companies and research labs have been pushing forward with work on this area. Progress on the quantum algorithm front is harder to track. While many groups are working on finding practical applications for early quantum computers, no such applications have yet been publicized.

Over the Horizon

The Impact of Technology

The semiconductor industry is poised for significant advancements in coming years, driven by the growing demands of AI, especially ML, and high-performance computing. The introduction of new technologies, such as 2.5-D integration, chiplets, and photonic interconnects, is expected to play a crucial role in meeting these demands. These innovations will enhance performance, increase bandwidth, and improve energy efficiency, addressing the limitations of traditional semiconductor designs.

Quantum computing remains an important area of development, with progress in error correction and qubit quality being made. However, its timeline remains uncertain. Even if it becomes successful, it will likely be useful for only a limited class of applications and won't replace today's semiconductor technology. Quantum computers will complement, rather than supplant, classical semiconductors, addressing specific complex problems in fields such as cryptography, materials science, and complex simulations.

Emerging memory technologies and advanced manufacturing techniques are also critical for the industry's growth. Innovations in 3-D memory stacking and integration with processors will improve data-transfer speeds and reduce latency, meeting the increasing data requirements of modern applications. The development of advanced materials and transistor architectures will further push the boundaries of semiconductor capabilities, enabling continued miniaturization and enhanced performance.

Finally, as Moore's law reaches its limits, future improvements in computing will rely more on optimizing algorithms, hardware, and technologies for specific applications rather than on general technology scaling. This requires innovation across the entire technology stack, from materials to design methods. However, the industry faces a paradox: The need for radical innovation conflicts with the high costs and long timelines of chip development, which can exceed $100 million and take over two years.

To address this, the industry must make system-design exploration easier, cheaper, and faster. Researchers are working to ensure that specific design changes to a chip do not require redesign of the entire chip. Solutions include enabling software designers to test custom accelerators without deep hardware knowledge and developing tools for application developers to make small hardware extensions to base platforms. This approach, described in more detail in the inaugural *Stanford Emerging Technology Review* (2023), depends on the involvement of major technology firms, who would need to participate in an app store–like model for hardware customization, balancing open innovation with profit motives.

CHALLENGES OF INNOVATION AND IMPLEMENTATION

A critical challenge facing the US semiconductor industry is its significant talent shortage, particularly in hardware design and manufacturing. For example, the Semiconductor Industry Association projects the number of jobs in the sector in the United States will grow by nearly 115,000 by 2030, to a total of approximately 460,000.[13] Moreover, it estimates that roughly 67,000, or 58 percent, of these new jobs risk going unfilled at current degree-completion rates. Looking at just the new jobs that are technical

A critical challenge facing the US semiconductor industry is its significant talent shortage, particularly in hardware design and manufacturing.

in nature, the percentage at risk of going unfilled is higher, at 80 percent. Almost two-thirds of the unfilled jobs would require at least a bachelor's degree in engineering.[14]

The pipeline of college graduates interested in semiconductors is also troubling. Student interest in hardware has diminished as graduating students flock to software companies.[15] Several factors appear to play a role, including the perception of higher salaries in software development and a lack of awareness about the diverse and exciting opportunities in hardware.

Since appropriately trained people are the only real source of new ideas, this trend does not bode well for the industry. Addressing this issue requires more and even closer collaboration among educational institutions, industry, and government to develop programs that attract and train the next generation of semiconductor engineers and researchers.

Policy, Legal, and Regulatory Issues

Supply chain resilience Building a resilient and diversified supply chain is critical for mitigating geopolitical risks. Investing in domestic manufacturing capabilities and promoting regional cooperation will enhance supply chain security and ensure a steady supply of essential components.

Geopolitical risks The extreme concentration of semiconductor manufacturing in Taiwan poses a significant risk to the global supply chain. Political tensions, trade disputes, and potential conflict in the region can disrupt the supply of critical components, impacting industries and economies worldwide. Diversifying supply chains and investing in domestic manufacturing capabilities are essential strategies for building resilience against geopolitical risks. Initial steps toward this were taken in the passage of the CHIPS and Science Act of 2022, which earmarked $52.7 billion for semiconductor manufacturing, research, and workforce development, plus significant tax credits for private investment in the field. Full implementation of the act has not yet occurred, partly because not enough time has elapsed and partly because the appropriations it called for have not been fully funded.

NOTES

1. Intel Corporation, "Inside an Intel Chip Fab: One of the Cleanest Conference Rooms on Earth," March 28, 2018, https://www.intc.com/news-events/press-releases/detail/165/inside-an-intel-chip-fab-one-of-the-cleanest-conference.

2. Jeremy Bowman, "This 1 Number May Ensure TSMC's Market Dominance," The Motley Fool, August 17, 2024, https://www.nasdaq.com/articles/1-number-may-ensure-tsmcs-market-dominance.

3. Counterpoint, "Global Semiconductor Foundry Revenue Share: Q1 2024," June 12, 2024, https://www.counterpointresearch.com/insights/global-semiconductor-foundry-market-share/.

4. Counterpoint, "Global Semiconductor Foundry Revenue Share."

5. Semiconductor Industry Association, *2021 State of the U.S. Semiconductor Industry*, 2021, https://www.semiconductors.org/wp-content/uploads/2021/09/2021-SIA-State-of-the-Industry-Report.pdf.

6. Electronics Resurgence Initiative 2.0, "2019 Summit Agenda, Videos, and Slides," Defense Advanced Research Projects Agency, accessed August 30, 2023, https://eri-summit.darpa.mil/2019-archive-keynote-slides.

7. In fact, transistor costs are usually plotted in a semi-log graph, where the log of the cost is plotted against time. In these plots the exponential decline in transistor costs becomes a straight line. The fact that the scale of the *y*-axis is linear is a clear indication that Moore's law is over.

8. David A. B. Miller, "Attojoule Optoelectronics for Low-Energy Information Processing and Communications," *Journal of Lightwave Technology* 35, no. 4 (February 2017): 34696, https://doi.org/10.1109/JLT.2017.2647779.

9. Paolo Cappaletti and Jon Slaughter, "Embedded Memory Solutions: Charge Storage Based, Resistive and Magnetic," in Andrea Redaelli and Fabio Pellizzer, eds., *Semiconductor Memories and Systems* (Cambridge, MA: Woodhead Publishing, 2022): 159–215, https://doi.org/10.1016/B978-0-12-820758-1.00007-8.

10. Cappaletti and Slaughter, "Embedded Memory Solutions."

11. Ziqian Li, Tanay Roy, David Rodríguez Pérez, et al., "Autonomous Error Correction of a Single Logical Qubit Using Two Transmons," *Nature Communications* 15, no. 1681 (February 2024), https://doi.org/10.1038/s41467-024-45858-z.

12. An overview of these technologies can be found in Eunmi Chae, Joonhee Choi, and Junki Kim, "An Elementary Review on Basic Principles and Developments of Qubits for Quantum Computing," *Nano Convergence* 11, no. 11 (March 2024), https://doi.org/10.1186/s40580-024-00418-5.

13. Dan Martin and Dan Rosso, "Chipping Away: Assessing and Addressing the Labor Market Gap Facing the U.S. Semiconductor Industry," Semiconductor Industry Association and Oxford

Economics, July 25, 2023, https://www.semiconductors.org/chipping-away-assessing-and-addressing-the-labor-market-gap-facing-the-u-s-semiconductor-industry/.

14. Martin and Rosso, "Chipping Away."

15. Tom Dillinger, "A Crisis in Engineering Education—Where Are the Microelectronic Engineers?," SemiWiki, July 3, 2022, https://semiwiki.com/events/314964-a-crisis-in-engineering-education-where-are-the-microelectronics-engineers.

STANFORD EXPERT CONTRIBUTORS

Dr. Mark A. Horowitz
SETR Faculty Council, Fortinet Founders Chair of the Department of Electrical Engineering, Yahoo! Founders Professor in the School of Engineering, and Professor of Computer Science

Dr. David Miller
W. M. Keck Foundation Professor of Electrical Engineering and Professor, by courtesy, of Applied Physics

Dr. David Schuster
Professor of Applied Physics

Dr. Asir Intisar Khan
SETR Fellow and Visiting Postdoctoral Scholar in Electrical Engineering

09

SPACE

KEY TAKEAWAYS

- A burgeoning "NewSpace" economy driven by private innovation and investment is transforming space launch, vehicles, communications, and key space actors in a domain that has until now been dominated by superpower governments.
- Space is a finite planetary resource. Because of dramatic increases in satellites, debris, and geopolitical space competition, new technologies and new international policy frameworks will be needed to prevent and manage international conflict in space and ensure responsible stewardship of this global commons.
- A race to establish a permanent human presence on the Moon is underway, with serious concerns that, despite Outer Space Treaty prohibitions against it, the first nation to reach the Moon may be in a strong position to prevent others from establishing their own lunar presences.

Overview

Sputnik 1 was the world's first artificial satellite, a technology demonstration placed into orbit by the Soviet Union in 1957. Sixty-eight years later, humankind operates many thousands of satellites to provide communications, navigation, and Earth observation imagery that are relied upon in many walks of life. A substantial amount of scientific discovery is also made possible with space-borne instrumentation. Additionally, space operations support military forces on Earth, and thus space itself is a domain in which international conflict and competition play out.

Today, the global space economy is growing at an average of 9 percent per annum.[1] Valued at $630 billion in 2023, it is forecast to potentially reach $1.8 trillion by 2035. This growth is driven by space-based technologies and their impacts on various industries, including defense, transportation, and consumer goods. It is supported by a shift in ownership of space assets from government to private providers.

Private-sector investment in space reached an all-time high of $70 billion in 2021–22, and commercial innovation continues to enhance the accessibility and diversity of space. The number of satellites launched per year has grown at a cumulative annual rate of over 50 percent from 2019 to 2023, supported by an increase in global rocket launches (223 in 2023).[2]

At its core, a space mission includes four components:

- The mission objectives, which can be scientific, commercial, military, or a combination thereof
- A space segment, which includes the spacecraft and the orbits that have been selected to accomplish the mission objectives
- A ground segment, which includes the rocket launcher, ground stations, and mission control centers
- A user segment, which includes all the users and stakeholders of the space mission

Space systems can be categorized in various ways. One is by whether they are crewed or uncrewed. The International Space Station (ISS) is currently the nexus for spaceflight; since 2011, US-crewed access to the ISS has been via rockets operated by Russia—and more recently through vehicles provided by SpaceX and Boeing. In the future, the NASA-operated Artemis program plans to launch its first crewed mission, a Moon flyby, in late 2025, followed by a Moon landing in 2026–27.

Uncrewed systems include those for Earth and planetary remote sensing (such as Planet Labs' Dove satellites); communication and navigation (such as the United States' Global Positioning System, or GPS, satellites); astronomy and astrophysics (such as the James Webb Space Telescope); space logistics and in-space assembly and manufacturing (such as Northrop Grumman's Mission Extension Vehicles, or MEVs); and planetary exploration (such as the Mars Perseverance rover).

Space systems can be characterized by size. Very large structures include the ISS, whose mass is

FIGURE 9.1 Fires and damage at Antonov Airport in Ukraine, as seen from a commercial satellite constellation

Source: © 2022 Maxar Technologies

420 tons, and proposed future space stations such as Blue Origin's Orbital Reef. Vastly smaller satellites, called smallsats, weigh under 500 kilograms.[3] CubeSats are the most popular smallsat format, with each CubeSat unit measuring 10 by 10 by 10 centimeters and weighing a couple of pounds. They can also be combined to build larger satellites. CubeSats support a growing commercial market, providing communications, Earth imagery, and other capabilities. Today, a large majority of functional satellites weigh between 100 and 1,000 kilograms.

Space systems can be characterized by their trajectories in space. For example, objects in orbit around Earth can be in low Earth orbit (LEO), which is less than 1,000 kilometers (km) in altitude; medium Earth orbit (MEO), which is between 1,000 and 35,000 km in altitude; high elliptical orbit (HEO); and geosynchronous orbit (GEO), with an orbit period equivalent to one Earth day. The image in figure 9.1 was obtained by a Maxar commercial satellite in LEO.

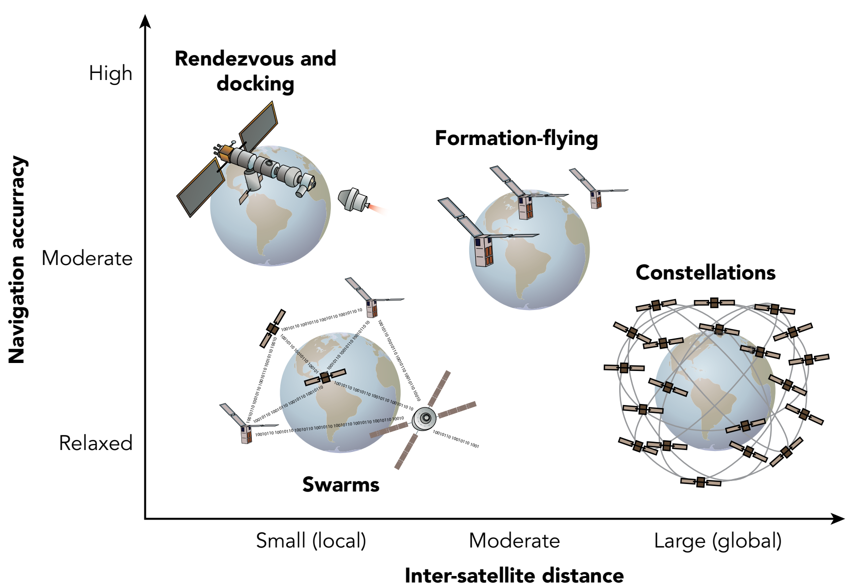
Another categorization of space systems focuses on their composition. Distributed space systems, comprising multiple interacting spacecraft, can achieve objectives that are difficult or impossible for single spacecraft. These systems take various architectural forms (see figure 9.2), defined by parameters like inter-spacecraft distances, required navigational accuracy, and number of satellites. They contrast with traditional single-spacecraft systems and offer expanded capabilities in space operations.

- Rendezvous and docking to support crew transportation, removal of space debris from orbit, in-orbit servicing of satellites, and assembly of larger structures in space. This involves small separations and high positional accuracy.

- Formation-flying architectures for observational missions that call for large effective apertures, such as space-based telescopes whose optical components are controlled very precisely at separations of tens to hundreds of meters

- Swarms that cooperatively sense the environment or share resources such as power, computation, and communications but whose components do

FIGURE 9.2 Characterizing distributed space systems



Source: Adapted from a diagram by Simone D'Amico

FIGURE 9.3 An artist's conception of a satellite swarm in space

Source: NASA / Blue Canyon Technologies

not necessarily need to be at fixed distances from one another (see figure 9.3)

- Constellations separated by tens of thousands of kilometers so that they may provide global ground coverage for navigation, communications, and remote sensing services

Key Developments

Impacts of Space Technologies

Space technology has proven its value to the national interest. Some of the most important applications today include the following:

Navigation This includes positioning, navigation, and timing (PNT) services around the world and in space. GPS and similar services operated by other nations help people know where they are and how fast and in which direction they are going, whether on land, on the ocean surface, in the air, or in space. Less well known is the timing information that GPS provides—timing that is accurate to the nanosecond is available anywhere in the world. This is a key tool for the financial sector, electric power grid, and transportation. Companies such as Xona Space Systems, a start-up founded by Stanford alumni, have begun developing GPS alternatives that aim to deliver even greater precision and robustness.

Communications Satellites provide vital communications in remote areas and for mobile users, complementing the terrestrial networks that carry most long-haul communications. Companies like SpaceX's Starlink, Amazon's Project Kuiper, Eutelsat OneWeb, and Astranis aim to offer low-latency, wide-coverage satellite internet. Recent innovations include optical communication systems, which use light for higher bandwidth and security. Kepler Communications is developing space-to-space and space-to-ground optical data relays, working toward an in-space internet for high-bandwidth satellite communications.

Remote sensing Remote sensing satellites, with their unique vantage point and sophisticated sensors, rapidly gather extensive data about areas and objects of interest. These data are integrated to create a "digital twin" of Earth, enhancing prediction, simulation, and response to terrestrial phenomena. Applications include disaster response, environmental monitoring, topographical mapping, and geospatial intelligence tracking human, animal, and marine activity. Governments are expanding remote sensing programs, complemented by commercial companies like BlackSky, Maxar, Planet Labs, and Capella Space. Recent efforts have focused on increasing data resolution, reducing response times, and exploring other valuable information modes such as hyperspectral imaging,[4] synthetic aperture radar,[5] and radio-frequency sounding (i.e., exploration of the environment through the use and exploitation of radio waves).[6]

Scientific research Space-based astronomy and exploration provide in-depth insights into the origins of planets, stars, galaxies, and life on Earth. The past year has seen significant strides in solar system exploration, particularly involving asteroids. NASA's OSIRIS-REx mission successfully returned asteroid samples to Earth and will soon be followed by the launch of the Psyche mission, which intends to examine at close range a metal-rich asteroid worth potentially quadrillions of dollars.

Space transportation The space transportation industry has seen launch costs drop by more than an order of magnitude over a couple of decades to $1,500 per kilogram in 2021.[7] Companies like SpaceX, Rocket Lab, Blue Origin, and Virgin Galactic have made progress in providing reliable launches and developing new vehicles. SpaceX's Starship—the most powerful rocket ever built—could dramatically reduce the cost of achieving LEO orbits, aspirationally between 10 and 100 times cheaper than today (see figure 9.4).[8] By the time this report is published, Blue Origin may have already launched the reusable high-volume heavy-lift New Glenn rocket.

Meanwhile, Blue Origin, Voyager Space, and Axiom Space are developing commercial space stations to replace the ISS, which NASA plans to decommission in 2030. These new stations aim to ensure continued orbital research and expand human presence in space.

National security Spacecraft constantly scan Earth for missile launches (both ballistic missiles and hypersonic missiles) aimed at the United States or its allies, nuclear weapons explosions anywhere in the world, radio traffic and radar signals from other countries, and the movements of allies and enemies in military contexts. US government investment in space for national security purposes continues to grow, including new commercial partnerships focused on data sharing for tracking objects in space and on Earth, satellite internet for battlefield communications, and research and development to maintain space superiority and safety as space becomes increasingly congested and contested because of an influx of nation-state and private-sector actors.

Trends in Space Technology

Privatization, miniaturization, and reusability The space sector is shifting from government-owned legacy systems and their long development timelines and mission lifetimes to a "NewSpace" economy driven by private companies. This privatization makes space technologies more accessible and less expensive. CubeSats and reusable rockets like SpaceX's Falcon 9 exemplify nongovernment innovations enabling new opportunities. Governments are also embracing small spacecraft and on-demand launches to expand space capabilities cost-effectively. The combination of smallsats and distributed architectures offers advantages in reduced costs, faster development timelines,

FIGURE 9.4 SpaceX's Starship could dramatically reduce the cost of achieving LEO orbits

Source: SpaceX. Used under CC BY-NC 2.0

frequent technology updates, and improved resilience, flexibility, and performance.

However, the private sector's rapidly increasing role in space also presents new challenges. These include dealing with risks inherent in dual-use space technologies; managing crises in a realm where lines separating individual private actors, the space sector as a whole, and government actors are increasingly blurred; differentiating between accidents and malevolent actions; and relying on companies whose interests may not be fully aligned with those of the US government. For example, technology for removing space debris could also be harnessed for antisatellite purposes.

The new Moon rush Recent years have seen a renewed desire to maintain a permanent human presence in lunar orbit and on the lunar surface. The abundance of certain materials on the Moon provides opportunities for mining and manufacturing. Such activities would reduce the amounts of material that would otherwise have to be transported from Earth. Combined with the significantly lower amount of fuel needed to launch from the Moon rather than from Earth, moon mining and manufacturing facilitates the construction of moon bases, the conduct of space exploration missions, and even launches into LEO that could be undertaken with hardware manufactured with materials from the Moon.

There have been a number of successful lunar landings recently. India became the first nation to touch down near the lunar south pole—a prime target for settlement—and China became the first nation to land on the Moon's far side. Japan became the fifth nation to successfully touch down on the Moon, and Intuitive Machines became the first private company to land on the Moon.[9] The NASA-led Artemis program is developing a new launch system, lunar base camp, and lunar terrain vehicles, among other things, as steps needed for establishing a permanent human presence on the Moon.[10]

Over the Horizon

New Applications of Space Technologies

Manufacturing For certain types of manufacturing, such as specialized pharmaceuticals, optics, and semiconductors, space offers two major advantages over terrestrial manufacturing. Because the vacuum of space is very clean, minimizing contamination is much easier. Further, the microgravity environment of space means that phenomena resulting from the effects of gravity—such as sedimentation, buoyancy, thermal convection, and hydrostatic pressure—can be minimized. This enables, for example, the fabrication of more perfect crystals and more perfect shapes. Production processes for biological materials, medicines, metallizations, polymers, semiconductors, and electronics may benefit.

Mining The Moon and asteroids may well have vast storehouses of useful minerals that are hard to find or extract on Earth, such as rare-earth elements that are used in batteries and catalytic converters as well as in guidance systems and other defense applications. Helium-3 found on the Moon may be an important source of fuel for nuclear fusion reactors. Future space-mining operations may bring these resources back to Earth to meet growing demand in a sustainable way. Mining of regolith (loose rock that sits atop bedrock) and ice on the Moon is also critical for enabling a permanent human presence there and supporting subsequent expansion into the solar system.

Power generation Above Earth's atmosphere, in certain orbits, the sun shines for twenty-four hours a day and even more brightly than it does on Earth's surface. To meet clean energy needs, this permanent sunlight could eventually be harnessed for space-based power generation. Solar panels could generate electricity in orbit, later beamed to the surface via microwaves, or orbital mirrors could be used to reflect sunlight onto Earth's nightside, to be collected by solar farms. If realized, dramatically lower launch costs (which may be possible thanks to new launch vehicles) make such concepts more technically and financially feasible today than they have been in the past.

Space situational awareness (SSA) The number of active satellites has increased from roughly 1,000 in 2014 to about 10,000 in 2024—a figure that will likely rise to several tens of thousands in the next decade. In addition, the European Space Agency estimates that about 170 million pieces of debris larger than 1 millimeter in size are in orbit, many of which are dangerous to satellites and space stations.[11] Ground-based stations are currently used to track space objects, but there is a push toward leveraging space-based sensors for more timely and accurate results. For example, a piece of space debris carrying an electrical charge and moving through plasma may create plasma signatures called solitons that signal its presence.[12]

Companies such as NorthStar are making initial strides toward creating satellite constellations for SSA. The emergence of low-cost, high-quality imagery and other information from space-based assets—increasingly launched and operated by private companies—will also be an important driver of open-source intelligence that data analysts can buy.

In-space servicing, assembly, and manufacturing (ISAM) Leadership, security, and sustainability in space require ISAM capabilities to approach, inspect, repair, refuel, or remove space assets without jeopardizing the space environment.[13] Spacecraft autonomy, in combination with rendezvous, proximity operations, and docking (RPOD), is a critical technology for ISAM. RPOD refers to the ability of spacecraft to operate autonomously, in combination with the ability to approach one another precisely and conduct close-up operations.

For example, orbital tugs and in-orbit fuel depots that can autonomously support in-orbit refueling are necessary components of a circular space economy that emphasizes the reuse and regeneration of products and materials in a sustainable manner. Hundreds of companies are developing new ISAM technology, but only a handful have demonstrated early RPOD capabilities in orbit—these include recent successes achieved by Astroscale (using its ADRAS-J smallsat in LEO) and Northrop Grumman (via its larger MEV satellites in GEO).

Energy harvesting in the space environment A piece of space debris moving in space carries significant kinetic energy. If it collides with another space structure, it generates energy that, in principle, can be harvested and put to constructive use. Rather than just asking, "Can we shield ourselves from the space environment?" we can also ask, "How can we harness it?"

Exploration A critical limitation for space exploration missions is travel time: Getting to the outer solar system can take ten years or more. As spacecraft fly ever farther from the Sun, they will need novel forms of power, such as sources driven by nuclear reactions, for the propulsive energy needed to make their missions possible.[14] Better propulsion systems that can be quickly deployed will also be needed to intercept interstellar objects so that samples can be collected from them.

On-demand space exploration missions Today, it takes a very long time to prepare for exploration missions, which means that targets of opportunity that suddenly appear cannot be visited by such missions. An on-demand capability would enable the close-up investigation of suddenly appearing targets such as the Oumuamua interstellar object, which passed through the solar system in 2017. Undertaking such a mission requires that a spacecraft can be made ready for launch shortly after the target is identified. Because such targets are likely to originate outside our solar system, the scientific return from bringing back a sample would be enormous.

Challenges of Innovation and Implementation

THE GRAND CHALLENGE OF SUSTAINABILITY

Sustainability encompasses both terrestrial sustainability *enabled by* space and the sustainability *of* humankind's use of space.

Sustainability *enabled by* space incorporates several of the technologies described above: for example, creating Earth's digital twin for disaster prevention and management, which requires integrating data from industry, government, and academia with advanced machine-learning techniques. Space-based solar power and resource extraction from the Moon and other celestial bodies are among other facets that illustrate the potential here.

Sustainability *of* space aims to create a circular, equitable space economy. Unlike Earth's organized transportation systems (which include traffic laws and gas stations), space lacks similar infrastructure. Addressing this requires making space assets reusable, establishing orbital services, managing debris, and quantifying orbital capacity. Space traffic management is essential to handle the increasing number of assets around the Earth and Moon. Developing guidelines for fair and safe orbital behavior—which don't currently exist—is essential.

The world ultimately faces a spaceflight sustainability paradox: The growing use of space to support sustainability and security on Earth will lead to more adverse impacts on the space environment itself. For example, multiple constellations of remote sensing satellites will contribute to greater space traffic challenges. Managing this complex issue will require advances in both policy and technology.

GAPS IN SMALL SATELLITE TECHNOLOGY

NASA has identified a list of issues that are restraining growth in usage of small spacecraft.[15] They include limitations in launch capacity, autonomous capabilities, PNT capabilities, and propulsion systems. The past year has seen several mission failures due to these and other shortfalls.

NASA is responding to these challenges via technology demonstration missions such as Starling, which in 2024 became the first successful in-orbit demonstration of several critical autonomous swarming technologies.[16] Among Starling's payloads is an optical PNT system newly developed by Stanford's Space Rendezvous Laboratory. This applies onboard cameras and advanced space robotics algorithms to navigate multiple satellites using only visual data. In doing so, it addresses a key technological gap for small spacecraft, which must typically rely on jammable GPS or expensive ground-based resources for navigation. However, more work is needed to take full advantage of smallsat architectures—and, by extension, distributed architectures featuring many smallsats working together.

TRIPLE-HELIX INNOVATION

Public entities have become risk averse in space, prioritizing accountability over innovation unless traditional methods fail. In contrast, private companies pursue innovation for profit and competitive advantage. Collaborative efforts between academia and industry often focus on technology commercialization and real-world demonstrations, frequently supported by governments. This cooperative model, known as triple-helix innovation, combines academia, industry, and government. Notable examples include the proposed $2 billion Berkeley Space Center collaboration with NASA's Ames Research Center and Stanford University's Center for AEroSpace Autonomy Research (CAESAR), which focuses on autonomy with Blue Origin and Redwire.

Policy, Legal, and Regulatory Issues

SPACE GOVERNANCE

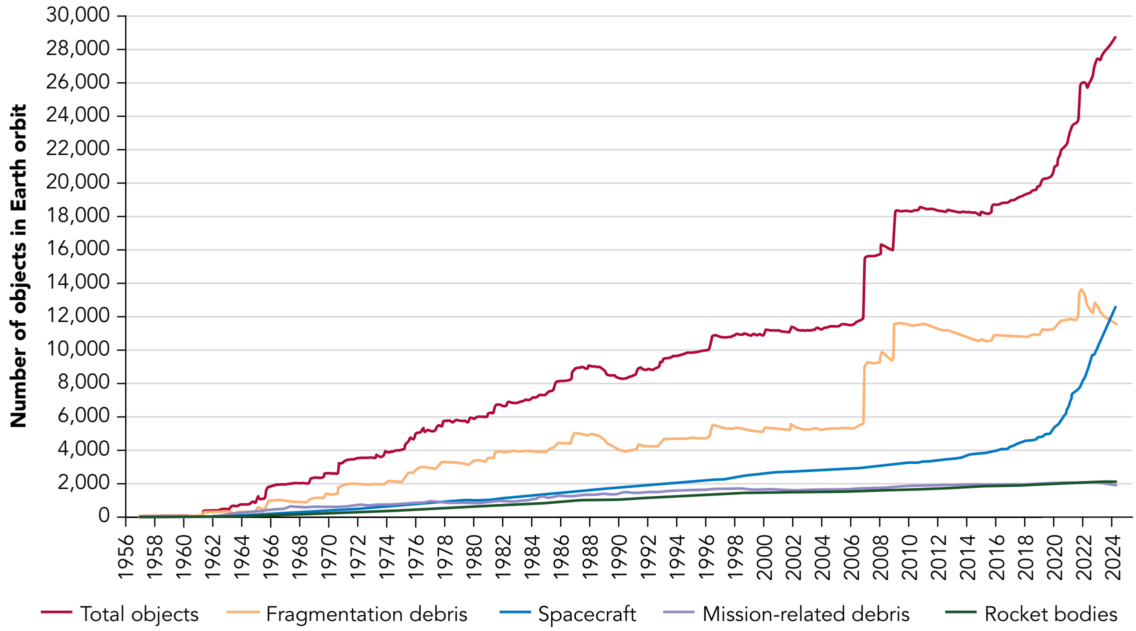
International and national space governance has not developed at the same rapid pace as space technology. Existing legal frameworks—many of which were products of the Cold War—do not address wide swaths of current activities and are often contested in scope and interpretation.[17] Attempts at improvement have often stagnated due to differing geopolitical aims. Even within the United States, space assets are not designated as critical infrastructure by the government despite their importance, and growth in space activity far outpaces the capabilities of current licensing processes run by the Federal Aviation Administration and the Federal Communications Commission (FCC).

Nonetheless, a number of developments in the past couple of years are notable. NASA released its strategy, including actionable objectives, for sustainability in space activities in Earth orbit.[18] It also promised to release similar strategies in the future for activities on Earth; the orbital area near and around the Moon known as cislunar space; and deep space, including other celestial bodies. In addition, the FCC issued its first-ever fine for a satellite not properly disposed of from geostationary orbit.[19] These short-term policy advances must be unified with a longer-term vision encompassing the next fifty to one hundred years to effectively address national security needs, support the space industry's continued development, and realize the responsible use of space as a global commons.

MAINTAINING SPACE ACCESS

The number of objects in space has grown rapidly. Figure 9.5 shows the total number of tracked space objects larger than 10 centimeters since 1959. Today, there are nearly 30,000 such objects, of which about 10,000 are working satellites.[20] There are also an estimated 1.1 million fragments between 1 and 10 centimeters in size. With so many objects in space, the risks of collision between them are growing. Each collision has the capacity to create even more debris, leading to a catastrophic chain reaction known as the Kessler syndrome, which would effectively block access to space. In addition, increasing volumes of space traffic (future mega-constellations will consist of tens of thousands of satellites) may lead to communications interference, and coordination of space

FIGURE 9.5 The number of objects in space larger than 10 centimeters has grown rapidly



Source: Adapted from *Orbital Debris Quarterly News* 28, no. 3 (July 2024): 10

activities such as orbit planning will be increasingly difficult to manage.

To tackle this issue, new domestic safety legislation and international cooperation will be needed for accurate tracking of space objects, facilitating the use of automated collision-avoidance systems, and removing debris from orbit. Similarly, more consistent guidelines will be needed to govern behavior in space, how space operations are conducted, and the sharing of data for situational awareness. Transparency and coordination among all players will be key, and the United States is in a good position to take a leading role among like-minded nations in advocating for these kinds of changes in space access.

GEOPOLITICS, SECURITY, AND CONFLICT IN SPACE

Many issues arise with respect to space and geopolitics. A key example is the Outer Space Treaty (OST), which entered into force in 1967; today, 115 countries are parties to the treaty, and 22 more have signed but not ratified it. Among other things, the treaty prohibits the placement of nuclear weapons or other weapons of mass destruction in space.

Recent evidence suggests the OST's norms are eroding—in 2024 Russia vetoed a United Nations resolution prohibiting the deployment of nuclear weapons in space, and senior US officials revealed that they believe Russia is developing a satellite to carry nuclear weapons into LEO, where a detonation could destroy all satellite activity there for up to a year.[21]

In addition, there is no treaty, OST or otherwise, that limits other military uses of space, including the placement of conventional weapons in orbit. And because space-based capabilities are integral to supporting modern warfighters, they may become targets of foreign counterspace threats. Rapid-launch

While prestige remains a factor, the current [Moon] race focuses on establishing a lunar presence for strategic and economic advantages.

capabilities to facilitate fast replacement of satellites rendered inoperative during conflict will increase the resiliency of critical national space assets.

A second issue relates to nonnuclear anti-satellite weaponry and capabilities. To date, four nations—China, Russia, India, and the United States—have successfully tested kinetic anti-satellite weapons capable of physically destroying satellites in space. (Every such test has produced a significant amount of space debris.) More broadly, countries are developing a range of capabilities, from the ground and in space, to degrade, deny, and even destroy satellites of other nations. Cyberattacks are an important element of the non-kinetic threat spectrum against space missions, which can lead to data corruption, jamming, and hijacking of space intelligence providers and customers.[22]

A third issue involves various national efforts to reach the Moon. To facilitate an orderly and peaceful exploitation of the Moon's resources that is consistent with the OST, NASA and its partners have also proposed the Artemis Accords, which define "principles for cooperation in the civil exploration and use of the Moon, Mars, comets and asteroids for peaceful purposes."[23] So far, forty-three nations have signed the accords—but notably not Russia or China, which are among the parties seeking to establish a permanent Moon presence.

However, nations today are engaged in a new "race to the Moon," though with different motivations than in the 1960s. While prestige remains a factor, the current race focuses on establishing a lunar presence for strategic and economic advantages. The first nation to successfully establish a lunar presence may well gain a first-mover advantage that enables it to be in a stronger position to set the terms for others to come. Although the OST prohibits claiming lunar sovereignty, there are concerns that nations might disregard this for national interests.[24] The possibility of a nation taking military action to prevent others from establishing their own lunar presence highlights the potential for conflict in this new space race.

Finally, in the past couple of years, the rise of the private sector's importance in providing capabilities for space launch and space-based communications has been dominated by SpaceX and Starlink, which are owned by the same person. In 2022, the CEO of Starlink, which Ukrainian military forces relied on for communications, denied its use to conduct military operations around Sevastopol, in Crimea—thus directly interfering with the execution of Ukrainian battle plans.[25] In September 2024, NASA turned to SpaceX to return to Earth two US astronauts left on the ISS when their Boeing-built Starliner spacecraft experienced operational failures and was brought to Earth without them.

Such incidents demonstrate the extreme dependence of the US government on capabilities provided by a very limited number of companies that are controlled by a single individual—and raises important policy questions of how to ensure that US space efforts can continue in accordance with US national interests.

NOTES

1. Alizée Acket-Goemaere, Ryan Brukardt, Jesse Klempner, et al., "Space: The $1.8 Trillion Opportunity for Global Economic Growth," World Economic Forum and McKinsey & Company, April 8, 2024, https://www.mckinsey.com/industries/aerospace-and-defense/our-insights/space-the-1-point-8-trillion-dollar-opportunity-for-global-economic-growth.

2. SpaceWorks, "Recap of All Global Launches for 2023," January 4, 2024, https://www.spaceworks.aero/recap-of-all-global-launches-for-2023.

3. Stevan M. Spremo, Alan R. Crocker, and Tina L. Panontin, *Small Spacecraft Overview*, NASA Technical Reports Server, May 15, 2017, https://ntrs.nasa.gov/api/citations/20190031730.

4. Anuja Bhargava, Ashish Sachdeva, Kulbushan Sharma, et al., "Hyperspectral Imaging and Its Applications: A Review," *Heliyon* 10, no. 12 (June 2024): e33208, https://doi.org/10.1016/j.heliyon.2024.e33208.

5. Kelsey Herndon, Franz Meyer, Africa Flores, et al., "What Is Synthetic Aperture Radar?," NASA Earthdata, accessed September 13, 2024, https://www.earthdata.nasa.gov/learn/backgrounders/what-is-sar.

6. J. Wickert, T. Schmidt, M. Schmidt, et al., "GNSS Radio Occultation," Global Geodetic Observing System, accessed September 13, 2024, https://ggos.org/item/gnss-radio-occultation/.

7. Ryan Brukardt, "How Will the Space Economy Change the World?," McKinsey & Company, November 28, 2022, https://www.mckinsey.com/industries/aerospace-and-defense/our-insights/how-will-the-space-economy-change-the-world.

8. Citi GPS, "Space: The Dawn of a New Age," Citi, May 9, 2022, https://www.citigroup.com/global/insights/space_20220509.

9. The landing was a mixed success. "Intuitive Machines Calls Moon Mission a Success Despite Soft Crash Landing," *Space & Defense*, February 26, 2024, https://spaceanddefense.io/intuitive-machines-calls-moon-mission-a-success-despite-soft-crash-landing/.

10. National Aeronautics and Space Administration, "Artemis Plan: NASA's Lunar Exploration Program Overview," September 2020, https://www.nasa.gov/wp-content/uploads/2020/12/artemis_plan-20200921.pdf?emrc=f43185.

11. European Space Agency, "How Many Space Debris Objects Are Currently in Orbit?," accessed September 13, 2024, https://www.esa.int/Space_Safety/Clean_Space/How_many_space_debris_objects_are_currently_in_orbit.

12. Abhijit Sen, Rupak Mukherjee, Sharad K. Yadav, et al., "Electromagnetic Pinned Solutions for Space Debris Detection," *Physics of Plasmas* 30, no. 1 (January 2023): 012301, https://doi.org/10.1063/5.0099201.

13. White House, "White House Office of Science and Technology Policy Unveils National In-Space Servicing, Assembly, and Manufacturing (ISAM) Implementation Plan," press release, December 16, 2022, https://www.whitehouse.gov/ostp/news-updates/2022/12/16/white-house-office-of-science-and-technology-policy-unveils-national-in-space-servicing-assembly-and-manufacturing-isam-implementation-plan.

14. Beth Ridgeway, ed., "Space Nuclear Propulsion," National Aeronautics and Space Administration, last modified August 19, 2024, https://www.nasa.gov/tdm/space-nuclear-propulsion.

15. Ames Research Center, "State-of-the-Art Small Spacecraft Technology," National Aeronautics and Space Administration, February 2024, https://www.nasa.gov/smallsat-institute/sst-soa/.

16. Justin Kruger, Simone D'Amico, and Soon S. Hwang, "Starling Formation-Flying Optical Experiment: Initial Operations and Flight Results," paper presented at 38th Small Satellite Conference, Logan, UT, August 8, 2024, https://digitalcommons.usu.edu/smallsat/2024/all2024/130/.

17. Sophie Goguichvili, Alan Linenberger, Amber Gillette, et al., "The Global Legal Landscape of Space: Who Writes the Rules on the Final Frontier?," Wilson Center, October 1, 2021, https://www.wilsoncenter.org/article/global-legal-landscape-space-who-writes-rules-final-frontier.

18. National Aeronautics and Space Administration, "NASA's Space Sustainability Strategy: Volume 1: Earth Orbit," April 9, 2024, https://www.nasa.gov/wp-content/uploads/2024/04/nasa-space-sustainability-strategy-march-20-2024-tagged3.pdf.

19. Space Bureau, Enforcement Bureau, and Media Relations Bureau, "FCC Takes First Space Debris Enforcement Action," Federal Communications Commission, October 2, 2023, https://www.fcc.gov/document/fcc-takes-first-space-debris-enforcement-action.

20. *Orbital Debris Quarterly News* 28, no. 3 (July 2024): 10, https://orbitaldebris.jsc.nasa.gov/quarterly-news/pdfs/ODQNv28i3.pdf.

21. Audrey Decker, "Russian Space Nuke Could Render Low-Earth Orbit Unusable for a Year, US Official Says," *Defense One*, May 1, 2024, https://www.defenseone.com/threats/2024/05/russian-space-nuke-could-render-low-earth-orbit-unusable-year-us-official-says/396245/; Robert Wood, "Remarks at a UN General Assembly Meeting on Russia's Veto of the U.S. and Japan-Drafted UNSC Resolution on Preventing Nuclear Weapons in Outer Space," US Mission to the United Nations, May 6, 2024, https://usun.usmission.gov/remarks-at-a-un-general-assembly-meeting-on-russias-veto-of-the-u-s-and-japan-drafted-unsc-resolution-on-preventing-nuclear-weapons-in-outer-space/.

22. Northern Sky Research, *Global Space Economy*, 3rd ed., January 2023, https://www.nsr.com/?research=global-space-economy-3rd-edition.

23. National Aeronautics and Space Administration, "The Artemis Accords," accessed September 13, 2024, https://www.nasa.gov/artemis-accords.

24. China has chosen to disregard the findings of a tribunal convened pursuant to the United Nations Convention on the Law of the Sea that ruled against Chinese claims regarding its land reclamation activities in the South China Sea, despite the fact that China is a formal signatory to UNCLOS. See In the Matter of the South China Sea Arbitration (Phil. v. China), Case No. 2013-19, Award of July 12, 2016, ITLOS Rep., https://pcacases.com/web/sendAttach/2086.

25. Wes J. Bryant, "When a CEO Plays President: Musk, Starlink, and the War in Ukraine," *Irregular Warfare Initiative*, October 17, 2023, https://irregularwarfare.org/articles/when-a-ceo-plays-president-musk-starlink-and-the-war-in-ukraine/.

STANFORD EXPERT CONTRIBUTORS

Dr. Simone D'Amico
SETR Faculty Council and Associate Professor of Aeronautics and Astronautics and, by courtesy, of Geophysics

Dr. Sigrid Elschot
Associate Professor of Aeronautics and Astronautics

Dr. Anton Ermakov
Assistant Professor of Aeronautics and Astronautics and, by courtesy, of Geophysics

Dr. Debbie Senesky
Associate Professor of Aeronautics and Astronautics and of Electrical Engineering

Dr. Tommaso Guffanti
SETR Fellow and Postdoctoral Scholar in Aeronautics and Astronautics

Dr. Justin Kruger
SETR Fellow and Postdoctoral Scholar in Aeronautics and Astronautics

10

SUSTAINABLE ENERGY TECHNOLOGIES

KEY TAKEAWAYS

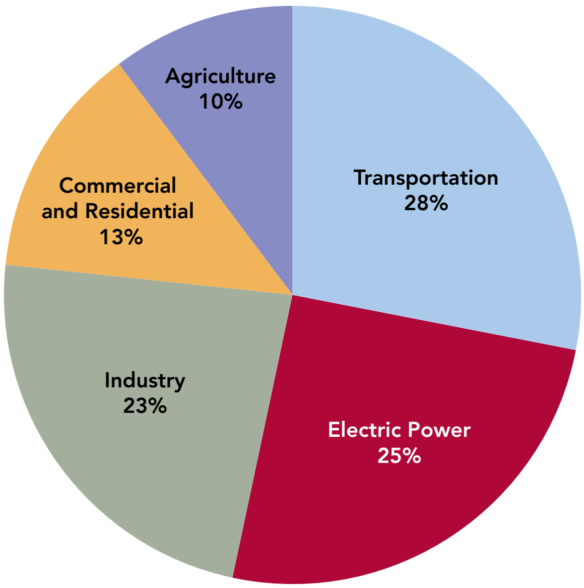
- Although many clean energy technologies are now available and increasingly affordable, scaling them to a meaningful degree and building the massive infrastructure needed to deploy them will take decades.
- The largest impact on reducing emissions in the near to medium term will come from building a no- to very-low-emission electricity grid, electrifying passenger cars and small commercial vehicles, and transitioning residential and commercial heating and industrial energy.
- In the long term, technologies for decarbonizing buses and long-haul trucks, decarbonizing carbon-intensive industries, and reducing greenhouse gases from refrigerants and agriculture will play key roles in a net-zero, emissions-free energy infrastructure.

Overview

Energy is a key strategic resource. Fossil fuels have dominated human energy use for centuries, but their emissions of carbon dioxide (CO_2) and other greenhouse gases have significantly altered the climate. Without a shift to sustainable energy sources in the coming decades, this trend will worsen, exacerbating climate change impacts globally. Figure 10.1 depicts the percentage of greenhouse gas emissions associated with each sector of the US economy.

Scaling solutions is a major global challenge. We need billions of zero-emission cars, hundreds of millions of emission-free trucks, and carbon-neutral fuels for tens of millions of airplane flights per year. Agricultural practices must change to feed billions sustainably. The global industrial infrastructure that mines and processes raw materials and turns them into manufactured products and then distributes these all over the world must adopt methods that eliminate CO_2 and other emissions.

FIGURE 10.1 Greenhouse gases emitted in the United States in 2022, by economic sector



Note: Due to rounding, total does not add up to 100 percent.

Source: Adapted from US Environmental Protection Agency, "Sources of Greenhouse Gas Emissions," April 2024

Another statistic underlines the magnitude of the challenge that lies ahead: Tens of billions of tons of CO_2 are produced every year from burning fossil fuels. Figure 10.2 demonstrates how this scale challenge is reflected in various parts of the economy.

This challenge has three major implications. First, no single technology can meet global energy demands in a net-zero emissions manner. Success requires diverse approaches that create a bridge from present sources, consumption trends, and infrastructure to a more sustainable future. Second, new energy technologies must match rising global power demand. Third, solutions must be cost-effective, ensuring energy affordability for people of all economic circumstances worldwide.

Renewable energy-generation technologies have experienced remarkable growth in recent years, primarily driven by declining costs of solar and wind power that make them competitive with conventional energy sources today. However, scaling distributed renewable generation capacity to meet global demand still presents challenges when it comes to storing and transporting energy.

The electric grid—a complex system that generates and transports electrical energy to end users—is central to the energy transition,[1] and it will require increased capacity and reliability as countries decarbonize. Building reliable, emission-free grids involves combining intermittent renewables like solar and wind (i.e., sources that are not always available) with clean, dispatchable power sources (i.e., ones that can be quickly ramped up or down in response to user demand) and improved energy storage. Despite the challenges posed by the variability of some renewables, many countries, including Switzerland, Denmark, and Brazil,[2] have successfully integrated a significant portion of these sources into their grids.

The economic impact of sustainable energy is significant. For example, one analysis indicates that clean energy accounted for 10 percent of global GDP growth in 2023,[3] or about $320 billion. Sustainable energy also contributes substantially to employment—an estimated 13.7 million direct and indirect global renewable-energy jobs existed in 2022, up from 7.3 million in 2012.[4]

Public health is also likely to benefit from emissions-free energy sources. For example, one report predicts that eliminating air pollution emissions from energy-related activities in the United States could prevent more than fifty thousand premature deaths each year and provide more than $600 billion in benefits annually from avoided illness and death.[5] A second report indicates that an electric grid producing 80 percent of its output from emission-free sources by 2030 could prevent an estimated 267,500 premature deaths between 2030 and 2050 and generate an estimated $1.13 trillion in present-value health benefits due to cleaner air.[6]

FIGURE 10.2 The scale challenge in global energy transition

Cars

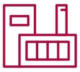
- There are an estimated 1.45 billion cars on the road today.
- Globally, one in every seven cars bought in 2022 was an electric vehicle.[a]
- Global lithium production will have to increase by a factor of 2.5 to 5 to meet expected demand for electric vehicles by 2030.[b]

Trucks

- Globally, there are approximately 217 million freight vehicles (including light commercial vehicles, medium- and heavy-duty trucks, and buses).[c]
- In 2022, 1.2% of trucks sold worldwide were electric (60,000 units).[d]

New Construction

- From 2020 to 2060, we expect to add about 2.6 trillion square feet of new floor area to the global building stock—the equivalent of adding an entire New York City to the world every month for 40 years.[e]
- After water, concrete is the second most-consumed material, with 30 billion tons poured each year.[f]
- If the cement industry were a country, it would be the third largest carbon dioxide emitter, behind only China and the United States.[g]



Industrial Energy Use

- Heavy industry—including steel, cement, and chemical production—accounts for nearly 40% of global carbon dioxide emissions. These emissions are the hardest to decarbonize and would require both an entire change in process and building new processing plants, which would need even more steel and cement.[h] Currently there are zero cement plants and only one steel plant that don't produce carbon dioxide.[i]

Source:

a. https://www.iea.org/energy-system/transport/electric-vehicles.

b. https://www.popularmechanics.com/science/energy/a42417327/lithium-supply-batteries-electric-vehicles/; https://www.sustainabilitybynumbers.com/p/lithium-electric-vehicles. Research conducted by Oxford University data scientist Hannah Ritchie.

c. https://www.shell.com/energy-and-innovation/the-energy-future/decarbonising-road-freight.html.

d. https://www.iea.org/reports/global-ev-outlook-2023/trends-in-electric-heavy-duty-vehicles.

e. https://www.iea.org/data-and-statistics/charts/global-buildings-sector-co2-emissions-and-floor-area-in-the-net-zero-scenario-2020-2050.

f. https://www.climateworks.org/blog/why-you-should-care-about-cement-and-concrete.

g. https://www.theguardian.com/cities/2019/feb/25/concrete-the-most-destructive-material-on-earth.

h. https://www.brookings.edu/wp-content/uploads/2021/06/FP_20210623_industrial_gross_v2.pdf.

i. https://www.gatesnotes.com/2022-State-of-the-Energy-Transition.

Key Developments

In 2021, the White House laid out a strategy to reach net-zero US emissions by 2050,[7] which includes an aspirational goal of 100 percent emissions-free electricity generation by 2035. Several key developments are shaping the transition toward a cleaner and more sustainable energy future, reflecting a combination of technological innovations and the impact of broader trends in energy systems, market dynamics, and policy interventions.

Substantial progress has already been made in several sustainable energy technologies:

- The cost of both wind-generated and solar-generated electricity is now substantially lower than that of fossil fuels.[8]
- The use of direct current (DC) for long-distance transmission of electricity can halve losses compared to transporting it using alternating current (AC), though AC transmission lines are much more common.

- Having dropped in cost by 90 percent over the past decade, lithium-ion (Li-ion) batteries are now being used on a massive scale to store excess energy from renewable generation for use during peak demand periods.[9]

- Cheaper Li-ion batteries are also making possible the production of electric vehicles (EVs) with ranges in excess of several hundred kilometers.

- Light-emitting diode (LED) lighting is up to ten times more efficient than incandescent lighting at converting electricity to usable light, and massive deployment of LED lightbulbs is taking place around the world.

- Heat pumps for heating and cooling are highly energy efficient, even at low temperatures, and heat-pump deployments are spreading rapidly.

The widespread deployment of such technologies in a net-zero emissions energy infrastructure depends on overcoming a variety of challenges, some of which include the following:

Public charging infrastructure for EVs Although US legislation has sought to fund the deployment of five hundred thousand public charging stations by 2030, the number of such stations that have been built since the legislation's passage is minuscule—just seven stations as of June 2024.[10]

The raw materials supply chain The United States relies heavily on imports of rare earths from China and cobalt from the Congo for clean energy technologies such as wind turbines and batteries—nations whose interests are not always aligned with US national interests. The United States also competes for the same materials with other nations that are themselves attempting to reduce energy-related emissions.

High up-front costs As is often true with capital investments, some clean energy technologies (e.g., nuclear reactors, hydroelectric dams, offshore wind farms, and grid-scale battery-energy storage systems) require significant up-front investment before their benefits are fully realized. In the absence of appropriate financial support, high up-front costs can be a barrier to adoption, especially in developing economies.

Over the Horizon

In the United States, energy research is a major focus of government, academia, and the private sector. The US Department of Energy (DOE) is a substantial funder of innovation in many energy technologies.[11] The private sector also invests heavily in clean energy. According to the American Clean Power Association, $271 billion of private-sector investments in domestic clean energy projects and manufacturing facilities were announced in 2023, exceeding the combined amount of clean energy investments made over the previous eight years.[12]

Government, academia, and the private sector also conduct research on energy technologies. For the federal government, the US National Laboratories conduct a substantial amount of research in this area. Academia conducts the bulk of early-stage research on energy technologies, and universities have been the source of many innovations in energy—including better batteries and more efficient solar cells and wind turbines. However, they do not have the resources to effect large-scale deployments. Big energy companies, including some that have previously built their businesses around fossil fuels, are increasingly involved in the sustainable energy ecosystem. Startups are also involved in the commercialization of research that emerges from academia. Larger companies also conduct research, but generally at later stages that are past proof of concept. Both large and small companies have entered partnerships with academic institutions such as Stanford.

The discussion below focuses on technologies that are not as mature and where policy is less developed.

Energy Technologies

LONG-DURATION ENERGY STORAGE

Long-duration energy storage from intermittent sources such as wind and solar is necessary to capture their full value. Storage allows excess energy produced during plentiful times to be used when intermittent sources are unavailable—for example, solar energy is about twice as abundant in summer than in winter. A variety of technological concepts for storage are being developed:

Hydrogen storage Electricity can convert water to hydrogen gas through electrolysis. The gas can then be kept in deep underground salt caverns for large-scale seasonal energy storage, after which it can be converted back to electricity in a fuel cell or gas turbine. The first purpose-built salt cavern seasonal energy-storage system is being built in Utah.[13]

Gravity storage Power that exceeds demand can be used to pump water from lower to higher levels and then recovered by letting that water flow back down through generators. Alternatively, large, multi-ton weights can be lifted hundreds of meters using excess energy and then later allowed to fall gently to drive electrical generators.

Thermal storage This approach stores excess power in the form of heat, such as heating a large volume of salts to a very high temperature. When needed, this stored heat can be released to generate power.

Large-scale, long-duration battery storage To improve important battery characteristics such as cost, tolerance to extreme temperature, safety, maintainability, and recyclability, battery technologies beyond Li-ion batteries are being developed, such as redox flow,[14] Ni-H_2 gas,[15] and Zn-MnO_2 chemistries.[16]

The chief challenges of all these forms of long-duration energy storage are scalability and cost. None of them will be a silver bullet that exponentially improves energy storage. If they can become economically feasible, however, each one of the technologies could satisfy niche applications.

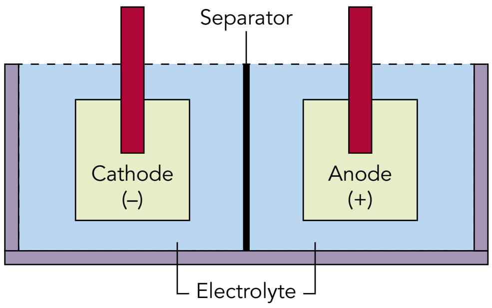
LOW-COST, HIGH-ENERGY-DENSITY BATTERIES

All batteries consist of three basic components: the cathode, made of one substance; the anode, made of a different substance; and the electrolyte, which consists of yet another substance (see figure 10.3). Batteries have different characteristics depending on the substances that constitute them—for example, the alkaline batteries used in a flashlight employ different materials than the lead-acid battery used in a standard car.

While batteries are now being deployed for short-term energy storage on the grid that lasts hours, they are currently not viable for longer-term storage. For example, the challenge of storing the world's electricity consumption for seventy-two hours for the long term (weeks or months) gives a very good sense of the scale challenge facing sustainable electricity generation.[17] Around 240 terawatt-hours (TWh) of battery storage would be needed; at about 200 watt-hours per kilogram, this would require over a billion tons of batteries.

Li-ion batteries, while currently the best large-scale production option, are not a good solution for long-duration storage. As measured in energy storage capacity, global production capacity of batteries in 2023 was about 2.8 TWh,[18] or barely enough to

FIGURE 10.3 Basic components of all batteries



cover 1 percent of the world's electrical consumption over three days. Producing enough capacity would take many decades, even if all available production were dedicated to this purpose. Moreover, such batteries would need to endure for thousands of cycles and retain charge for weeks or months. Current Li-ion technology doesn't meet these requirements, making it unsuitable for long-term, grid-scale energy storage solutions.

Cost is also a significant barrier for Li-ion batteries in grid storage. At current prices of $139 per stored kilowatt-hour, grid-scale Li-ion storage is prohibitively expensive. Aqueous chemistries, like manganese-hydrogen batteries, offer more cost-effective solutions for long-duration storage due to cheaper materials and lower maintenance needs.[19] Sodium-ion batteries also show promise, currently matching Li-ion costs but with potential for further price reductions as the technology matures. These alternatives could provide more economical options for large-scale grid energy storage in the future.

RENEWABLE FUELS: COMBUSTIBLE HYDROCARBONS AND BIODIESEL

Research on renewable fuels aims to create energy sources that do not rely on fossil fuels such as oil, gas, or coal. Renewable fuels include combustible hydrocarbons such as biodiesel, which can be produced from animal fats or vegetable oils; bioethanol produced from corn or algae; hydrogen, which can be produced from many sources; and ammonia produced using green hydrogen (see definition in the column at right).

Because it burns without CO_2 emissions, hydrogen is important for transitioning to renewable fuels. However, for most transportation applications, vehicles must carry hydrogen as a liquid or highly compressed gas to avoid frequent refueling. Hydrogen fuel cells are twice as efficient as hydrocarbon combustion engines, but compressed hydrogen contains only a quarter of the energy of an equivalent volume of hydrocarbon fuels. Consequently, hydrogen-powered vehicles need fuel tanks at least twice the size of conventional ones for the same range.

Research into hydrogen storage must therefore focus on developing cost-effective storage technologies with improved energy density that do not depend on liquification or compression.[20] It may also be possible to use captured CO_2 combined with sustainably produced hydrogen to create renewable hydrocarbon fuels.[21] Ammonia is another form of hydrogen storage and is being prototyped as a fuel for oceangoing shipping,[22] but concerns about its safety and impact on air pollution need to be addressed before wide-scale adoption occurs.

Producing and transporting hydrogen cost-effectively and with minimal leakage is challenging. Current hydrogen from fossil fuels, known as gray hydrogen, is unsustainable due to its carbon footprint. Blue hydrogen, based on methane with carbon capture, and green hydrogen, which uses renewable electricity for electrolyzing water, are emerging alternatives. Geologic hydrogen from natural water-rock interactions and hydrogen production from geothermal resources are also being explored.

CARBON CAPTURE AND REMOVAL

Emission-free energy production at the scale required will take many decades to accomplish, and fossil fuels will still be an appreciable (though declining) fraction of society's mix of energy sources for some time to come. In the meantime, carbon capture and storage (CCS) is gaining momentum. CCS involves capturing CO_2 emissions from industrial processes and power plants and then storing them underground, preventing their release into the atmosphere.

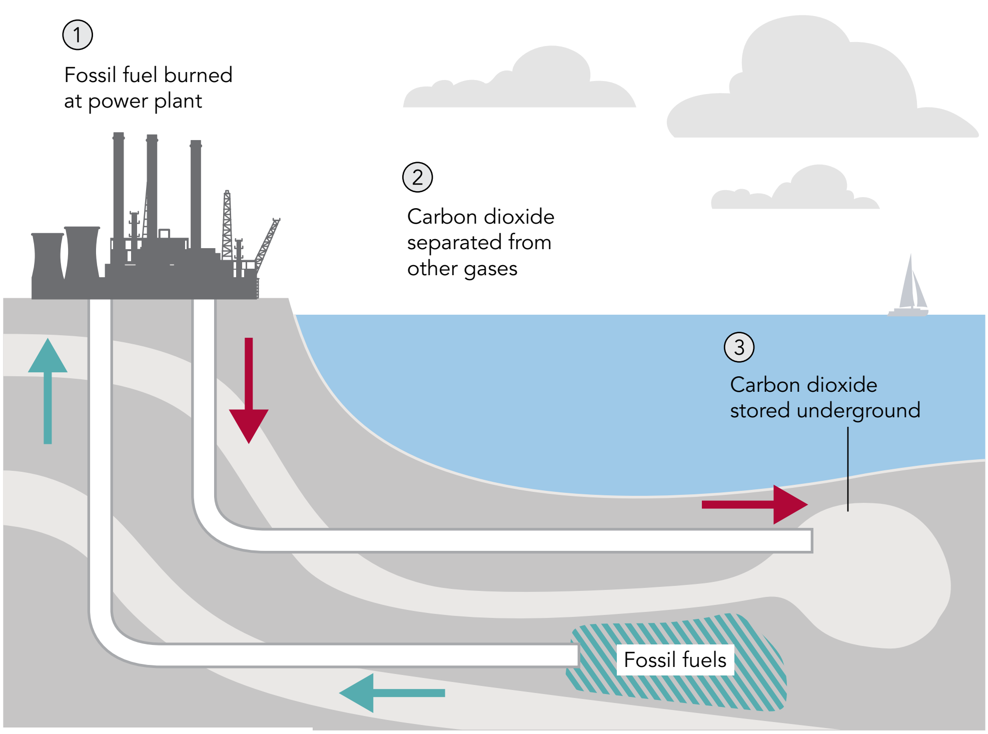
CCS in fossil-fuel power plants aims to reduce CO_2 emissions while maintaining the use of these sources. Currently, only 45 million tons of CO_2 are captured annually,[23] a very small fraction of global emissions. However, new government incentives, especially in the United States, have sparked increased CCS permit applications, and a growing number of projects are in various stages of development, from planning to already operational. If all planned CCS projects were completed, global capacity would increase eightfold. Yet this would still account for less than 1 percent of global CO_2 emissions from fossil fuels.

Carbon capture usually takes place at the source of emissions, such as the smokestack of a fossil-fuel-burning power plant. Here, CO_2 emissions are much more concentrated; once dispersed by the wind into the atmosphere, they become much harder to capture. Technologies to trap CO_2 at the source include liquid and solid materials, as well as membranes that extract CO_2 from the gas. After the CO_2 is captured, it can be permanently stored underground or used in the production of other materials (see figure 10.4).

Research into source capture based on liquid and solid materials is focused on developing inexpensive, energy-efficient materials for membranes that can be used for long periods of time to rapidly separate CO_2 from other gases. Membrane-development challenges include finding ways to reduce the cost and increase the stability of these approaches, as well as increasing membranes' permeability and their ability to extract CO_2.

Carbon removal calls for capturing CO_2 directly from the atmosphere—also known as direct air capture (DAC)—at concentrations much lower than at a

FIGURE 10.4 How carbon capture and storage works



Source: Adapted from BBC Research

The electric grid of the future will need to be more extensive and complex, with power generation, consumption, and storage distributed throughout the system.

smokestack. DAC uses significantly more energy in capturing a ton of CO_2 than capturing it at a power plant or other source. Several commercial-scale projects are now under development and will be operational by the end of the decade.[24] Potentially scalable DAC approaches include biomass storage; mineralization of CO_2 using silicates; ocean alkalinization; and storage in algae.[25] Research challenges involve improving the scalability of these approaches, the length of time over which carbon can be stored using them, and their affordability.

GRID TECHNOLOGIES

Today's electric grid in the United States is highly centralized and operates as a single unit through the real-time coordination of power plants spread across many states. Such coordination must constantly balance supply with demand. This creates the potential for widespread instability in the event of outages that would otherwise be highly localized.

The electric grid of the future will need to be more extensive and complex, with power generation, consumption, and storage distributed throughout the system. Renewable sources of electricity will be more varied and decentralized. Electrical demand will increase as electrically operated systems displace ones powered by fossil fuels. This will require a significant expansion in the electricity grid within the next few decades, which could see it double or triple in size. Energy storage and electricity demand will have to be more dynamically managed as well.

The grid must adapt to handle a significantly larger and more variable energy flow, which will require substantial upgrades to existing infrastructure, including transmission lines, substations, distribution networks, and information systems. Addressing all these challenges securely is the goal of what is widely known as the smart grid, which will coordinate these different elements to increase efficiency, reliability, and resilience against attack or natural disaster. In the United States, the Office of Electricity in the DOE has the primary federal responsibility for development of technologies, tools, and techniques for the smart grid.

Some of the important technologies for a smart grid include the following:

Reconductoring This involves replacing existing conductors on power lines with advanced ones to improve capacity, reduce losses, and/or enhance reliability without increasing land use.[26] Reconductoring is crucial for grid maintenance, meeting demand, and renewable energy integration.

End-use energy management There is increasing emphasis on making energy demand more flexible, which involves managing when energy is used. Grids dominated by renewable energy typically have excess supply at certain hours (for example, during the day for solar-dominated grids) and excess demand at other hours, necessitating the use of expensive energy storage. Demand management can reduce this disparity by using a variety of approaches to change when electricity consumption occurs:

- Scheduling devices and systems to operate during periods of high renewable-energy generation
- Implementing dynamic pricing structures that reflect the actual cost of electricity at any given time, incentivizing consumers to shift their energy use
- Allowing consumers to schedule their energy use to coincide with periods of low carbon intensity on the grid, further reducing emissions
- Incentivizing consumers to reduce their electricity consumption during peak demand periods, enhancing grid stability

Vehicle-to-grid (V2G) technology V2G enables EVs to return energy to the grid, enhancing stability, and lowers electricity costs for EV owners by allowing them to resell energy. However, implementing V2G requires significant upgrades to the grid as well as financial investment and stakeholder coordination. Hardware interfaces and communication protocols need to be standardized. A key concern here is that if V2G is very successful, it could degrade EV batteries faster. Further research is also needed to understand how the approach affects battery lifespan and warranties.

Second-life battery applications As EV adoption increases, more EV batteries will be retired. These often retain significant capacity and are particularly well suited for stationary energy storage applications, such as battery backup for solar power. Research is underway to develop specialized battery-management systems (BMSs) tailored to the unique characteristics and requirements of second-life batteries, optimizing their performance and further extending their lifespan.

AI and data-driven energy systems The growing field of energy artificial intelligence leverages AI, machine learning, and data analytics to optimize energy systems, improve grid management, and accelerate the integration of renewable energy sources. An important application involves AI algorithms that dynamically adjust grid operations in response to changing conditions, such as fluctuations in renewable energy generation or unexpected demand surges. Data analytics can also anticipate and prevent equipment failures in power plants and grid infrastructure.

NUCLEAR POWER

Energy can be released from certain atomic nuclei through fission, which is usually the splitting of the nuclei of uranium or plutonium, and fusion, the merging of the nuclei of hydrogen isotopes into one nucleus. (In this report, *nuclear power* will refer to power generated through fission.)

The technical feasibility of generating electricity through fission has been established for many decades, and today it generates just under a fifth of electricity consumption in the United States.[27] In addition, nuclear power—an emissions-free energy source—is experiencing renewed interest, particularly as a source of dispatchable power that can complement intermittent renewable energy sources.

At the same time, the widespread deployment of nuclear (fission) reactors has been inhibited by several factors, including the following:

Cost and timescales The construction of nuclear reactors in the United States has a long history of significant cost overruns and construction delays. These have significantly driven up the cost of fission-produced electricity compared with original estimates.

Fuel security Today, over 90 percent of the uranium used in US nuclear reactors is imported; Kazakhstan and Russia supply nearly half of all American uranium consumption. In addition, most new reactor designs call for uranium fuel enriched with U-235 at a higher level than that used in most of today's operating reactors.[28] Such fuel, known as high-assay low-enriched uranium (HALEU), is available only from Russia today.

Manufacturing capability A few hundred gigawatts of fission generation capacity will have to be brought online by 2050[29] if the United States is to achieve its goal of tripling its production of such electricity by that year from its 2020 baseline.[30] This corresponds to nearly three hundred gigawatt-scale reactors—or many more smaller ones—which would demand a historically unprecedented rate of reactor construction. Achieving this will depend on multiple factors, including the availability of a skilled construction labor force and heavy-manufacturing capacity.[31]

Safety A number of well-publicized, safety-related reactor incidents, including the Chernobyl disaster in 1986 and the Fukushima accident in 2011, damaged public trust in nuclear power worldwide, leading to increased regulatory pressure in the United States to improve safety at existing and future power plants.

Nuclear weapons proliferation Most fission reactors require some enrichment of natural uranium. Enriching uranium from its natural state to being reactor-ready takes much more work than enriching it further, from being reactor-ready to bomb-ready. Thus, the widespread deployment of reactors raises concerns about the spread of nuclear technologies and material that can be used for weapons by states and nonstate actors such as terrorist groups.

Waste management All nuclear reactors produce radioactive waste that must be safely managed for tens of thousands of years. Long-term waste management depends on technologies for waste storage and locations where waste can be stored. The latter problem has not been solved in the United States despite many years of effort.

In part to address some of these issues, new reactor technologies—most prominently the small modular reactor (SMR)—have been developed. Advocates argue that SMRs are potentially safer than traditional nuclear reactors and offer advantages in terms of scalability, flexibility, and reduced waste production and capital costs. Several companies and research institutions are developing SMR designs, with some projects nearing deployment.

Nevertheless, SMRs remain an unproven technology in America, though the US Navy employs them in some ships and submarines (see sidebar on the Naval Nuclear Propulsion Program). The first company to pursue SMRs in the United States, NuScale Power, produced a design that was approved by the US Nuclear Regulatory Commission.[32] A plan to produce a working demonstration plant at Idaho National Laboratory recently fell through, however, after rising costs led the prospective customers to pull out.[33]

THE NAVAL NUCLEAR PROPULSION PROGRAM

While land-based small modular reactors (SMRs) remain an unproven technology in the United States, small nuclear reactors generating a few hundred megawatts have powered American submarines and aircraft carriers for decades. Of particular significance is that they are fueled with bomb-grade, highly enriched uranium, which limits their utility as civilian, land-based power plants. Nevertheless, the US Naval Nuclear Propulsion Program (NNPP) has considerable expertise in the design, manufacture, operation, decommissioning, and disposal of naval nuclear reactors, all of which are relevant to civilian SMR programs. Moreover, the NNPP maintains an extensive technical and industrial base through its laboratories, factories, shipyards, and training facilities to enable continued design, construction, and operation of these platforms.[a]

a. US Department of Energy and US Department of the Navy, "The United States Naval Nuclear Propulsion Program 2020," accessed August 13, 2023, https://www.energy.gov/sites/default/files/2021-07/2020%20United%20States%20Naval%20Nuclear%20Propulsion%20Program%20v3.pdf.

Analysts have also raised concerns about whether SMRs will be able to deliver their promised benefits. For example, SMRs tend to be less efficient than larger reactors and generate a greater volume of waste per unit of energy produced because more neutrons escape from the core, activating more of the surrounding material.[34] Also, a significant fraction of the cost of reactors, such as the cost of site preparation, is fixed and therefore independent of reactor size. Thus, such costs will be higher for the deployment of a number of SMRs than for the deployment of a single larger reactor with the same overall power output.

Addressing the issues described above is at least as much a question of policy as technology. Just as various public subsidies have been provided for certain kinds of renewable energy sources, such as residential solar panels, subsidies for various aspects of fission-based power generation could reduce costs. For example, the DOE supported a US company that produced 20 kilograms of HALEU in late 2023.[35]

As for fusion power, it continues to hold promise as a potentially limitless and inherently safe source of energy. In December 2022, the Lawrence Livermore National Laboratory demonstrated for the first time a better-than-breakeven fusion outcome: An experiment produced more energy than the laser energy used to "ignite" the deuterium-tritium fuel. Since then, the repeatability of this outcome has been demonstrated twice.[36] Nevertheless, significant scientific and engineering breakthroughs are still needed to make fusion generation of electricity commercially viable. Even the most optimistic private investors in fusion do not believe that commercial-scale fusion power plants are any closer than ten to fifteen years away, with large-scale deployments even further out.

Growth in Electrical Demand

One of the most significant trends in the energy arena is the anticipated growth in electricity demand over the coming decades—as much as 50 percent to 100 percent more than US demand in 2022.[37] This projected surge in demand will be driven by:

- The rise of AI and the consequent increase in energy-intensive data centers
- The electrification of transportation, with EVs becoming increasingly popular in many countries
- Efforts to boost domestic manufacturing, which increase industrial electricity consumption
- Government policies supporting green technologies and renewable energy sources, which further drive the shift toward electrification

Meeting this demand will require significant scale-up and investment in electricity generation and grid infrastructure.

Manufacturing and Supply Chains

With many corporate and government net-zero targets established for 2040 or 2050, the next decade or so will be a crucial period for the development of sustainable energy technology and policy. The United States has an unparalleled capacity for fundamental research in energy, much of it based in universities. However, in the equally important domain of manufacturing at scale, it is no longer the world leader. China and other countries that offer substantial government subsidies and support to businesses control most of the manufacturing, supply chain, and critical minerals for battery and solar cell production.

As these technologies will be directly tied to the energy security of the United States, promoting domestic production will be vitally important. Of concern is that China dominates much of the supply chain for battery materials, owing to its ability to supply such minerals at relatively low cost. This raises questions about supply chain security for key energy-technology materials. US government incentives have mobilized some investment in domestic materials production, but the United States will need to work with allies and partners to develop alternative sources for materials and processes.

Policy, Regulatory, and Legal Issues

ECONOMICS AND EMPLOYMENT

Energy and economics are deeply interconnected, with energy costs and efficiency directly impacting economic prosperity. The energy sector's significance means transitions in policy or sources can create economic winners and losers, at least in the short term. For instance, the shift to sustainable energy threatens well-paying fossil-fuel jobs. These transitions also involve cultural and other changes. For example, education plays a critical role in facilitating such transitions, particularly as manufacturing skills essential for technology scalability are declining in the US workforce. Balancing economic impacts, job transitions, and skill development is vital for a successful sustainable energy future.

ENVIRONMENTAL IMPACTS

Energy production, including sustainable methods, often generates harmful waste. Large-scale deployment of renewable technologies will result in significant end-of-life waste, such as old windmill blades and dead solar cells. Addressing these concerns proactively can minimize negative environmental impacts from decommissioned equipment. The next generation of energy technologies also presents an opportunity to incorporate recyclability into their design, moving toward a zero-waste economy.

Also relevant to environmental impact is the fact that many forms of sustainable energy require new acquisitions of land to build generating stations and storage facilities. For example, wind energy often requires the construction of many wind turbines on large tracts of land (see figure 10.5). Residents may support windmills in principle but then adopt a "not in my backyard" mentality. Early consultation and engagement with landowners and communities will be needed to build the social license to operate for land-intensive clean energy projects.

SUSTAINED FUNDING THROUGH THE VALLEY OF DEATH

The *valley of death* refers to the period after research has demonstrated the engineering feasibility of a particular innovation (a step beyond scientific feasibility) but before the innovation achieves adoption on a scale large enough to establish the viability of a business model using it.

In some fields, a significant gap exists between prototype development and market viability, requiring pilot projects that bridge the gap between academic research and development (R&D) and widespread use. Such projects address technical issues that emerge only at larger scales, beyond typical prototype development. However, venture capitalists often hesitate to invest in such projects, creating a funding gap for this critical commercialization stage. To address this valley of death for promising technologies, the Department of Energy's Loan Program offers debt financing for clean energy projects.[38] This initiative supports the transition from academic R&D to widespread commercial use, filling a crucial role in technology development.

SUSTAINED POLICY SUPPORT FOR EMISSIONS-FREE ENERGY INNOVATION

The federal government plays an important and large role in funding energy R&D. However, this research requires sustained support with a long-term vision. Commercial technologies such as solar cells and batteries stem from fundamental research in America that began decades ago, with these technologies only now reaching fruition. For many such innovations, large fluctuations in research funding and inconsistent support are damaging American research enterprises that depend on the ability to retain knowledgeable and experienced scientists and engineers to do the relevant work. As a result, key innovations such as solar energy and Li-ion batteries were invented in the United States and then commercialized overseas.

For the next generation of emission-free technologies, America must sustain a stable innovation

FIGURE 10.5 A cluster of wind turbines

Source: National Renewable Energy Laboratory / Dennis Schroeder

ecosystem over several decades. At stake is leadership in technologies such as fusion energy, next-generation nuclear reactors, carbon-neutral fuels, and long-duration energy storage. Success depends on achieving a consensus that combines vigorous academic and national laboratory innovation with effective public-private partnerships.

NOTES

1. The term *grid* always includes the means for transporting electrical energy from one location to another. Some sources include various means of producing electrical energy, while others do not; this report adopts the former convention.

2. Charlie King, "Top 10 Countries Using Renewable Energy," *Energy Digital*, November 8, 2023, https://energydigital.com/top10/top-10-countries-using-renewable-energy.

3. Laura Cozzi, Timur Gül, Thomas Spencer, et al., "Clean Energy Is Boosting Economic Growth," International Energy Agency, April 18, 2024, https://www.iea.org/commentaries/clean-energy-is-boosting-economic-growth.

4. International Renewable Energy Agency and International Labour Organization, *Renewable Energy and Jobs: Annual Review 2023* (Abu Dhabi, UAE, and Geneva, CH: 2023), https://www.irena.org/Publications/2023/Sep/Renewable-energy-and-jobs-Annual-review-2023.

5. Nicholas A. Mailloux, David W. Abel, Tracey Holloway, et al., "Nationwide and Regional PM2.5-Related Air Quality Health Benefits from the Removal of Energy-Related Emissions in the United States," *Geohealth* 6, no. 5 (May 2022), https://doi.org/10.1029/2022GH000603.

6. Charles Driscoll, Kathy Fallon Lambert, Peter Wilcoxen, et al., "An 80x30 Clean Electricity Standard: Carbon, Costs, and Health Benefits," Harvard Chan C-CHANGE and Clean Energy Futures, July 12, 2021, https://www.hsph.harvard.edu/c-change/wp-content/uploads/sites/2343/2021/07/CEF-80x30-7.15.21.pdf.

7. US Department of State and the US Executive Office of the President, "The Long-Term Strategy of the United States: Pathways to Net-Zero Greenhouse Gas Emissions by 2050," November 18, 2021, https://www.whitehouse.gov/wp-content/uploads/2021/10/us-long-term-strategy.pdf.

8. For wind, see Max Roser, "Why Did Renewables Become So Cheap So Fast?," *Our World in Data*, December 1, 2020, https://ourworldindata.org/cheap-renewables-growth. For solar, see International Renewable Energy Agency, *Renewable Power Generation Costs in 2022* (Abu Dhabi, UAE: 2023), https://www.irena.org/Publications/2023/Aug/Renewable-Power-Generation-Costs-in-2022.

9. An important milestone was achieved on April 29, 2024, when, for several hours, storage was the single largest source of energy to California's power grid.

10. David Shepardson, "Democrat Calls Only 7 EV-charging Stations Deployed Under US Program 'Pathetic,'" Reuters, June 5, 2024, https://www.reuters.com/world/us/democrat-calls-only-7-ev-charging-stations-deployed-under-us-program-pathetic-2024-06-05/.

11. US Department of Energy, "Innovation," accessed October 9, 2023, https://www.energy.gov/innovation.

12. American Clean Power, "New Report: Private Sector Investments in U.S. Clean Energy Sector Exceeded $270 Billion in Last Year," August 7, 2023, https://cleanpower.org/news/investing-in-america/.

13. Henry Fountain, "A Huge Underground Battery Is Coming to a Tiny Utah Town," *New York Times*, January 12, 2024, https://www.nytimes.com/2024/01/12/climate/green-hydrogen-climate-change.html.

14. Robert K. Emmett and Mark E. Roberts, "Recent Developments in Alternative Aqueous Redox Flow Batteries for Grid-Scale Energy Storage," *Journal of Power Sources* 506 (September 2021): 230087, https://doi.org/10.1016/j.jpowsour.2021.230087.

15. Prachi Patel, "NASA Battery Tech to Deliver for the Grid," *IEEE Spectrum*, September 24, 2023, https://spectrum.ieee.org/grid-scale-battery-storage-nickel-hydrogen.

16. Erik D. Spoerke, Howard Passell, Gabriel Cowles, et al., "Driving Zn-MnO_2 Grid-scale Batteries: A Roadmap to Cost-Effective Energy Storage," *MRS Energy & Sustainability* 9 (February 2022): 13–18, https://doi.org/10.1557/s43581-021-00018-4.

17. In reality, not all sources of electricity even in a sustainable energy world will be intermittent, but many will be. The required storage may be only forty-eight hours rather than seventy-two, and so on. But these figures set the required scale.

18. Aude Marjolin, "Lithium-Ion Battery Capacity to Grow Steadily to 2030," S&P Global, July 27, 2023, https://www.spglobal.com/marketintelligence/en/news-insights/research/lithium-ion-battery-capacity-to-grow-steadily-to-2030.

19. Wei Chen, Guodong Li, Allen Pei, et al., "A Manganese–Hydrogen Battery with Potential for Grid-Scale Energy Storage," *Nature Energy* 3 (April 2018): 428–35, https://doi.org/10.1038/s41560-018-0147-7.

20. US Department of Energy, "Hydrogen Storage," accessed August 30, 2023, https://www.energy.gov/eere/fuelcells/articles/hydrogen-storage-fact-sheet.

21. "Hydrogen," International Renewable Energy Agency, accessed September 10, 2024, https://www.irena.org/Energy-Transition/Technology/Hydrogen.

22. See, for example, Mitsubishi Heavy Industries, "Mitsubishi Shipbuilding Receives Order for Ammonia Fuel Supply System for Ammonia-Powered Marine Engine—Delivery Scheduled in 2025, Together with Ammonia Gas Abatement System," April 10, 2024, https://www.mhi.com/news/24041002.html.

23. Katie Lebling, Ankita Gangotra, Karl Hausker, et al., "7 Things to Know About Carbon Capture, Utilization and Sequestration," World Resources Institute, November 13, 2023, https://www.wri.org/insights/carbon-capture-technology.

24. Office of Clean Energy Demonstrations, "Regional Direct Air Capture Hubs Selected and Awarded Projects," US Department of Energy, accessed August 13, 2024, https://www.energy.gov/oced/regional-direct-air-capture-hubs-selected-and-awarded-projects.

25. National Academies of Sciences, Engineering, and Medicine, *Negative Emissions Technologies and Reliable Sequestrations: A Research Agenda* (Washington, DC: National Academies Press, 2019), https://doi.org/10.17226/25259.

26. Emilia Chojkiewicz, Umed Paliwal, Nikit Abhyankar, et al., "2035 The Report: 2035 4.0 Reconductoring," GridLab, April 2024, https://www.2035report.com/wp-content/uploads/2024/04/GridLab_2035-Reconductoring-Technical-Report.pdf.

27. US Energy Information Administration, "Frequently Asked Questions: What Is US Electricity Generation by Energy Source?," last modified February 29, 2024, https://www.eia.gov/tools/faqs/faq.php?id=427&t=3.

28. Natural uranium is 99.3 percent U-238 and about 0.7 percent U-235, a lighter and more fissionable isotope than U-238.

29. In August 2023, the United States was operating ninety-three reactors for commercial power production with an average age of about forty-two years (US Energy Information Administration, "Nuclear Explained: US Nuclear Industry," last modified August 24, 2023, https://www.eia.gov/energyexplained/nuclear/us-nuclear-industry.php). The average productive lifetime of early reactors is about thirty years, though that is extendable through refurbishment and for newer reactors between forty and sixty years (World Nuclear Association, "Decommissioning Nuclear Facilities," last modified May 3, 2022, https://world-nuclear.org/information-library/nuclear-fuel-cycle/nuclear-waste/decommissioning-nuclear-facilities). Taken together, these statements suggest that a significant fraction of all currently operating reactors in the United States will need to be replaced by 2050.

30. Bureau of International Security and Nonproliferation, "Declaration to Triple Nuclear Energy," US Department of State, December 2, 2023, https://www.state.gov/declaration-to-triple-nuclear-energy/.

31. Procuring nuclear-heavy equipment presents several hurdles due to the specialized nature of the components. One of the primary difficulties lies in the limited number of manufacturers capable of producing these highly specialized parts, such as reactor pressure vessels, steam generators, and control rod drive mechanisms. The manufacturing process is time-consuming, often requiring years to go from raw material to finished component. Advanced forging, precision machining, and complex welding each require specialized expertise and facilities. Furthermore, quality control is paramount, given the consequence of defects. These factors aggregate to drive high costs and long lead times.

32. Office of Nuclear Energy, "NRC Certifies First US Small Modular Reactor Design," US Department of Energy, January 20, 2023, https://www.energy.gov/ne/articles/nrc-certifies-first-us-small-modular-reactor-design.

33. Adrian Cho, "Deal to Build Pint-Size Nuclear Reactors Is Canceled," *Science* 382, no. 6672 (November 2023): 749–50, https://doi.org/10.1126/science.adm9503.

34. Lindsay M. Krall, Allison M. Macfarlane, and Rodney C. Ewing, "Nuclear Waste from Small Modular Reactors," *Proceedings of the National Academy of Sciences* 119, no. 23 (May 2022), https://doi.org/10.1073/pnas.2111833119.

35. Office of Nuclear Energy, "Centrus Produces Nation's First Amounts of HALEU," US Department of Energy, November 7, 2023, https://www.energy.gov/ne/articles/centrus-produces-nations-first-amounts-haleu.

36. Benny Evangelista, ed., "The Age of Ignition," Lawrence Livermore National Laboratory, December 5, 2022, https://lasers.llnl.gov/sites/lasers/files/2023-06/Age_of_Ignition.pdf.

37. US Energy Information Administration, "International Energy Outlook 2023," US Department of Energy, October 11, 2023, https://www.eia.gov/outlooks/ieo/.

38. US Department of Energy, "Overview," Loan Programs Office, accessed August 13, 2023, https://www.energy.gov/lpo/overview.

STANFORD EXPERT CONTRIBUTORS

Dr. Sally Benson
SETR Faculty Council, Precourt Family Professor, and Professor of Energy Science Engineering

Dr. Inês Azevedo
Associate Professor of Energy Science Engineering and, by courtesy, of Civil and Environmental Engineering

Dr. Simona Onori
Associate Professor of Energy Science Engineering

Dr. Karan Bhuwalka
SETR Fellow and Research Engineer in Energy Science and Engineering

Alexander Lubkin
MBA Candidate at the Stanford Graduate School of Business

11

CROSSCUTTING THEMES AND COMMONALITIES

None of the individual technology areas covered in chapters 1 through 10 operates in a vacuum. It is crucial that policymakers consider broader, crosscutting themes that influence how technology progresses over time as well as the key common drivers that can accelerate or hinder progress. To give just one example, synergies between different technologies mean that a policy initiative taken in one domain may have large positive or negative impacts in other technological areas—and policymaking will improve if these potential cascade effects can be considered at the outset.

Crosscutting themes can sometimes get overlooked in the desire to take action to advance an emerging technology. By devoting an entire chapter to them, we want to underline that there are important similarities in how people and institutions make progress and that, when crafting policy for individual domains, it is essential to take a holistic view of the emerging tech landscape and the factors affecting it.

This chapter organizes these important crosscutting themes into two broad categories. The first category includes key observations about how technological development generally unfolds, including in ways that may be surprising. These observations should be regarded as important considerations in the process of formulating technology policy to promote innovation. The second category is inspired in part by the experiences of the research faculty who contributed to this report and includes major enablers of innovation. Notably, enablers come in many varieties, including technical issues, human capital, the structure and strength of innovation funding, and cognitive blind spots that can lead policymakers astray and increase the likelihood of national technological surprise.

CATEGORY 1: KEY OBSERVATIONS ABOUT HOW TECHNOLOGIES EVOLVE OVER TIME

- The Goldilocks Challenge: Moving Too Quickly, Moving Too Slowly
- Increasing Access to New Technologies Worldwide
- Large and Growing Synergies Between Different Technologies
- Nonlinear Paths from Research to Useful Application
- Punctuated Technological Progress
- Nontechnological Influences on Technological Innovation
- The Changing Role of Government in Technological Innovation
- The Relationship of Political Regime Type to Technological Progress

CATEGORY 2: COMMON INNOVATION ENABLERS AND INHIBITORS

- The Central Importance of Ideas and Human Talent in Science and Technology
- Frontier Bias in Policymaking
- Optimism, Pessimism, and Realism in Technology Policy
- Universities and Technology Innovation
- The Structure of Research and Development Funding
- Cybersecurity

Key Observations About How Technologies Evolve over Time

The Goldilocks Challenge: Moving Too Quickly, Moving Too Slowly

Takeaway ***Innovation that emerges too fast threatens the legitimate interests of those who might be negatively affected, while innovation that moves too slowly increases the likelihood that a nation will lose first-mover advantages.***

Innovation typically brings two types of benefits: enhancing or improving existing processes and enabling entirely new functions that solve problems people did not even know they had.

Technological progress also brings risks. Chief among them are the risks of moving too fast or too slowly. Innovation that emerges too fast threatens to disrupt the often delicate balance that has been established among many national, organizational, and personal interests. A push to deploy new capabilities may give short shrift to issues such as safety, security, employment, values, ethics, societal impact, and geopolitics.

Examples abound. Genetically modified organisms (GMOs) have lost favor in much of Europe, largely due to safety concerns. The Concorde supersonic passenger airliner stopped flying due mostly to concerns about noise and its high operating costs. As noted in this report, concerns about the downside societal impacts of artificial intelligence (AI) are prompting calls for increased regulation of the technology. Similar concerns about the potential weaponization of synthetic biology inhibit a full-throated societal endorsement of a biotechnology-enabled future.

On the other hand, innovation that is too slow increases the likelihood that a nation will lose the technical, economic, and national security advantages that often accrue to first movers in a field. Such concerns are apparent in reports asserting that the United States is falling behind China in the development of key technologies essential for national security, such as AI.[1] Similar worries have been expressed regarding America's competition with China in technology domains considered critical to economic security.[2]

To fully realize the benefits of innovation, policy measures are often necessary to address both sets of challenges effectively.

Increasing Access to New Technologies Worldwide

Takeaway ***National monopolies on technology are increasingly difficult to maintain. Even innovations that are solely American born (an increasingly rare occurrence) are unlikely to remain in the exclusive control of American actors for long periods.***

Technologies such as synthetic biology, robotics, space exploration, and blockchain often spread from wealthy nations and large corporations to less affluent countries, smaller companies, and individual actors. Innovations that are American born are unlikely to remain in the exclusive control of American actors for long. The dissemination of many of these technologies is, in part, driven by the long-term trend of decreasing technology costs, which make these advancements accessible to a broader array of players despite efforts, such as export controls, to delay this spread.

Several key implications arise from this trend:

- Winning isn't winning anymore. The old model of achieving lasting national technological dominance is being replaced by a paradigm of continuous competition where technological advantages are not sustained for long periods.
- Greater policy complexity results from more actors. Both state and nonstate actors gain new tools to challenge US interests.
- Technological advantages are narrowing, even on the frontier. Although the United States may possess the most technologically advanced capabilities, other actors with less sophisticated versions of them can eliminate monopolies and narrow the relative advantages the United States previously had.
- There are more actors with different ethical thresholds, constraints, and perspectives. Actors with fewer bureaucratic and ethical constraints may exploit and adapt technology faster and more effectively than those with more stringent regulations.

To be sure, there are exceptions to this trend of technological diffusion. One of them is the first appearance of an emerging technology. At such a point in time, the diffusion process has not yet begun—at least not in full force—and it may indeed be that the technology in question will be characterized by the dominance of a few key actors. This is true in AI today, where a small number of private-sector actors clearly dominate the creation of large language models (LLMs). A second exception is when scale is a critical aspect of widespread innovation—actors without access to the natural resources, such as rare-earth metals, or the financial capital to support large-scale deployments are less likely to be able to take advantage of the technology.

It may be possible to extend periods of American monopoly on certain technologies, but these periods cannot be prolonged indefinitely. Extensions can help to buy time for US policymakers to better anticipate a world of technological diffusion. But all too often, buying time becomes an end unto itself, and actions to craft a better policy—such as targeted immigration reform to create a "brain gain" for American universities and companies to draw the world's best talent—are not taken.

Large and Growing Synergies Between Different Technologies

Takeaway ***The synergies between different technologies are large and growing, as advances in one technology often support advances in other technologies.***

The synergies between different technologies are both significant and expanding, as advances in one field often enhance progress in others (see sidebar on examples of synergy among emerging technologies). This point is more obvious when a field such as AI or materials science is seen as a foundational technology that impacts a variety of application domains. For example, this report has discussed how AI has facilitated innovations in battery technology and in protein folding. But what is less obvious is that AI itself has benefited greatly from advances in semiconductor technology, which has itself benefited from developments in materials science.

EXAMPLES OF SYNERGY AMONG EMERGING TECHNOLOGIES

- Artificial intelligence (AI) contributes to advances in synthetic biology by predicting the structures of various biomolecules, such as proteins, nucleic acids, and small molecules singly or joined together in various complexes.[a]
- AI helps to screen many candidate compounds to predict the ones most likely to exhibit desirable properties for materials science.[b]
- Materials science is central to the identification of new semiconductors that may be useful in developing more energy-efficient chips, which in turn can reduce the cost of training AI models.[c]
- Materials science is important in space research to create new materials for the construction of spacecraft and satellites[d] and to enable the development of neural probes that can send and receive electrical signals in neural tissue.[e]
- Energy technologies help to improve the performance of robotics and spacecraft.[f]
- Synthetic biology can build organisms that produce certain specialized materials.[g]
- Cheaper semiconductors have driven down the cost of DNA sequencing, which itself is a fundamental technology for synthetic biology.[h]

a. Josh Abramson, Jonas Adler, Jack Dunger, et al., "Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3," *Nature* 630 (May 2024): 493–500, https://doi.org/10.1038/s41586-024-07487-w.

b. This report, chapter 5 on materials science.

c. Po-Wen Chan and Balaji Chandrasekaran, "Materials Engineering: The True Hero of Energy-Efficient Chip Performance," *Applied Materials*, August 1, 2024, https://www.appliedmaterials.com/us/en/blog/blog-posts/the-true-hero-of-energy-efficient-chip-performance.html.

d. Frontiers in Space Technology, "Next Generation of Materials for Space Applications," Frontiers, accessed September 18, 2024, https://www.frontiersin.org/research-topics/55679/next-generation-of-materials-for-space-applications.

e. Yongli Qi, Seung-Kyun Kang, and Hui Fang, "Advanced Materials for Implantable Neuroelectronics," *MRS Bulletin* 48 (May 2023): 475–83, https://doi.org/10.1557/s43577-023-00540-5.

f. Anil D. Pathak, Shalakha Saha, Vikram Kishore Bharti, et al., "A Review on Battery Technology for Space Application," *Journal of Energy Storage* 61 (May 2023): 106792, https://doi.org/10.1016/j.est.2023.106792.

g. Orlando Burgos-Morales, M. Gueye, Laurie Lacombe, et al., "Synthetic Biology as Driver for the Biologization of Materials Sciences," *Materials Today Bio* 11 (May 2021): 100115, https://doi.org/10.1016/j.mtbio.2021.100115.

h. James M. Heather and Benjamin Chain, "The Sequence of Sequencers: The History of Sequencing DNA," *Genomics* 107, no. 1 (January 2016): 1–8, https://doi.org/10.1016/j.ygeno.2015.11.003.

Nonlinear Paths from Research to Useful Application

Takeaway ***The traditional linear model of R&D, in which basic research leads to applied research, which then leads to development and prototyping, which finally leads to novel and useful products or services, is only one model for how societies obtain value for investments in technology innovation.***

The traditional linear model of research and development (R&D), which progresses from basic research to applied research, leading to development and then to marketable products, represents just one way societies derive value from technological investments. This model starts with basic research focused on fundamental scientific understanding without immediate application, progresses to applied research with specific problems in mind, and moves through development to create prototypes and proofs of concept.

However, some argue that this linear approach is unrepresentative of how scientific progress actually occurs. Other models are less linear in nature, acknowledging and extolling the need for feedback between the various activities. For example, some challenging problems require a deeper fundamental scientific understanding, known as use-inspired basic research.

Research in AI on the hallucinations of LLMs fits into this category. The mechanisms that lead LLMs to generate entirely false statements known as hallucinations are not well understood today, and research continues on this topic. However, despite the fact that these errors occur frequently, LLMs continue to be useful in many walks of life.

In other cases, technology convergence can have a big impact on synergy and innovation. Here, convergence means that several key technologies each advance to the point that they are useful for a specific application, at which stage a useful innovation that draws on all of them can be produced or developed. For example, deep knowledge of drug delivery mechanisms is crucial before chemical compounds with strong anticancer properties can be turned into effective cancer treatment medications. Electric cars are another example: They were made possible by the convergence of advances in battery technology, lightweight materials, sensors, and computing power. A final example comes from chapter 6 on neuroscience in this report, which discusses how effective neurological interventions depend not only on a fundamental theoretical understanding of brain function, but also on the development of neural probes that can be implanted into the brain without causing serious damage to brain tissue.

Punctuated Technological Progress

Takeaway ***Technology often progresses in fits and starts, with long periods of incremental results followed by sudden breakthroughs. As a result, the speed of change is hard even for leading researchers to anticipate.***

Technological progress generally exhibits a variety of patterns. Progress in semiconductors has been fairly predictable historically, progressing consistently with Moore's law, which predicts a steady exponential decrease in cost over time. But, as noted in chapter 8 on semiconductors, this steady decline is coming to an end, if it hasn't expired already. Solar cells and light-emitting diode (LED) lighting have followed similar cost reduction curves, except that these cost decreases are usually represented as a function of manufacturing experience and expertise rather than time.[3]

But most other technologies have demonstrated much more uneven progress, characterized by extended phases of gradual development interrupted by sudden, transformative bursts of innovation. Sometimes, these bursts result from particular breakthroughs, such as the personal computer revolution of the 1980s and the emergence of the World

Wide Web in the early 1990s. Other times, they are due to the simultaneous availability and maturity of several key technologies that are required to make significant progress in some other technological domain—the convergence phenomenon mentioned earlier in this chapter.

Predicting future progress can be challenging and misleading due to the pattern of punctuated innovation. Even experts in a given field can be surprised by the rapidity of progress. For instance, Geoffrey Hinton, one of the pioneers in AI and a winner of the 2024 Nobel Prize in Physics for his application of tools and concepts from statistical mechanics to machine learning, recently expressed astonishment at the swift progress in AI and predicted that it will surpass human intelligence in the future. At the same time, chapter 1 on artificial intelligence notes that the transformative impact of AI on the US military has been slower than is required to ensure that the United States will retain the technological advantages it currently has over adversary forces.

Another example is nuclear fusion for sustainable energy. The previous edition of the *Stanford Emerging Technology Review* (*SETR*) described breakthroughs in nuclear fusion—two better-than-breakeven experiments conducted at the Lawrence Livermore National Laboratory. A flurry of excitement at the time—quite justified on scientific grounds—led to many overly optimistic news stories about inexhaustible supplies of energy whose tone conveyed a sense of rapid change in the field.[4] Since then, two more experiments have been performed, demonstrating repeatability. While repeatability is in itself an important scientific milestone, fusion is a long way from technical viability as a sustainable energy source, and most experts believe that prospects for commercialization remain decades away.

The punctuated nature of technological change suggests that expectations of regular and rapid change in most fields are generally not realized, despite what headlines in the news might lead one to believe.

Nontechnological Influences on Technological Innovation

Takeaway ***Technology applications in society require scientific proof of concept, engineering feasibility, economic viability, and societal acceptability.***

Scientific advancements are frequently highlighted in the news for their promise to address societal challenges and enhance our quality of life. However, there is often a large gap between a demonstration of scientific feasibility and the creation of a product or service based on the technology that is useful to society.

To be sure, scientific feasibility is a necessary prerequisite, but it may well be that other necessary forms of feasibility do not follow. First, after achieving scientific proof of concept, a given technology application must demonstrate engineering practicality. An example is the idea of a single-stage-to-orbit, chemically fueled spacecraft launched from Earth. With current rocket fuels and materials, it is generally believed that it is possible—though barely—to launch a spacecraft using a single rocket stage rather than multiple stages. What has not yet been demonstrated is a feasible engineering design that would reliably accomplish this task.

Economic viability and practicality come after engineering feasibility, and these involve considerations such as cost and ease of use. Early attempts to build supercomputers with superconducting components demonstrated technical success but faced practical challenges due to the need for liquid helium for cooling. This requirement made the computers difficult and costly to deploy, and the development of alternative technologies offering comparable performance at lower cost doomed the approach in the marketplace.

Next, there is the cost of manufacturing. It may prove too difficult to develop a manufacturing process to build a product or service based on the initial scientific proof of concept, or the materials used to demonstrate engineering feasibility may turn out to be too expensive or rare to support large-scale production. Cheaper alternatives may also become available, undermining the commercial viability of the original concept.

An illustration of this phenomenon can be found in the competition between lithium-ion (Li-ion) and sodium-ion batteries. Sodium-ion batteries have potential cost advantages over Li-ion ones because sodium is much more abundant than lithium. But Li-ion battery technology has a head start of a couple of decades that has driven down the cost of these batteries significantly. Thus, the economics of procurement today favor Li-ion in many of the most common applications—although any significant disruption of the lithium supply chain could make sodium-ion batteries more broadly competitive.

Societal acceptability matters as well. The psychology of individuals, as well as cultural practices and beliefs of a community or society, contribute to the adoption and use of any given technology application. For instance, in Europe, GMOs as food are highly controversial, and concerns over their safety have prevented the uptake of GMO foods that are consumed widely in the United States.

Finally, the road from scientific discovery to useful application is often rockier than expected, with would-be innovators finding that the realization of the benefits promised to investors and customers actually entails greater costs than planned, takes longer than anticipated, and delivers fewer capabilities than expected. They may stumble for other reasons, too, ranging from difficulty finding sufficient funding to advance innovations and challenges over environmental or other impacts. Moreover, risks associated with things such as ethics and equity, privacy, and increased challenges to health, safety, and security—all issues that could lead to an erosion of trust in a product or service—may become apparent only once it is being sold to customers.

The Changing Role of Government in Technological Innovation

Takeaway ***The US government is no longer the primary driver of technological innovation or funder of R&D.***

Plenty of technological innovations, including satellites, jet engines, and semiconductors, have their roots in US government financial support and advocacy. But in many fields today, the government is no longer the primary driver of innovation. Private companies have taken up much of the slack. These companies, however, may be under the jurisdiction of nations—or controlled by senior executives—whose interests are not aligned with those of the users of their services. The Starlink satellite communications network has been an essential part of Ukrainian battlefield communications; however, the CEO of Starlink curtailed Ukrainian access on a number of occasions in ways that affected Ukraine's battlefield strategy.[5] Such concerns are most serious when there is only one or just a small number of private-sector providers of the services in question.

The growing influence of private companies in critical technologies has led US officials to emphasize the need for closer public-private cooperation and government regulation. Even if the government cannot lead in innovation, it still plays a crucial role in funding R&D, promoting key innovations, setting standards, and forming coalitions domestically and internationally.

No better example of the growing influence of private companies in setting the R&D agenda can be found than in the current scene for the funding of AI research today. Whereas the federal government talks in terms of billions of dollars in federal support for AI research, the private sector is talking in terms of amounts ten to a hundred times larger. Similar trends seem to apply to biotechnology and synthetic biology research, though not quite as starkly. And, as chapter 9 on space discusses, services related to space are increasingly delivered by private companies.

It's also true that national priorities change with the evolution of the geopolitical environment. In the 1990s, there was widespread optimism about the triumph of liberal democracy and free market capitalism. Much of US economic policy was characterized by efforts to support free trade, accelerate globalization, and promote China's integration into the world economy as a way of facilitating its transition to more democratic rule.

During this time, the global manufacturing landscape for key technologies, particularly semiconductors, underwent significant shifts. Over three decades, the US share of global semiconductor production dropped from 37 percent to 12 percent, as noted in chapter 8 on semiconductors. Meanwhile, Asian manufacturers, especially in South Korea and Taiwan, emerged as major players, supported by government policies and regional demand shifts. Asia had become the dominant region for semiconductor production, laying the groundwork for the current global supply chain.

This shift in manufacturing capabilities, coupled with China's economic and military rise, is a key element of changes in the geopolitical environment and drives many Western concerns about technological dependencies in the twenty-first century. Accordingly, national policies that were seen as useful and appropriate in the environment of thirty years ago may need reassessment today.

The story of advanced chip fabrication is instructive. The most advanced chips made today require a laser technology called extreme ultraviolet (EUV) lithography, which uses tiny explosions of molten tin made at extreme speeds and bounced off the flattest mirrored surfaces in the world to produce an extremely short wavelength of light. Shortening light wavelengths is the key to shrinking the size and increasing the density of transistors on a chip, making it faster and more powerful.

Major breakthroughs in this laser technology were developed by researchers at Lawrence Livermore National Laboratory, Lawrence Berkeley National Laboratory, and Sandia National Laboratory in the 1990s, and intellectual property rights were owned by the US government but licensed under approval by Congress and the Department of Energy (DOE). The Dutch company ASML applied for a license, and at the time no objections were raised.

Today, ASML is the only company in the world that can manufacture and service the sophisticated machines using EUV technology, each of which costs about $200 million.[6] The future development of advanced semiconductor manufacturing equipment to shrink transistor dimensions even more—as well as balancing market access with the national security concerns of exporting to and servicing ASML equipment in China—are major geopolitical and economic concerns.[7]

The Relationship of Political Regime Type to Technological Progress

Takeaway ***Democracies provide greater freedom for scientific exploration, while authoritarian regimes can direct sustained funding and focus on technologies they believe are most important.***

Technological innovation occurs in both democracies and autocracies, but different regime types face different advantages and challenges. Democracies benefit from the rule of law, a free flow of ideas and people, and the freedom for individuals to pursue their own research interests. Perhaps most importantly, because failure in a democracy does not lead to persecution or necessarily result in professional ostracism, individuals are freer to experiment and explore. By contrast, authoritarian regimes are characterized by the rule of the state and dire consequences for failure, which can restrict the flow of ideas, force adherence to state-approved research areas, and prompt scientists to focus only on what are considered safe topics.

On the other hand, authoritarian regimes can direct sustained funding and attention to areas deemed crucial by the state more easily than democracies and maintain focus on these areas for extended periods, independent of short-term profit or political considerations. For example, it is widely accepted that Chinese AI efforts have access to the personal data of individuals on a far broader scale than those in the West, which generally has stronger privacy protections against government intrusion than China does.

In the United States, recent attempts to adopt more centralized technology-policy approaches—such as the Creating Helpful Incentives to Produce Semiconductors (CHIPS) and Science Act of 2022 supporting the semiconductor sector—aim to capture some benefits associated with authoritarian direction while preserving democratic values. Critics argue that such measures interfere with free markets and create inefficiencies by selecting winners and losers. Conversely, advocates believe these policies are necessary to counterbalance the advantages of authoritarian regimes. The challenge is to find a balance that aligns with American values while enhancing economic competitiveness.

Common Innovation Enablers and Inhibitors

The Central Importance of Ideas and Human Talent in Science and Technology

Takeaway ***Human talent plays a central role in generating the ideas for innovation; such talent can be found all over the world and cannot be manufactured at will.***

Scientific progress thrives on new ideas, which are generated daily by the most talented individuals worldwide. But human talent capable of creating ideas in science and technology cannot be generated on demand. Such talent must be acquired from foreign sources and/or nurtured domestically.

Foreign sources of talent make critical contributions to US science, technology, engineering, and mathematics (STEM). Although America remains the single most prominent contributor to global R&D, other nations—most notably China—are rapidly increasing their investments in this area. Geographic concentration of R&D expenditure continues its shift from the United States and Europe to East, Southeast, and South Asia.

This trend highlights the increasing importance of international collaboration. US researchers undoubtedly benefit from ideas developed abroad by reading scientific literature from other countries, but direct interactions with foreign researchers are often more valuable because they provide more comprehensive and expansive insights. Such interactions help American researchers acquire tacit knowledge that is not captured in published papers, including research directions that appeared promising but did not ultimately bear fruit. These interactions also offer a deeper understanding of foreign scientific progress. This point about the importance of tacit knowledge in scientific advancement has been made by many scholars[8] and was strongly expressed in a multitude of interviews with Stanford faculty working in the technology areas addressed in this report.

America's ability to attract and retain foreign talent is essential for maintaining its innovation edge, and domestic innovation is hindered when limitations are imposed on interactions with foreign scientists and their research. Skilled immigrants play a crucial role in American innovation, with immigrant college graduates receiving patents at twice the rate of native-born Americans.[9] More generally, a broader pool of people will yield higher quality talent than digging more deeply into an existing pool of people simply because the broader pool is more likely to have a greater number of individuals at the high end of the talent distribution.

However, US policies that discourage immigration can reduce the influx of skilled workers, impacting the country's capacity for innovation.[10] They also shift skilled talent and multinational R&D investment to other countries, including both strategic competitors such as China and close allies such as Canada.[11] Such a shift can also force US companies to relocate abroad due to worker shortages.[12] Finally, many academic researchers are immigrants on student visas. Without a clear route to permanent residence, the United States could lose key teaching and research talent in vital STEM domains.

Immigration policies affecting the labor force can make it harder to meet goals in industries like semiconductors, biotechnology, and sustainable energy. US workers still make up the majority of the US STEM workforce, although foreign-born talent accounts for an increasingly large fraction of it. Even if the trend toward greater foreign participation in STEM jobs continues—and especially if it were to reverse—strengthening the domestic pipeline of STEM workers is essential for several reasons.

First, a number of studies indicate a strong correlation between a nation's STEM education and economic growth and productivity.[13] Correlation is not causation, but the connection is unlikely to be accidental or spurious.

Second, other nations—including nations such as China and India that are presently important providers of US STEM talent—are investing more heavily in scientific R&D. Individuals who have previously chosen to work and study in the United States may well take advantage of opportunities at home in greater numbers. Foreign-born individuals working in the US STEM workforce may have family or personal ties in their nations of origin that tempt them to return. Those nations may also take steps that explicitly discourage their scientists and engineers from studying or working in the United States.

Third, a STEM worker educated in the United States is more likely to have personal and citizenship loyalties to the United States and can fairly be regarded as more likely to remain in the country.

Fourth, many security-sensitive jobs require US citizenship. In 2021, the US Department of Defense (DOD) noted that improving the capacity and resilience of the defense industrial base requires more workers trained in STEM,[14] also observing that the dearth of trained software engineers working on classified projects was in part because of the requirement that they must be US citizens.

In promoting a more robust domestic contribution to building STEM expertise in America, it is sobering to realize that the United States is facing a decades-long decline in K–12 (kindergarten to twelfth grade) STEM proficiency,[15] with standardized testing revealing declining scores in fourth- and eighth-grade mathematics.[16] While COVID-19 disruptions account for some decline,[17] the 2023 scores follow a twenty-year trend of diminishing STEM proficiency.[18] In 2022, American thirteen-year-olds had the lowest math and reading scores in decades, with math literacy declining across all demographics and regions.[19] A 2023 American College Testing (ACT) report found that 70 percent of high school seniors failed to meet college readiness benchmarks in math,[20] highlighting a critical challenge for the nation's economic and technological competitiveness.

Of particular concern is that at the top end, only 7 percent of American teens scored in the highest level of math proficiency as measured by the Program for International Student Assessment, a test to assess student ability to apply knowledge in real-world situations administered by the Organisation for Economic Co-operation and Development. This compares to 12 percent of Canadians and 41 percent of Singaporean teens scoring in the same top category.[21]

Adding to the challenge is a growing shortage of qualified STEM educators in the United States—about 28 percent of grade-seven to grade-twelve science teachers do not have a degree in the sciences, and about 40 percent of math teachers in high-poverty-area schools do not have a math degree.[22] Teachers in low-performing schools and in schools with higher concentrations of minority students are more likely to have less experience and report feeling underprepared for STEM instruction.[23]

Today, both paths to growing the requisite talent base to sustain and grow US innovation face serious and rising challenges. The global competition for talent means that the United States must adopt a more strategic approach to leveraging international expertise, as connections between American science and technology efforts and those of the rest of the world will accelerate the nation's progress in critical technology fields. To maintain and enhance its innovation capacity, the United States urgently needs to improve its own STEM education across all demographic groups, provide better pathways for skilled immigrants to remain in the United States, and invest more in human capital.

Concerns about foreign appropriation of American intellectual efforts are not without foundation. But using a meat axe to make widespread and blunt cuts in opportunities to collaborate with foreign scientists when a surgical scalpel is needed to curb only the issues that warrant serious concern is a sure way to undermine the effectiveness of US scientific endeavors.

Frontier Bias in Policymaking

Takeaway ***A policy bias toward science or technology at the frontiers of knowledge tends to overestimate the benefits accruing from such advances, at least in the short term. Many technologies with transformational potential are not necessarily on the technical frontier.***

A frontier bias is a tendency among analysts, commentators, and policymakers to focus on the significance of the newest and most recent innovations. Such a trend has been apparent even in the uptake of the inaugural *SETR* report—requests for briefings arising from its publication have most often focused on what's newest and most advanced in various fields.

But a frontier bias, while understandable, carries with it the risk of overlooking "old" technologies that can be used in novel and impactful ways. Innovation using proven and known technologies is a powerful way of advancing national and societal interests and by definition does not rely on scientific or technological breakthroughs.

One prominent example is technology being used in the present Russia-Ukraine war. For instance, the drones having the most effect on the battlefield area are a diverse mix of moderately sophisticated ones and off-the-shelf commercial drones. And in response to US trade sanctions on advanced semiconductors, Russia is making use of chips designed for home and commercial use to control its weapons.

Another example is the widespread use of the AK-47 automatic rifle. Unlike other popular guns, the AK-47 was deliberately designed to be low-tech—cheap, simple, and durable, as well as easy to manufacture and with few moving parts. It has since proliferated: Some seventy-five million of these guns are in operation today, and they have had an enormous impact on forces around the world,[24] most notably insurgent groups and terrorists.[25]

The story in chapter 10, on sustainable energy technologies, about a second life for electric vehicle (EV) batteries is also relevant here. As EVs become more widely used, the batteries powering them—still with significant capacity for power storage—will constitute a waste stream if other uses cannot be found for them. With specialized battery management systems tailored to their unique characteristics, these batteries can serve in stationary energy storage applications, such as acting as backup power sources for the grid. Their age will mean the batteries may not be at the cutting edge of battery technology when they are converted, but they will have significant capacity that would otherwise be thrown away.

A frontier bias, while understandable, carries with it the risk of overlooking "old" technologies that can be used in novel and impactful ways.

A second consequence of frontier bias is a misunderstanding of the difference between scientific or technological advances and adoption at scale. For example, in the couple of decades after the first generation of commercial nuclear power in October 1956, there was considerable optimism that further technological advancements in the field would bring about an era in which electrical energy was too cheap to meter. But as discussed in chapter 10, nuclear fission has not been widely adopted as a source of energy for a variety of technical, economic, and political reasons.

For an innovation to have significant societal impact, it needs to be broadly available and widely used. At one extreme, some innovations can be acquired on a small scale by individuals. The rapid spread of personal computers in the 1980s and of rooftop solar panels for home electricity generation are examples—people were willing to spend money out of their own pockets to derive the benefits of these innovations and the result was rapid uptake and adoption throughout society.

By contrast, advanced technology that requires a significant degree of centralized planning and/or funding for realization is likely to be adopted on much longer timescales. Nuclear energy requires the construction of nuclear reactors costing billions of dollars. State-of-the-art semiconductor plants come at the cost of tens of billions. Medicines for treating neurodegeneration are available only at the end of a very expensive drug approval and manufacturing process. Carbon capture and sequestration is too expensive to be widely adopted and is of marginal benefit for individuals, though it is of use to industrial facilities. For such innovations, it is unrealistic to expect rapid and widespread adoption throughout society.

Optimism, Pessimism, and Realism in Technology Policy

Takeaway ***Good public policy anticipates wide variations in perspectives on any given technology.***

This publication focuses on ten emerging technologies of significance. In putting together the latest edition of *SETR*, every faculty member interviewed from each of these technology areas was broadly optimistic about the societal and scientific value of work that is being conducted in their chosen domains.

This is hardly a surprise. If they were not optimistic about the value of work in those fields, why would they continue to work in them? Technological optimism—a belief that technological advances will continue to accrue and that such advances will solve important problems—is virtually a requirement for people to spend large parts of their lives working to invent or invest in a new technology.

But an optimistic view of technology cannot and should not be the only voice shaping technology policy. When everyone in a decision-making organization shares similar perspectives on technology—creating analytical blind spots and, potentially, groupthink (unwarranted conformity in beliefs)—the risks associated with innovation can be underestimated. It is therefore important to have voices in the decision-making process that anticipate and articulate potential downsides of a given technology. The people with these voices play a critical role in identifying possible negative consequences of new technologies—such as privacy concerns, ethical dilemmas, or societal disruptions—and in devising ways to mitigate those risks before technological lock-in makes it difficult or impossible to address their root causes. A set of individuals with different knowledge and backgrounds who are involved in the decision-making process will generate a more comprehensive understanding of potential risks and benefits and avoid damaging groupthink.

The optimists and advocates of a particular technology will almost always dismiss early attention to potential downside risks as being premature and likely to stifle or retard innovation. But such downsides will almost always exist when a new technology is adopted at a large scale. By actively searching for voices that can explain the potential downsides early in a technology's life cycle, policymakers can create a more balanced and forward-thinking policy environment that reduces risks while fostering innovation.

Another important issue when it comes to developing a framework for shaping policy is the double-edged nature of leapfrog developments in technology. For individual businesses, such breakout events can offer competitive advantages and open new markets. However, for nations, these rapid changes can pose strategic challenges when the scale of adoption reaches a sufficiently high level. At that point, countries that did not anticipate and/or prepare sufficiently may find themselves vulnerable to strategic surprises that could impact national security, economic stability, or social cohesion.

Finally, there needs to be a recognition that the process of creating effective technology policy can be slow and complex due to the need to mesh various perspectives together. Only by engaging in thorough deliberation and incorporating diverse viewpoints can policymakers shape an environment that is both well suited to the speed of technological progress and able to adapt to the inevitable uncertainties that accompany it. This approach ensures that public policy remains adaptable and capable of addressing both the ups and downs of technological advancements.

Universities and Technology Innovation

Takeaway ***US universities play a pivotal role in the innovation ecosystem that is increasingly at risk.***

The US R&D infrastructure is extensive, with significant contributions from both the private sector and federal government. Historically, private-sector research entities such as the former Bell Laboratories, IBM's Thomas J. Watson Research Center, and Xerox PARC have performed a substantial amount of foundational scientific research. However, such organizations now focus primarily on applied research tied to their commercial interests. This shift means that their research often targets immediate, practical applications rather than long-term, fundamental breakthroughs. Moreover, corporate R&D outputs tend to be proprietary, limiting their accessibility and broader impact.

The federal government also operates a large number of laboratories and Federally Funded Research and Development Centers (FFRDCs), including those run by the DOE, the DOD, and NASA. These mission-driven laboratories aim to address particularly difficult problems that go beyond the capacity of private industry or individual universities.

Universities play two unique and pivotal roles within the innovation ecosystem that are often underappreciated. First, they have the mission of pursuing high-risk research that may not pay off in commercial or societal applications for a long time, if ever.[26] (The sidebar on the long-term reach of university research offers some examples.) Unlike mission-driven federal labs, universities have a broader scope and breadth of research, and unlike the private sector, they conduct open, transparent research that promotes accountability, collaboration, and broad impact. This openness accelerates discovery by making study details, data, and results accessible to others. One significant data point is that more than 80 percent of the algorithms used today—not just in AI but in all kinds of information technology—originated from sources other than industrial research.[27] University openness magnifies educational and societal benefits by enabling other researchers to build on prior work, thus driving innovation forward.

THE LONG-TERM REACH OF UNIVERSITY RESEARCH

Research in number theory—a branch of pure mathematics—was undertaken for decades before it became foundational to modern cryptography.

In the 1960s, academic research on perceptrons sought to develop a computational basis for understanding the activity of the human brain. Although this line of research was abandoned after a decade or so, it ultimately gave rise to the work in AI on deep learning several decades later.

The term *mRNA vaccines* entered the public lexicon in 2021 when COVID-19 vaccines were released. Yet development of these vaccines was built on university research with a thirty-year history.[a]

Magnetic resonance imaging (MRI) was first discovered in university studies in the 1940s, but it took another three decades of research, much of it university based, for the first medical MRI imagers to emerge.

a. Elie Dolgin, "The Tangled History of mRNA Vaccines," *Nature*, October 22, 2021, https://www.nature.com/articles/d41586-021-02483-w.

Second, as educational institutions, universities play the central role in producing meaningful STEM expertise in the next generation. Any long-term plan for STEM leadership globally has to take into account how to sustain any advantages that the United States has—and US higher education in STEM is still the best in the world. This leadership is reinforced by the strength of America's university-based research enterprise: There is no better way to learn how to do state-of-the-art research in STEM than to actively participate in such work. By providing students with hands-on research experiences, access to cutting-edge facilities, and mentorship from leading experts, US universities create an environment where the next generation of STEM leaders can flourish.

Throughout history, government-supported university research has played a key role in technological advancements, from radar and proximity fuses during World War II to modern developments like AI and mRNA vaccines. It has generated knowledge whose exploitation creates new industries and jobs, spurs economic growth, and supports a high standard of living while also achieving national goals for defense, health, and energy.[28] It has also been a rich source of new ideas, particularly for the longer term, and universities are the primary source of graduates with advanced science and technology skills.

University research and development funding from all sources has grown significantly, reaching nearly $90 billion in 2021.[29] While private-sector investment in technology and university research has increased, it cannot replace federal funding, which supports R&D focused on national and public issues rather than commercial viability.[30] The US government remains uniquely capable of making large investments year after year in basic science at universities and national laboratories, which is essential for future applications. However, its share of academic R&D funding has declined over the past decade, standing at 55 percent of total support for academic R&D in 2021, the most recent year with available data.

As measured against GDP, funding trends are also negative. The fraction of GDP that goes to R&D could fairly be regarded as a seed corn investment in the future, yet federal R&D funding has gone from 1.86 percent of GDP in 1964 to just 0.66 percent of GDP in 2016.[31]

A constrained budget environment largely accounts for these negative funding trends. For example, the CHIPS and Science Act of 2022 authorized dramatically increased funding for basic research—about $53 billion—but Congress provided only $39 billion.[32] The United States still funds more basic research than China, but Chinese investment is rising much more rapidly and will likely overtake that of the United States within a decade.[33]

Moreover, despite their vital contributions, universities face challenges due to the blurring line between fundamental and export-controlled research, which complicates international collaboration in fields such as semiconductors, nanotechnology, AI, and neuroscience. For example, some researchers worry that fundamental research, which should be a less sensitive area, could now be considered export controlled and may shy away from foreign collaboration out of an abundance of caution. While well intended, these kinds of expanding restrictions may backfire in the long term, holding back US progress in key technological domains. Restrictions are not the only challenge; policy ambiguity is also harmful because it can discourage or deter collaboration with non-US researchers wishing to contribute to work in the United States.

All of these policy issues, widely recognized among the research community and apparent in interviews with Stanford faculty for this publication, underscore the urgent need for clarification and reform to advance research and promote effective international collaborations.

The Structure of Research and Development Funding

Takeaway ***Sustaining American innovation requires long-term government R&D investments with clear strategies and sustained priorities and not the wild swings from year to year that have become increasingly common.***

Budget is an obvious aspect of government funding for R&D, but three other aspects deserve at least as much attention. First, the government plays an important role in funding long-term precompetitive research that industry is not structured to support. Second, frequent shifts in funding levels, which are becoming increasingly common in government funding, undermine systematic R&D efforts and drive away scientific talent that opts to find employment elsewhere. Third, the so-called valley of death, a period occurring after demonstrating the engineering feasibility of an innovation but before it achieves large-scale adoption and commercial viability, is a significant problem.

When a new innovation is first offered to customers, its cost relative to what it is capable of can impact its success. High initial costs can deter the public from purchasing or using the innovation, potentially leading to a firm's commercial failure in the absence of external funding. However, as production volume increases, per-unit costs typically decrease due to the learning curve in manufacturing. This cost reduction is critical, especially in sectors like energy production where large-scale deployment offers significant societal benefits.

The problem is that researchers and young companies trying to get to this point must first find ways to scale their activities, and raising money to do this can be challenging. Research funding typically ceases once the feasibility of a technology has been demonstrated. If no alternative sources of money are found—or those that are available are not sufficient to get projects to critical scale—then those projects may have to stop or progress much more slowly. In some cases, innovations never scale beyond the initial stages, regardless of their technical sophistication or desirability.

For a firm to get through this valley of death, it must either secure investors who believe in the innovation's potential or attract enough customers to sustain operations. True commercial viability typically requires reducing per-unit costs to an affordable level for most customers. This can be particularly challenging for projects that require very large capital investments.

Bridge funding, which could come from government entities, banks, or other sources, may help to establish commercial viability, but an ongoing challenge is distinguishing between genuinely promising innovations and those that merely appear to be innovative but are not commercially viable. Firms failing to cross the valley of death could be acquired by foreign competitors—in particular, China—that have a greater willingness to invest in a technology that has not yet proven itself in the marketplace.

Focused research organizations (FROs) are a new funding model designed to bridge the valley of death by providing financial support to teams of scientists and engineers for rapid prototyping and testing. Convergent Research, a nonprofit established in 2021 to support FROs, received $50 million in philanthropic donations in March 2023 to start two new FROs.[34]

Cybersecurity

Takeaway ***Researchers working in highly competitive environments who neglect cybersecurity place their research progress at risk.***

Cybersecurity refers to technologies, processes, and policies that help to protect computer systems, networks, and the information contained therein from malicious activities undertaken by adversaries or unscrupulous competitors. Traditionally, cybersecurity has focused on protecting computer systems from unauthorized access and misuse, emphasizing the core principles of data confidentiality, integrity, and availability. Confidentiality ensures data privacy, preventing unauthorized disclosure. Integrity maintains data and program accuracy, guarding against unauthorized alterations. Availability ensures data and computing resources are accessible to authorized users, especially during critical times.

Initially, cybersecurity as a technical discipline focused on secure programming languages and robust software architectures, which created systems more resistant to threats like malware and advanced cyberattacks. The internet's growth and the proliferation of networked devices required security that also encompassed protecting the infrastructure that supports data transmission and storage.

Cybersecurity has continued to evolve, focusing on challenges from social engineering attacks and digital misinformation, recognizing that human interaction with computer systems is a threat vector. It has also focused on the risks of increasing information warfare and cyber-enabled information operations.[35] Further research addresses threats posed by the emergence of AI at both human and system levels.

As a national-level issue, cybersecurity policy measures are often associated with private-sector businesses and government. But cybersecurity is also a critical concern for R&D in academia and industry.

One major cybersecurity interest is ensuring the integrity of data. Data generated from scientific experiments are fundamental to research progress, and the deletion or destruction of data can severely hinder scientific advancement. More insidiously, if data are subtly altered, this can skew results in ways that are difficult to detect, potentially leading researchers down the wrong path and wasting valuable effort. Computer programs are also susceptible to such risks, as minor changes can go unnoticed for long periods of time and call into question the validity of previously collected or analyzed data.

A second interest in cybersecurity is protecting the confidentiality of work products, such as datasets and working papers. Unauthorized access to confidential datasets can breach agreements and compromise academic integrity, while premature disclosure of draft working papers can undermine claims of priority and reveal incomplete, inconsistent, or inaccurate information.

Computers managing data collection from laboratory instruments are also vulnerable to attacks that could disrupt data collection, corrupt data, and damage equipment. Such attacks can have severe repercussions for research continuity.

Technical safeguards exist to address these cybersecurity challenges, but maintaining them in academic settings requires substantial management effort. This effort can clash with the informal, collegial, and flexible cultures common in labs, where rigorous security practices may be perceived as disruptive.

Cybersecurity issues also arise with the development and use of the foundational models of AI. These models are generally trained on enormous amounts of data. A malicious actor could attack a large language model (LLM) trained on textual data by posting corrupted data online that is subsequently used to train the model. Publicly accessible LLMs can also be attacked by users seeking to circumvent the models' safety guidelines to obtain outputs that they would otherwise not disclose, such as how to make a deadly poison.

Another cybersecurity threat involves the selective targeting of personnel involved in key research projects. It is quite possible that these individuals may be harassed via cyberspace. This activity may include compromising personal finances or threatening researchers' families, which can be distressing for both the scientists and their family members. Another tactic involves questioning an investigator's professional ethics on social media and in other online forums, which can damage their reputation and productivity.

NOTES

1. See, for example, Josh Luckenbaugh, "Just In: U.S. Falling Behind China in Critical Tech Race, Report Finds," *National Defense*, July 17, 2023, https://www.nationaldefensemagazine.org/articles/2023/7/17/us-falling-behind-china-in-critical-tech-race-report-finds; Jeremy Neufeld, "STEM Immigration Is Critical to American National Security," Institute for Progress, March 30, 2022, https://ifp.org/stem-immigration-is-critical-to-american-national-security/.

2. Craig Cohen and Alexander Kisling, eds., "Part II: Winning the Economic and Tech Race," in *2024 Global Forecast: A World Dividing* (Washington, DC: Center for Strategic and International Studies, 2024), https://csis-website-prod.s3.amazonaws.com/s3fs-public/2024-01/240130_GlobalForecast_2024_WinningEconomic_TechRace.pdf.

3. Jean-Lou Chameau, William F. Ballhaus, and Herbert Lin, eds., *Emerging and Readily Available Technologies and National Security: A Framework for Addressing Ethical, Legal, and Societal Issues* (Washington, DC: National Academies Press, 2014), https://doi.org/10.17226/18512.

4. On solar panels, see François Lafond, Aimee Gotway Bailey, Jan David Bakker, et al., "How Well Do Experience Curves Predict Technological Progress? A Method for Making Distributional Forecasts," *Technological Forecasting and Social Change* 128 (March 2018): 104–117, https://doi.org/10.1016/j.techfore.2017.11.001. On LEDs, see Brian F. Gerke, Allison T. Ngo, Andrea L. Alstone, et al., "The Evolving Price of Household LED Lamps: Recent Trends and Historical Comparisons for the US Market," Office of Science and Technical Information, US Department of Energy, October 14, 2014, https://doi.org/10.2172/1163956.

5. As one illustration, on December 4, 2023, John Kerry, Special US Envoy for Climate, said to the COP28 climate summit in Dubai, "There is potential in fusion to revolutionize our world" as he launched an international plan to boost nuclear fusion. Valerie Volcovici, "US Envoy Kerry Launches International Nuclear Fusion Plan at COP28," Reuters, December 6, 2023, https://www.reuters.com/business/energy/us-envoy-kerry-launches-international-nuclear-fusion-plan-cop28-2023-12-05/.

6. Adam Satariano, Scott Reinhard, Cade Metz, et al., "Elon Musk's Unmatched Power in the Stars," *New York Times*, July 28, 2023, https://www.nytimes.com/interactive/2023/07/28/business/starlink.html.

7. Mat Honan and James O'Donnell, "How ASML Took Over the Chipmaking Chessboard," *MIT Technology Review*, April 1, 2024, https://www.technologyreview.com/2024/04/01/1090393/how-asml-took-over-the-chipmaking-chessboard/.

8. Arjun Kharpal, "Netherlands Takes on U.S. Export Controls, Controlling Shipments of Some ASML Machines," CNBC, September 6, 2024, https://www.cnbc.com/2024/09/06/netherlands-expands-export-curbs-on-advanced-chip-tools.html.

9. Michael Polyani, *Personal Knowledge* (Chicago, IL: University of Chicago Press, 1962); Thomas S. Kuhn, *Structure of Scientific Revolutions*, 2nd ed. (Chicago, IL: University of Chicago Press, 1970); for more recent discussions, see, for example, Vincent-Wayne Mitchell, William S. Harvey, and Geoffrey Wood, "Where Does All the 'Know How' Go? The Role of Tacit Knowledge in Research Impact," *Higher Education Research & Development* 41, no. 5 (2022): 1664–78, https://doi.org/10.1080/07294360.2021.1937066; Tim Thornton, "Tacit Knowledge as the Unifying Factor in Evidence Based Medicine and Clinical Judgement," *Philosophy, Ethics, and Humanities in Medicine* 1, no. 2 (March 2006) https://doi.org/10.1186/1747-5341-1-2.

10. Jennifer Hunt and Marjolaine Gauthier-Loiselle, "How Much Does Immigration Boost Innovation?," *American Economic Journal: Macroeconomics* 2, no. 2 (April 2010): 31–56, https://doi.org/10.1257/mac.2.2.31.

11. Dany Bahar, Prithwiraj Choudhury, Sara Signorelli, et al., *Talent Flows and the Geography of Knowledge Production: Causal Evidence from Multinational Firms*, Harvard Business School, Technology and Operations Management Unit, Working Paper no. 22-047, last modified December 23, 2022, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4005693.

12. Immigration, Refugees, and Citizenship Canada, "Canada's Tech Talent Strategy," Government of Canada, last modified June 27, 2023, https://www.canada.ca/en/immigration-refugees-citizenship/news/2023/06/canadas-tech-talent-strategy.html.

13. Britta Glennon and David R. Dollar, "What's Behind the Globalization of R&D?," April 26, 2021, in *Dollar & Sense*, podcast produced by Brookings Institution, MP3 Audio, 25:02, https://www.brookings.edu/articles/whats-behind-the-globalization-of-rd.

14. Maja Bacovic, Zivko Andrijasevic, and Bojan Pejovic, "STEM Education and Growth in Europe," *Journal of the Knowledge Economy* 13 (2022): 2348–71, https://doi.org/10.1007/s13132-021-00817-7.

15. US Department of Defense, *Industrial Capabilities Report to Congress: 2020 Annual Report*, January 2021, https://www.businessdefense.gov/docs/resources/USA002573-20_ICR_2020_Web.pdf.

16. Gabrielle Athanasia and Jillian Cota, "The U.S. Should Strengthen STEM Education to Remain Globally Competitive," Center for Strategic & International Studies, April 1, 2022, https://www.csis.org/blogs/perspectives-innovation/us-should-strengthen-stem-education-remain-globally-competitive; Darrel M. West, "Improving Workforce Development and STEM Education to Preserve America's Innovation Edge," Brookings Institution, July 26, 2023, https://www.brookings.edu/articles/improving-workforce-development-and-stem-education-to-preserve-americas-innovation-edge/; Julia Yoon, "Innovation Lightbulb: Strengthening K-12 STEM Education for a Robust U.S. Technology Workforce," Center for Strategic & International Studies, June 28, 2024, https://www.csis.org/analysis/innovation-lightbulb-strengthening-k-12-stem-education-robust-us-technology-workforce.

17. National Science Board, "Talent is the Treasure: Who Are We Leaving on the Bench?," March 2024, https://www.nsf.gov/nsb/publications/2024/2024_policy_brief.pdf.

18. National Science Board, "Talent is the Treasure"; Yoon, "Innovation Lightbulb."

19. Athanasia and Cota, "The U.S. Should Strengthen STEM Education"; Steven Deitz and Christina Freyman, "The State of U.S. Science and Engineering 2024," National Science Board, National Science Foundation, March 13, 2024, https://ncses.nsf.gov/pubs/nsb20243/; Yoon, "Innovation Lightbulb"; National Science Board, "Talent is the Treasure."

20. Dana Goldstein, "What the New, Low Test Scores for 13-Year-Olds Say About U.S. Education Now," *New York Times*, June 21, 2023, https://www.nytimes.com/2023/06/21/us/naep-test-results-education.html.

21. ACT Newsroom & Blog, "Fewer High School Seniors Ready for College as ACT Scores Continue to Decline," ACT, October 10, 2023, https://leadershipblog.act.org/2023/10/act-scores-decline.html.

22. Program for International Student Assessment, "PISA 2022 Mathematics Literacy Results," Institute of Education Sciences, National Center for Education Statistics, accessed September 18, 2024, https://nces.ed.gov/surveys/pisa/pisa2022/mathematics/international-comparisons/.

23. Athanasia and Cota, "The U.S. Should Strengthen STEM Education"; Dan Martin, "Understanding the Gaps in the US STEM Labor Market," Oxford Economics, October 5, 2023, https://www.oxfordeconomics.com/resource/understanding-the-gaps-in-the-us-stem-labor-market/.

24. Martin, "Understanding the Gaps."

25. Stephan Wilkinson, "How the AK-47 Became the 'Weapon of the Century,'" *Military Times*, December 12, 2017, https://www.militarytimes.com/off-duty/gearscout/2017/12/12/how-the-ak-47-became-the-weapon-of-the-century/.

26. Audrey Cronin, *Power to the People: How Open Technological Innovation Is Arming Tomorrow's Terrorists* (New York: Oxford University Press, 2020).

27. National Academies of Sciences, Engineering, and Medicine, *Information Technology Innovation: Resurgence, Confluence, and Continuing Impact* (Washington, DC: National Academies Press, 2020), https://doi.org/10.17226/25961.

28. Neil C. Thompson, Shuning Ge, and Yash M. Sherry, "Building the Algorithm Commons: Who Discovered the Algorithms That Underpin Computing in the Modern Enterprise?," *Global Strategy Journal* 11, no. 1 (2021): 17–33, https://doi.org/10.1002/gsj.1393.

29. National Research Council, *Research Universities and the Future of America: Ten Breakthrough Actions Vital to Our Nation's Prosperity and Security* (Washington, DC: National Academies Press, 2012), https://doi.org/10.17226/13396.

30. Jaquelina C. Falkenheim and Jeffrey M. Alexander, "Funding Sources of Academic R&D," in *Academic Research and Development* (Alexandria, VA: Science and Engineering Indicators, National Science Board and National Science Foundation, 2023), https://ncses.nsf.gov/pubs/nsb202326/funding-sources-of-academic-r-d.

31 Council on Foreign Relations, *Innovation and National Security: Keeping Our Edge*, Independent Task Force Report No. 77, James Manyika and William H. McRaven, chairs (New York: Council on Foreign Relations, Inc., 2019), 21, https://www.cfr.org/report/keeping-our-edge/pdf/TFR_Innovation_Strategy.pdf.

32. Council on Foreign Relations, *Innovation and National Security*, 10.

33. James Pethokoukis, "Broken Promises: CHIPS Act Funding for Science Research Falls Short," American Enterprise Institute, March 13, 2024, https://www.aei.org/articles/broken-promises-chips-act-funding-for-science-research-falls-short/.

34. OECD, "OECD Data Explorer: Gross Domestic Expenditure on R&D by Sector of Performance and Type of R&D ," OECD.org, 2024, https://data-viewer.oecd.org/?chartId=200b7feb-69b3-4502-8839-2f415a04ae97. Analysis was conducted by comparing two methods of measuring annual increases of US and Chinese spending on basic research and projecting rates into future.

35. Alex Knapp, "Why Billionaires Ken Griffing and Eric Schmidt Are Spending $50 Million on a New Kind of Scientific Research," *Forbes*, March 17, 2023, https://www.forbes.com/sites/alexknapp/2023/03/17/why-billionaires-ken-griffin-and-eric-schmidt-are-spending-50-million-on-a-new-kind-of-scientific-research.

36. Herbert Lin, "The Existential Threat from Cyber-Enabled Information Warfare," *Bulletin of the Atomic Scientists* 75 (4): 187–96, doi:10.1080/00963402.2019.1629574, 2019.

STANFORD EXPERT CONTRIBUTOR

Dr. Herbert S. Lin

SETR Director and Editor in Chief, Hank J. Holland Fellow in Cyber Policy and Security at the Hoover Institution

12

TECHNOLOGY APPLICATIONS BY POLICY AREA

This chapter explores applications from each technology field described in the report as they relate to five important policy themes: economic growth, national security, environmental and energy sustainability, health and medicine, and civil society. For each area, we extract from the technology discussions of chapters 1 through 10 applications or consequences that speak to it. Readers are invited to refer to the relevant technology chapter for more information about an application or consequence mentioned as well as for discussions of technical and other terms that may be unfamiliar.

Economic Growth

Artificial intelligence (AI) AI may significantly boost productivity across many sectors of the economy. Large language models such as ChatGPT have already demonstrated how they can be used in a variety of diverse fields, including law, customer support, computer programming, and journalism. Generative AI, a form of AI that creates new text, images, and other content, is expected to raise global GDP by $7 trillion and lift productivity growth by 1.5 percent over a ten-year period, if adopted widely.

Biotechnology and synthetic biology Biotechnology is poised to emerge as a general-purpose technology that can be applied broadly, with the capacity to revolutionize areas such as healthcare and manufacturing. Biological processes could ultimately produce as much as 60 percent of the physical inputs to the global economy. Already, biotechnology and synthetic biology are enablers for advances in medicine and healthcare (e.g., vaccines and cancer treatments), agriculture (e.g., drought-resistant crops), food (e.g., nutritionally enriched vegetables), and energy production (e.g., biofuels). Potential applications also include biotic semiconductors, magnets, fiber optics, and data storage.

Cryptography Blockchain technologies can effectively provide provenance in supply chains as well as personal identity management that curbs fraud and identity theft, leading to more secure and efficient transactions. Blockchain technology also underpins cryptocurrencies. A US central bank digital currency (CBDC), a form of digital currency that does not necessarily use blockchains, could help reduce inefficiencies in US deposit markets, promoting broader participation in the financial system.

Lasers Lasers are a key component for a variety of economically significant applications across manufacturing, communications, high-end chip production, defense, and medicine. For example, in precision manufacturing, lasers play an important role in cutting and shaping materials. Another application is in long-distance fiber-optic communications, where they are the providers of the light pulses that carry very high volumes of data.

Materials science Lighter and stronger materials will increase the energy efficiency of vehicles used to transport people and cargo. New semiconductor materials enable new types of chips and other information processing hardware. Technological innovations are also offering new ways to produce low-carbon steel and cement.

Neuroscience Interventions for those with neural disorders include pharmaceuticals that curb, treat, or reverse neurodegenerative conditions; diagnostics to identify early onset of such conditions; and rehabilitation therapies that help those suffering from them engage in the activities of daily living. By helping to address neurogenerative diseases more effectively, research in the field could allow people to remain in the workforce longer and be more productive, as well as reduce the burden on caregivers, who often need to take time off work to look after relatives and friends.

Robotics Robots are used widely today, including in manufacturing; on-demand delivery services; surgery; science and exploration; food production; disaster assistance; security and military services; and transportation. Innovations in robotics have enormous potential to increase productivity in many fields and perhaps to create new types of jobs. But robots involving physical labor and presence may also eliminate some jobs and change others, creating the need for retraining and other measures to address short-term impacts.

Semiconductors Semiconductors are an enabling technology for any application that can be improved through the use of information. They provide the computing capabilities that many sectors of the economy rely on. As such, they are key drivers of economic activity and growth. However, reductions in the cost of semiconductors and increases in processing power are likely to become less frequent or regular in the future—and predictions about economic growth in the years ahead attributable to improvements in semiconductor technology may prove to be overly optimistic.

Space Space activities play critical roles in our daily lives and the economy, from enabling global navigation systems to providing precise time information for financial transactions. Expanding commercial activities are expected to drive high growth in the space sector. In the future, space activities could become even bigger drivers of economic growth on Earth, through things such as asteroid mining and space-based power production.

Sustainable energy technologies In 2023, clean energy accounted for 10 percent of global GDP growth. Doubling the share of renewables by 2030 would increase global GDP by over $1 trillion in addition to creating 24 million new jobs in the renewable energy sector. Although up-front expenses remain high, the cost of both wind-generated and solar-generated electricity is now substantially lower than that of fossil fuels. Nuclear-generated electricity is widely considered a necessary part of a net-zero emissions energy mix in the longer-term future. However, economic considerations such as the cost and timelines for constructing reactors and the lack of an actual long-term US nuclear-waste disposal policy are a substantial impediment to more widespread deployment of nuclear power in the United States.

National Security

Artificial intelligence Because AI enables more rapid processing of an expanded range of data inputs, all aspects of military operations potentially benefit from it. Possible applications include managing military logistics; improving the effectiveness and efficiency of maintaining equipment; managing electronic medical records; navigating autonomous vehicles; operating drone swarms; recognizing targets; performing intelligence analysis; developing options for command decisions; and enhancing war gaming to develop and refine plans. However, the US Department of Defense's ethical considerations for the development and deployment of AI capabilities (especially in nuclear command and control) may not be shared by adversaries.

Biotechnology and synthetic biology With synthetic biology becoming increasingly available to state and nonstate actors, there are concerns that a malicious actor could create or deploy weaponized organisms or threaten the provision of biologically developed foods, medicines, fuels, or other products to coerce others. Conversely, the prospect of distributed biomanufacturing offers possibilities for localized biodefense and a larger degree of independence from foreign suppliers of many raw materials. China is investing considerably more resources in biotechnology than the United States, creating the potential for a Sputnik-like strategic surprise.

Cryptography Adversaries are likely to have been storing encrypted data, hoping that future advances in quantum computing and other digital capabilities will allow them to crack the encryption protecting the information. Efforts are already underway to create new encryption methods that would be quantum resistant. Separately, zero-knowledge proof methodology to cooperatively track and verify numbers of tactical nuclear warheads may benefit future arms control agreements.

Lasers Operational laser weapons systems are starting to be fielded for military applications such as short-range air defense against drones and to counter artillery, missiles, and other threats. But they have not yet been deployed widely, and their battlefield effectiveness in the face of countermeasures has not yet been tested fully.

Materials science Improvements in materials science and nanotechnology can advance capabilities in stealth technology, camouflage, and body armor and can increase the energy content in explosives. Quantum dots—materials that are smaller than about 100 nanometers in all dimensions—can be used in sensors for detecting agents associated with chemical and biological warfare.

Neuroscience Neuroscience may help illuminate the nature of traumatic brain injuries and post-traumatic stress disorder, thereby leading to better treatments for these conditions. Brain-machine interfaces could also enable new prostheses for wounded combatants.

Robotics Advances in robotics can assist military forces with transporting equipment and supplies, urban warfare, autonomous vehicle deployment, and search-and-rescue efforts. Additionally, robotics can assist with mine clearance, disaster recovery, and firefighting. Some military robots, such as lethal autonomous weapons systems, also raise questions of roboethics on the battlefield. Given the pressure for militaries to act more rapidly, many observers believe that decisions of lethal force will be turned over to computers, while others insist that life-and-death decisions must remain with humans.

Semiconductors Modern military hardware is critically dependent on semiconductor technology for information processing. The primary fabricator of semiconductor chips globally is Taiwan, which houses two of the three leading manufacturers, the Taiwan Semiconductor Manufacturing Company and the United Microelectronics Corporation. China's long-held interest in reunification with Taiwan and its rising military capabilities and assertiveness toward Taiwan are raising deep concerns about the potential for a Chinese blockade or other actions that could disrupt the global semiconductor supply chain and raise the risk of military conflict between the United States and China. The Creating Helpful Incentives to Produce Semiconductors (CHIPS) and Science Act of 2022 is intended to reduce that risk, but major initiatives called for in the legislation have not been fully funded.

Space Communications, surveillance, and navigation in denied areas are essential functions for military forces. In the future, nonnuclear weapons may be based in space and used to attack terrestrial and/or space targets. Satellites are also essential for the detection of launched ballistic missiles, nuclear weapons explosions, and electromagnetic emissions from other nations. The emergence of low-cost, high-quality information from space-based assets (almost entirely commercial) is a driver of open-source (unclassified) intelligence (OSINT), which has the potential to upend traditional intelligence processes built on classified information collection and analysis. The net effect of OSINT could be a declining US intelligence advantage, as more countries, organizations, and individuals can collect, analyze, and disseminate high-quality intelligence without expensive, space-based government satellite capabilities. The commercialization of space also puts powerful capabilities in the hands of individuals and organizations who are not accountable to voters and whose interests may not be aligned with those of the US government.

Sustainable energy technologies The United States is no longer the world leader in energy manufacturing at scale. For instance, China and other countries with lower operating costs control most of the manufacturing, supply chain, and critical minerals for battery and solar cell production. US energy security will require expansion of domestic production and manufacturing, as well as collaboration with allies and partners to better protect energy supply chains. Moreover, there are concerns that a global increase in fission reactors will result in a greater risk of nuclear proliferation (i.e., the spread of nuclear weapons), especially to nonnuclear states or nonstate actors, while some believe that the emissions-free potential of fission reactors is worth the risk of proliferation, which can be minimized through carefully implemented safeguards. Fuel security for nuclear power remains an issue as well—America currently imports more than 90 percent of its uranium, with about half coming from Kazakhstan and Russia.

Environmental and Energy Sustainability

Artificial intelligence AI capabilities can greatly improve global sustainability efforts, from helping farmers identify which produce or livestock are appropriate to harvest to helping analyze weather patterns to prepare populations and infrastructure for extreme or unusual conditions. At the same time, training and using AI models requires a large amount of energy, and demand for power to support these activities is expected to grow significantly in the future.

Biotechnology and synthetic biology Synthetic biology can contribute to new methods for energy production and environmental cleanup. Electrobiosynthesis is a biotechnology that enables plant-free bioproduction in places where soils are poor, water is scarce, or climate and weather are too variable to support traditional agriculture.

Cryptography Blockchain technologies can provide a transparent and secure way to track the movement of goods, including their origin, quantity, and other relevant information, thereby improving efficiency in global supply chains and limiting illegal extractions of certain materials. Although some established cryptocurrencies such as Bitcoin require massive amounts of energy, newer ones require far less.

Lasers Lasers can help monitor the environment. For example, when integrated into appropriate instruments, they can help measure levels of toxins or biological agents and enhance the study of microbes in soil or algae in water. In all these cases, laser light stimulates responses in the targets it illuminates that help other instruments to make more accurate measurements of the state of the environment.

Materials science Innovations in materials science and engineering are creating new and sustainable plastics that are easier to recycle. New materials can also advance the electrification of transportation and industry, which is integral to decarbonization strategies, and can support the design of relatively cheap batteries that last a long time and can be quickly recharged. Nanomaterials such as quantum dots can further improve the efficiency of solar cells and biodegradable plastics. However, some innovations in the field have potential downsides, too. For instance, the long-term dangers of nanoparticles released into the environment at the end of their life cycle are unknown.

Neuroscience Sustainability on a planet with finite resources requires that decision makers and the people they represent are able to make tradeoffs between immediate rewards and future gains. Neuroscientists have found evidence for cognitive predisposition favoring short-term gains over long-term rewards, based on functional resonance magnetic imaging (fMRI) brain scans of people making choices between immediate and delayed reward.[1] (This example is not further discussed in chapter 6.)

Robotics The deployment of robots primarily for the Three Ds—dull, dirty, or dangerous jobs—enables robotic cleanup of environmentally hazardous materials and their operation in environments that can be dangerous for humans, such as nuclear reactors. Robots are also valuable in the construction, maintenance, and management of solar and wind farms.

Semiconductors Transitioning to renewable energy sources will require vast amounts of semiconductors. Advanced chips are integral to electric vehicles, solar arrays, and wind turbines. Design innovations will continue to improve the energy efficiency of chips.

Space Remote sensing data can create a "digital twin" of Earth to track and model environmental change and the movement of humans and animals, informing disaster response and sustainable development policies. In the future, the development of space technologies will help to address food security, greenhouse gas emissions, renewable energy, and supply chain optimization. Satellite imagery, combined with weather data and powered by predictive optimization algorithms, could increase crop yields and also detect greenhouse gas emissions to identify natural-gas leaks and verify compliance with regulations. Advancing space technologies could also enable mining from the Moon and asteroids of minerals that are hard to find on Earth, as well as transmission of sustainable solar energy directly to Earth from space.

Sustainable energy technologies New investments in energy research and development are enabling advances in clean electricity generation, long-distance transmission lines, lighting based on light-emitting diodes (LEDs), and electric car batteries. Long-duration energy storage is a critical field for climate and sustainability goals. The development of batteries for electric grids that can store energy for weeks or months is needed to support the use of solar and other intermittent renewable energy sources. Renewable fuels, especially hydrogen, can replace hydrocarbons in transportation and industry. However, new hydrogen production and storage methods are needed to make its use cost-effective at scale. Nuclear power could help the United States reach sustainability goals, too, but it is unclear whether enough reactors can become operational in time to meet commitments to triple nuclear generation of electricity by 2050 compared to the 2020 baseline. Moreover, nuclear waste remains an environmental policy issue, and the United States has no enduring plan for a long-term solution to storing it.

Health and Medicine

Artificial intelligence AI data analytics are already improving the accuracy of healthcare assessments and procedures. Continued advancement could place AI-monitored cameras and sensors in the homes of elderly or at-risk patients to provide prompt attention in case of emergency while protecting patient privacy. AI-operated mobile robots can potentially replace basic nursing care.

Biotechnology and synthetic biology Synthetic biology has remarkable potential to contribute to the creation of new drugs as well as to pathogen detection and neutralization. Synthetic biology can also help to reduce disease transmission, personalize medicine through genetic modifications, improve cancer treatment, and offer custom lab-grown human tissue for medical testing. DNA sequencers and synthesizers using the internet allow researchers around the world to obtain information on viruses—and potentially vaccines or cures—even faster than a pandemic spreads. However, that same speed and accessibility raise concerns about potential misuse of the technology by bad actors. It is also unclear how some new biological organisms will interact with the natural and human environments.

Cryptography Blockchain technology can securely store all data from a person's important documents, including medical records, in encrypted form while facilitating selective data retrieval that protects a patient's privacy. This approach enables data analytics to be performed on aggregated and anonymized datasets, thus enabling researchers and internal auditors to access information without violating patients' privacy rights.

Lasers Lasers have a host of applications in medicine. Laser-based measurements of biological tissue can yield information about tissue composition and structure without the need for invasive biopsies. Lasers can also replace surgical scalpels in many instances, making cleaner and more precise incisions with less collateral damage to surrounding tissue.

Materials science Materials science and nanotechnology are improving the capabilities and effectiveness of medical devices and the delivery of treatments. For example, wearable electronic devices made from flexible materials can conform to skin or tissues to provide specific sensing or actuating functions; devices like "electronic skin," or e-skin, can sense external stimuli such as temperature or pressure; and "smart bandages" with integrated sensors and simulators can significantly accelerate healing of chronic wounds. Injectable hydrogels can fine-tune long-term delivery of medications, which can lead to improvements in the administration and efficacy of essential medicines such as insulin. Nanomaterials like quantum dots are being used as fluorescent markers in biological systems to improve the contrast of biomedical images. Finally, biosensors allow the rapid testing of blood for bacterial pathogens.

Neuroscience Advances in neuroscience may help address neurodegeneration and related diseases, such as chronic pain, depression, opioid dependency, and Alzheimer's disease, dramatically improving the quality of life of patients (and their families) and potentially reversing the anticipated rising costs associated with care. However, too many fundamental gaps still remain in our understanding of the brain to be confident of rapid progress in treating such illnesses.

Robotics Some robotics are already deployed in the healthcare industry, such as assisted laparoscopic surgical units and equipment. Improvements in haptic technology, which provides doctors using robots to operate on patients remotely with the tactile sensation of actually holding surgical tools, can increase the effectiveness and safety of these robots. Robotics will also be increasingly useful to support aging populations. Assistive robots could help people move around, while other robots can help nursing and homecare workers provide essential functions such as bathing or cleaning.

Semiconductors Semiconductor chips are ubiquitous in modern medical equipment. Imaging devices such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound use embedded computers to generate images from electromagnetic radiation and sound waves penetrating or emanating from the human body.

Space The potential for space manufacturing can improve development of specialized pharmaceuticals, which can be made in a microgravity environment with minimal contaminants.

Sustainable energy technologies A transition from fossil fuel energy to a renewable energy–based world economy would reduce greenhouse gas emissions and prevent thousands of premature deaths from pollution and extreme weather events. Eliminating energy-related air pollution in the United States alone could prevent more than fifty thousand deaths annually and save hundreds of billions of dollars a year from avoided illness. Reducing carbon dioxide emissions will result in less extreme climates, which in turn will lead to fewer health problems from extreme heat.

Civil Society

Artificial intelligence Because AI models are trained on existing datasets, they are likely to encode any biases present in these datasets, affecting model-based outcomes and decision-making. Many facial recognition algorithms are better at identifying lighter-skinned faces than darker-skinned ones, leading to discrimination against people with the latter. Indiscriminate data collection can violate privacy and copyrights. Deepfakes used for misinformation and disinformation have personal, legal, and political impacts. The long-term nature and extent of AI's impact on employment—in terms of displacing some jobs and improving productivity in others—is still unknown.

Biotechnology and synthetic biology Different religious traditions may have different stances toward life or living systems, as well as different opinions as to whether the engineering of new lifeforms violates any of their basic precepts. Another deliberation will be over who should have access to the benefits from synthetic biology given the risks to human and environmental safety from both malicious and unintentional acts.

Cryptography The nature of cryptography and encrypted communications raises questions about exceptional access regulations, which would require communications carriers and technology vendors to provide access to encrypted information to law enforcement agents or other bodies under specific legal conditions on the basis that encryption technology is also accessible to criminals and other malefactors. Opponents of exceptional access argue that implementing this capability weakens the security provided by encryption. Its supporters argue that the reduction in personal encryption security is worth the benefits to law enforcement of being able to catch and prosecute bad actors.

Lasers Lasers per se do not particularly raise issues of relevance to civil society. Instead, such issues often arise for specific applications in which lasers play a central role. Different ones arise if lasers are being used in a military context, a medical context, a manufacturing context, and so on. And even within each of these domains, concerns can vary. For example, certain types of laser can produce a toxic exhaust, which obviously raises environmental concerns, while others do not produce noxious outputs.

Materials science Given the many uncertainties about the long-term dangers and health concerns of nanoparticles released into the environment, important questions arise about how and to what extent regulations should be adopted to mitigate risks that might accompany such releases. Resolving them will require seeking consensus on the magnitude and severity of these risks, as well as on appropriate remedies.

Neuroscience Neuroscience development is influenced by existing legal frameworks. The Controlled Substances Act, for instance, limits medical research on some substances that may have therapeutic effects. Meanwhile, cognitive and behavioral neuroscience have broad implications for public policy because a basic aspect of criminal law is the nature and extent of an individual's responsibility for a criminal act. Minors under eighteen years of age, for example, cannot be subject to the death penalty for crimes they committed because adolescent brains are not considered fully developed, putting minors at higher risk of impulsive, irrational thoughts and behaviors. As neuroscience advances, it could find evidence that reinforces or contradicts this and other principles.

Robotics Greater adoption of robotics will require moving workers to new roles as well as setting standards for human safety around robots. As robots assume more tasks, human workers will need education and training programs to undertake new roles and to benefit from robotics. Standards will also be needed to clarify limits to robotic applications. Ethical considerations warranting policy development include how to ensure data acquisition for training robots respects privacy and inclusiveness and how to set safety standards (i.e., Should the requirement be that a robot's performance is comparable to that of an average human, or should it be near perfect?). Safety considerations for human-robot interactions will be an ongoing challenge.

Semiconductors Student interest in hardware design has dropped precipitously in favor of software-oriented jobs. Some estimates suggest that by 2030, 60 to 80 percent of jobs in semiconductor manufacturing will be unfilled given current rates at which students with relevant degrees are graduating in the United States.

Space In space, the rapid expansion of commercial assets and applications is raising important new policy considerations not covered by current norms. The increasing dependence of government on the private sector to provide space-based capabilities—including launch, vehicles, and space-based communications and internet access—that are vital to national security and economic growth raises questions about how to align public and private interests. Attempts at improvement have often stagnated due to nations' differing geopolitical aims. Dual-use space technologies and the challenge of getting private and government actors to cooperate will complicate crisis response.

Sustainable energy technologies Continued creation of sustainable energy infrastructure requires new acquisitions of land to build generating stations and storage facilities, which can displace residents from private property and impact local property values, encouraging some to adopt a position of supporting windmills but "not in my backyard." The construction of nuclear power plants and facilities for storing radioactive waste is often met with opposition from those concerned about exposure to radiation in the environment.

NOTES

1. Emmanuel Guizar Rosales, Thomas Baumgartner, and Daria Knoch, "Interindividual Differences in Intergenerational Sustainable Behavior Are Associated with Cortical Thickness of the Dorsomedial and Dorsolateral Prefrontal Cortex," *NeuroImage* 264, no. 119664 (2022), https://doi.org/10.1016/j.neuroimage.2022.119664.

CONCLUSION

This new edition of the *Stanford Emerging Technology Review* has spotlighted ten pivotal technological domains that are shaping the future of science and innovation. Our extensive consultations with leading Stanford academics across scientific disciplines make clear that the coming decade will witness an unprecedented convergence of multiple technologies, driving progress at extraordinary speeds. Artificial intelligence (AI), fueled by increased computing power and more data, has potential to enhance human productivity dramatically and accelerate advancements across scientific fields, from drug discovery to breakthroughs in new materials. Synthetic biology and biotechnology promise groundbreaking applications in agriculture, healthcare, and industrial production. Technological progress spans from the vast expanses of space to the microscopic world of nanoparticles. While technology itself is neither good nor bad, it's crucial for decision makers to comprehend the scale of technological change, its potential to either improve or disrupt societal norms—and the imperative for American leadership in navigating these expanding frontiers.

For decades, the prevailing approach to US science and technology policy has been to fund research at academic institutions and national laboratories, anticipate breakthroughs, and hope for positive outcomes. However, the new landscape we are navigating demands a shift in strategy. On the path toward the decisive establishment of US leadership in science and technology, much has been accomplished in the past. Yet much more needs to be done in the future, and policymakers must engage more actively with the technology community in both academia and industry to shape the ecosystem in a way that serves the interests of the American people.

Ultimately, humans develop and use technology, and effective governance to maximize benefits and mitigate risks requires human guidance. Policymakers can establish frameworks that encourage innovation, set priorities and strategies, align economic policies to foster innovation and maintain leadership, and bolster America's position in international competition. For instance, innovations in renewable energy offer a path to both energy independence and sustainability, but they require ongoing government attention and funding to attract innovators and overcome developmental challenges.

As well as offering a look at individual emerging technologies, this publication also highlights common themes that emerge across them related to the development of science and technology. The importance of universities in the American innovation trifecta—government, academia, and industry—stands out as a crucial factor. As we noted at the

Ultimately, humans develop and use technology, and effective governance to maximize benefits and mitigate risks requires human guidance.

start of this report, the US government is the only funder capable of making the large, sustained, and sometimes risky investments in the basic science conducted at universities that will be essential for future applications. Such support is going to be even more critical in the years ahead as the clock speed of technological change continues to accelerate and as other nations step up their own investments in fundamental research.

Gaining a technological lead in a domain is distinct from maintaining it. Engaging with expertise around the world, leveraging the potential of highly skilled immigrants, and sustaining robust domestic development of scientific expertise are essential to reinforce American leadership in an increasingly competitive global landscape. Recognizing the evolving role of government in technology development is also vital. Innovations are no longer created and protected solely by state-backed research groups; private corporations and even individuals are developing more and more transformative technologies.

This paradigm shift is most evident in fields like AI and space exploration, where private companies are spearheading the creation of large language model systems and deploying innovative, highly advanced assets into space—a domain previously dominated by governments. The concentration of power in different hands has significant implications for technology access, priorities, and policy.

This edition of *SETR* started by asking the question: "What do policymakers need to know about emerging technologies from Stanford?" It serves as an initial step in providing the necessary and rapidly changing knowledge about these crucial technologies, their key takeaways, future implications, and potential policy concerns. The goal is to foster meaningful and ongoing discussions that can lead to effective and timely policymaking, even as technologies continue to evolve. We hope you found it useful, and we welcome feedback on how to make the publication even more impactful in the future—send your thoughts to SETReview2025@stanford.edu.

LEADERSHIP

Co-chairs

Condoleezza Rice is the Tad and Dianne Taube Director of the Hoover Institution and a senior fellow on public policy. She is the Denning Professor in Global Business and the Economy at the Stanford Graduate School of Business. She served as the sixty-sixth secretary of state (2005–9) and the nineteenth national security advisor (2001–5). Rice served as Stanford University's provost from 1993 to 1999. She is author and coauthor of nine books. She received her PhD in political science from the University of Denver.

John B. Taylor is the George P. Shultz Senior Fellow in Economics at the Hoover Institution and the Mary and Robert Raymond Professor of Economics at Stanford University. He is also the director of Stanford's Introductory Economics Center. He specializes in monetary policy, fiscal policy, and international economics. He served as under secretary for the Treasury for international affairs (2001–5) and on the President's Council of Economic Advisers. He has authored and coauthored twenty-one books, including *First Principles: Five Keys to Restoring America's Prosperity*, winner of the 2012 Hayek Prize. He received his PhD in economics from Stanford University.

Jennifer Widom is the Frederick Emmons Terman Dean of the School of Engineering and the Fletcher Jones Professor in Computer Science and Electrical Engineering at Stanford University. Her research interests span many aspects of nontraditional data management. She is a fellow of the Association for Computing Machinery and a member of the National Academy of Engineering and the American Academy of Arts and Sciences. She received her PhD in computer science from Cornell University.

Amy Zegart is the Morris Arnold and Nona Jean Cox Senior Fellow at the Hoover Institution and professor, by courtesy, of political science at Stanford University. She is also a senior fellow at Stanford's Institute for Human-Centered Artificial Intelligence and at its Freeman Spogli Institute for International Studies. The author of five books, she specializes in US intelligence, emerging technologies and national security, grand strategy, and global political risk management. She received her PhD in political science from Stanford University.

Director and Editor in Chief

Herbert S. Lin is senior research scholar at the Center for International Security and Cooperation and Hank J. Holland Fellow in Cyber Policy and Security at the Hoover Institution, both at Stanford University. His research interests relate broadly to emerging technologies and national security. He is an elected fellow of the American Association for the Advancement of Science. He received his doctorate in physics from the Massachusetts Institute of Technology.

Managing Editor

Martin Giles is a policy fellow at the Hoover Institution and the assistant director of Hoover's Technology Policy Accelerator at Stanford University. He is a seasoned executive, with editing and publishing experience at *The Economist* and *MIT Technology Review*, where he focused on developments in Silicon Valley and on emerging technologies. Giles was also the first editorial director of In-Q-Tel. He received his MBA from the University of Chicago Booth School of Business.

Faculty Council

Zhenan Bao is the K. K. Lee Professor in Chemical Engineering and professor, by courtesy, of chemistry and of materials science and engineering at Stanford University. She is a member of the National Academy of Sciences, the National Academy of Engineering, the American Academy of Arts and Sciences, and the National Academy of Inventors. Bao is known for her work on artificial electronic skin, which is enabling a new generation of skin-like electronics for regaining sense of touch for neuroprosthetics, human-friendly robots, human-machine interfaces, and seamless health-monitoring devices. She received her PhD in chemistry from the University of Chicago.

Sally Benson is the Precourt Family Professor in the Department of Energy Science and Engineering in the Stanford Doerr School of Sustainability and a senior fellow in the Precourt Institute for Energy and the Woods Institute for the Environment at Stanford University. Her research focuses on technologies and pathways to reducing greenhouse gas emissions. She received her PhD in materials science and mineral engineering from the University of California–Berkeley.

Dan Boneh is a professor of cryptography and electrical engineering at Stanford University, codirector of the Stanford Computer Security Lab, and a senior fellow at the Freeman Spogli Institute for International Studies. His research focuses on applied cryptography and computer security. He has authored more than one hundred publications. He received his PhD in computer science from Princeton University.

Simone D'Amico is an associate professor of aeronautics and astronautics and, by courtesy, of geophysics at Stanford University, where he serves as the W. M. Keck Faculty Scholar of Engineering. He is the founding director of Stanford's Space Rendezvous Laboratory and its Center for AEroSpace Autonomy Research (CAESAR) and is a science fellow at the Hoover Institution. His research explores the intersection of advanced astrodynamics, spacecraft navigation and control, and machine learning to enable future distributed space systems. He currently leads four satellite swarm and formation-flying projects for NASA and the National Science Foundation. He received his PhD in aerospace engineering from Delft University of Technology.

Drew Endy is the Martin Family University Fellow in Undergraduate Education (bioengineering), faculty codirector of degree programs for the Hasso Plattner Institute of Design (the d.school), core faculty at the Center for International Security and Cooperation, senior fellow (courtesy) of the Freeman Spogli Institute for International Studies, and senior fellow (courtesy) and science fellow at the Hoover Institution at Stanford University. He serves as president and director of the Biobricks Foundation and director of the iGEM Foundation and the Biobuilder Educational Foundation. His research focuses on the foundations of synthetic biology along with broader societal aspects. He received his PhD in biotechnology and biochemical engineering from Dartmouth College.

Siegfried Glenzer is a professor of photon science and, by courtesy, of mechanical engineering at Stanford University, where he serves as director of the High Energy Density Science division at SLAC National Accelerator Laboratory. He is also deputy director of the Inertial Fusion Science and Technology hub, also known as RISE, funded by the Department of Energy. His research explores matter in extreme conditions and the development of fusion energy sciences and technologies. He has authored or coauthored more than 550 journal publications. He received his PhD in physics from Ruhr University Bochum.

Mark A. Horowitz is the Fortinet Founders Chair of the Department of Electrical Engineering and the Yahoo! Founders Professor in the School of Engineering at Stanford University. His research has contributed to early RISC (reduced instruction set computer) microprocessors, multiprocessor designs, and high-speed interfaces, and he currently works to create new agile design methodologies for analog and digital VLSI (very-large-scale integration) circuits. He received his PhD in electrical engineering from Stanford University.

Fei-Fei Li is the inaugural Sequoia Professor in the Computer Science Department at Stanford University and founding codirector of Stanford's Institute for Human-Centered Artificial Intelligence. Her current research focuses on cognitively inspired AI, machine learning, deep learning, computer vision, robotic learning, and AI and healthcare. She received her PhD in electrical engineering from the California Institute of Technology and holds an honorary doctorate from Harvey Mudd College.

Allison Okamura is the Richard M. Weiland Professor of mechanical engineering in the School of Engineering and professor, by courtesy, of computer science at Stanford University. She is a deputy director of the Wu Tsai Neurosciences Institute, affiliated faculty at Stanford Bio-X and Stanford's Institute for Human-Centered Artificial Intelligence, a senior fellow (courtesy) and science fellow at the Hoover Institution, and a fellow of the Institute of Electrical and Electronics Engineers. Her research interests include haptics, teleoperation, mixed reality, and medical and soft robotics. She received her PhD in mechanical engineering from Stanford University.

Kang Shen is the Frank Lee and Carol Hall Professor of biology and professor of pathology at Stanford University, where he serves as the Vincent V. C. Woo Director of the Wu Tsai Neurosciences Institute and affiliated faculty at Stanford Bio-X. His research focuses on neuronal cell biology and developmental neuroscience. He has authored or coauthored more than one hundred journal articles. He received his PhD in cell biology from Duke University.

Advisory Board

Steven Chu is the William R. Kenan Jr. Professor of Physics and professor of molecular and cellular physiology and of energy science and engineering at Stanford University. He previously served as US secretary of energy (2009–13) and as director of the Lawrence Berkeley National Laboratory. He received his PhD in physics from the University of California–Berkeley.

Robert Gates served as the twenty-second secretary of defense (2006–11) and as the director of central intelligence (1991–93) following nearly twenty-seven years in the CIA and on the National Security Council. He is the current chancellor of William & Mary and previously served as president of Texas A&M University. He received his PhD in Russian and Soviet history from Georgetown University.

Susan M. (Sue) Gordon is the former principal deputy director of national intelligence at the Office of the Director of National Intelligence (2017–19). She was formerly deputy director of the National Geospatial-Intelligence Agency (2015–17). She served twenty-seven years in the CIA, including as director of the CIA's Information Operations Center, and is now a leading voice on intelligence, cyber, space, and disruptive technologies as an independent director, advisor, and university fellow. She continues to shape national security as a member of the Council on Foreign Relations and the Defense Innovation Board.

John L. Hennessy is a professor of electrical engineering and computer science at Stanford University and the chairman of Alphabet Inc. He previously served as the tenth president of Stanford University (2000–16) and is the founder and director of the Knight-Hennessy Scholars Program, the world's largest fully endowed, university-wide graduate fellowship. He received his PhD in computer science from the State University of New York at Stony Brook.

Norbert Holtkamp is a science fellow at the Hoover Institution and professor of particle physics and astrophysics and of photon science at the SLAC National Accelerator Laboratory. He previously served as SLAC's deputy laboratory director (2014–22). He received his PhD in physics from the Technical University of Darmstadt in Germany.

Jerry McNerney is the former US representative for California's ninth congressional district (2007–23), where he served on the Committee on Energy and Commerce, the Veterans' Affairs Committee, and the Space, Science, and Technology Committee. He was elected to the California State Senate in 2024 and is a senior policy advisor at Pillsbury Winthrop Shaw Pittman LLP.

He received his PhD in mathematics from the University of New Mexico–Albuquerque.

Mary Meeker is a cofounder and general partner of BOND and serves on the boards of Block/Square, Genies, and Plaid. Her investments at BOND (and, previously, Kleiner Perkins) include Spotify, Waze, DocuSign, Ring, Checkr, Ironclad, and On Running. She was previously a managing director at Morgan Stanley, focused on emerging technology companies, coauthor of *USA Inc.: A Basic Summary of America's Financial Statements*, and publisher of the widely distributed Internet Trends Report.

Lloyd B. Minor is the Carl and Elizabeth Naumann Dean of the School of Medicine and vice president for medical affairs at Stanford University. He is a professor of otolaryngology–head and neck surgery and professor, by courtesy, of bioengineering and neurobiology at Stanford University. Prior to joining Stanford, he was provost and senior vice president for academic affairs at Johns Hopkins University. He is an elected member of the National Academy of Medicine.

Peter L. Scher is the vice chairman of JPMorgan Chase & Co. He previously served in senior government positions, including as US special trade ambassador under President Clinton and staff director of the US Senate Committee on Environment and Public Works. Prior to joining JPMorgan Chase & Co. in 2008, he was the managing partner of the Washington, DC, office of law firm Mayer Brown LLP.

Eric Schmidt is cofounder, with his wife, Wendy, of Schmidt Sciences, a nonprofit organization working to advance science and technology to address global challenges, and currently chairs the Special Competitive Studies Project, a nonprofit initiative focused on strengthening America's long-term AI and technological competitiveness in national security, the economy, and society. Previously, he was CEO and chairman of Google (2001–11) and executive chairman and technical advisor to Alphabet Inc. (2015–20). He received his PhD in electrical engineering and computer science from the University of California–Berkeley.

Thomas M. Siebel is the chairman and CEO of C3 AI and was the founder and CEO of Siebel Systems. He is a member of the colleges of engineering boards at the University of Illinois Urbana-Champaign and the University of California–Berkeley and is an elected member of the American Academy of Arts and Sciences.

Senior Research Staff

Malaysia Atwater is a research assistant for the Stanford Emerging Technology Review and the Hoover Institution's Technology Policy Accelerator at Stanford University. Her research focuses broadly on the policy considerations and implications of emerging technologies. She received her BA in political science from Stanford University.

Tanner Braman is a senior research program manager for the Bio-Strategies and Leadership Initiative at the Hoover Institution at Stanford University. His research focuses on policy issues in biosecurity and the emerging bioeconomy. He received his BS in chemistry and BA in international relations from William & Mary.

Corie Wieland is a cyber research associate at the Hoover Institution at Stanford University. In addition to cyber policy and security, her research focuses include emerging technologies, gray area tactics, and national security policy. She received her BA in international relations and her MA in international policy from Stanford University.

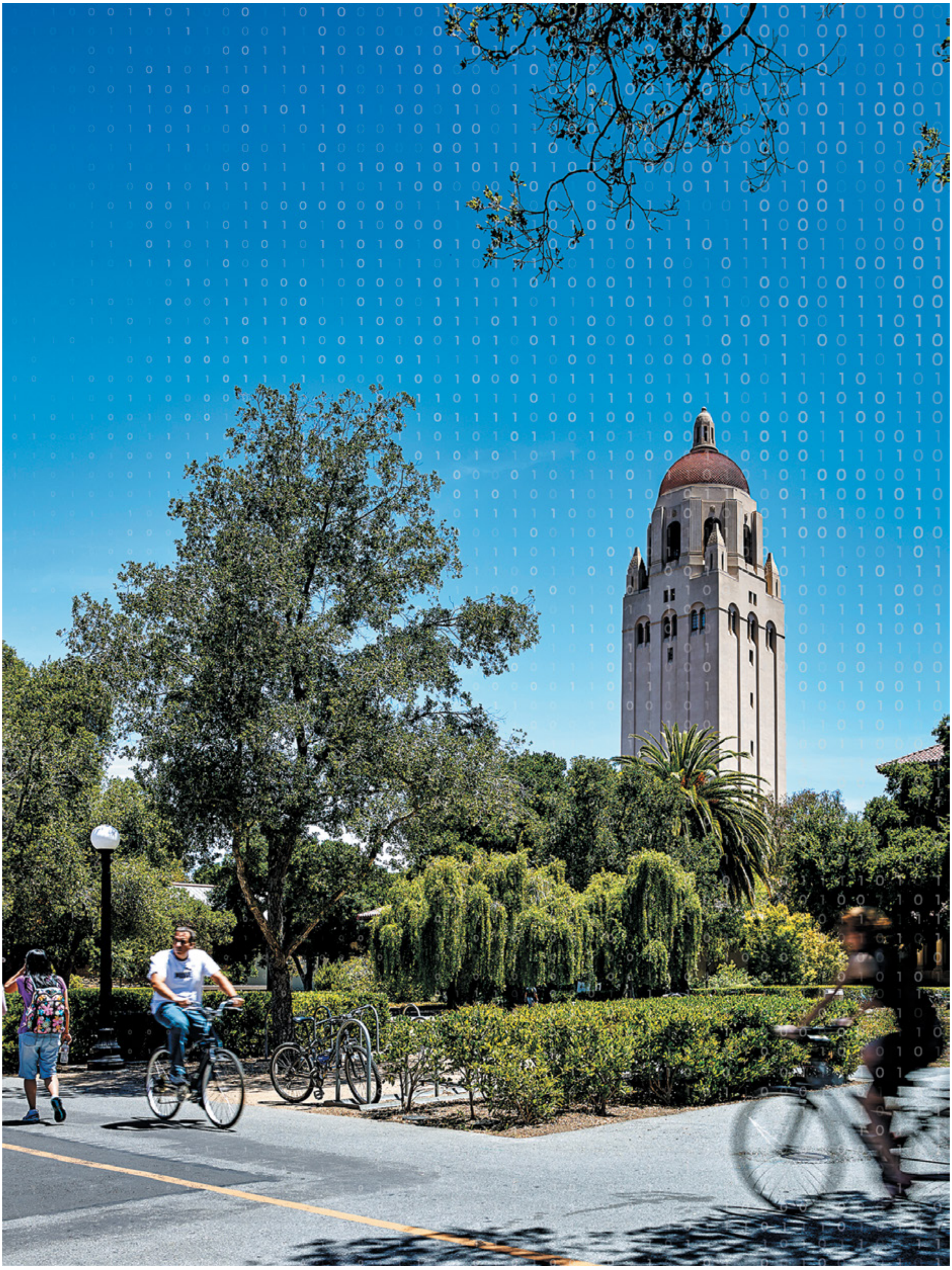
ACKNOWLEDGMENTS

Production of the *Stanford Emerging Technology Review* (*SETR*) is a team effort.

The members of the SETR Faculty Council gave seminars on leading-edge work going on in their respective fields, and many of their faculty colleagues offered expert commentary and insights at these seminars. SETR faculty and fellows both contributed much effort in crafting entire chapters based on their deep understanding of their respective domains, including original ideas drawn from their own research. The SETR Advisory Board gave invaluable advice on how to interface with policymakers and the policy process. SETR staff and undergraduate research assistants did heroic work fact-checking and tracking down references, among many valuable tasks they performed. The Hoover Press did a wonderful job turning the manuscript into a final report.

For all of this, we are grateful.

Condoleezza Rice, Co-chair
John B. Taylor, Co-chair
Jennifer Widom, Co-chair
Amy Zegart, Co-chair
Herbert S. Lin, Director and Editor in Chief
Martin Giles, Managing Editor



31 30 29 28 27 26 25 7 6 5 4 3 2 1

Designer: Howie Severson
Typesetter: Maureen Forys

Image credits: iStock.com/monsitj (cover); iStock.com/mofuku (p. 20); iStock.com/wacomka (p. 38); iStock.com/FeelPic (p. 52); iStock.com/DanBrandenburg (p. 64); iStock.com/Chartchai Sansaneeyashewin (p. 76); iStock.com/ArtemisDiana (p. 90); iStock.com/imaginima (p. 102); iStock.com/Floriana (p. 114); iStock.com/dima_zel (p. 126); iStock.com/JONGHO SHIN (p. 140); Tim Griffith (p. 189)