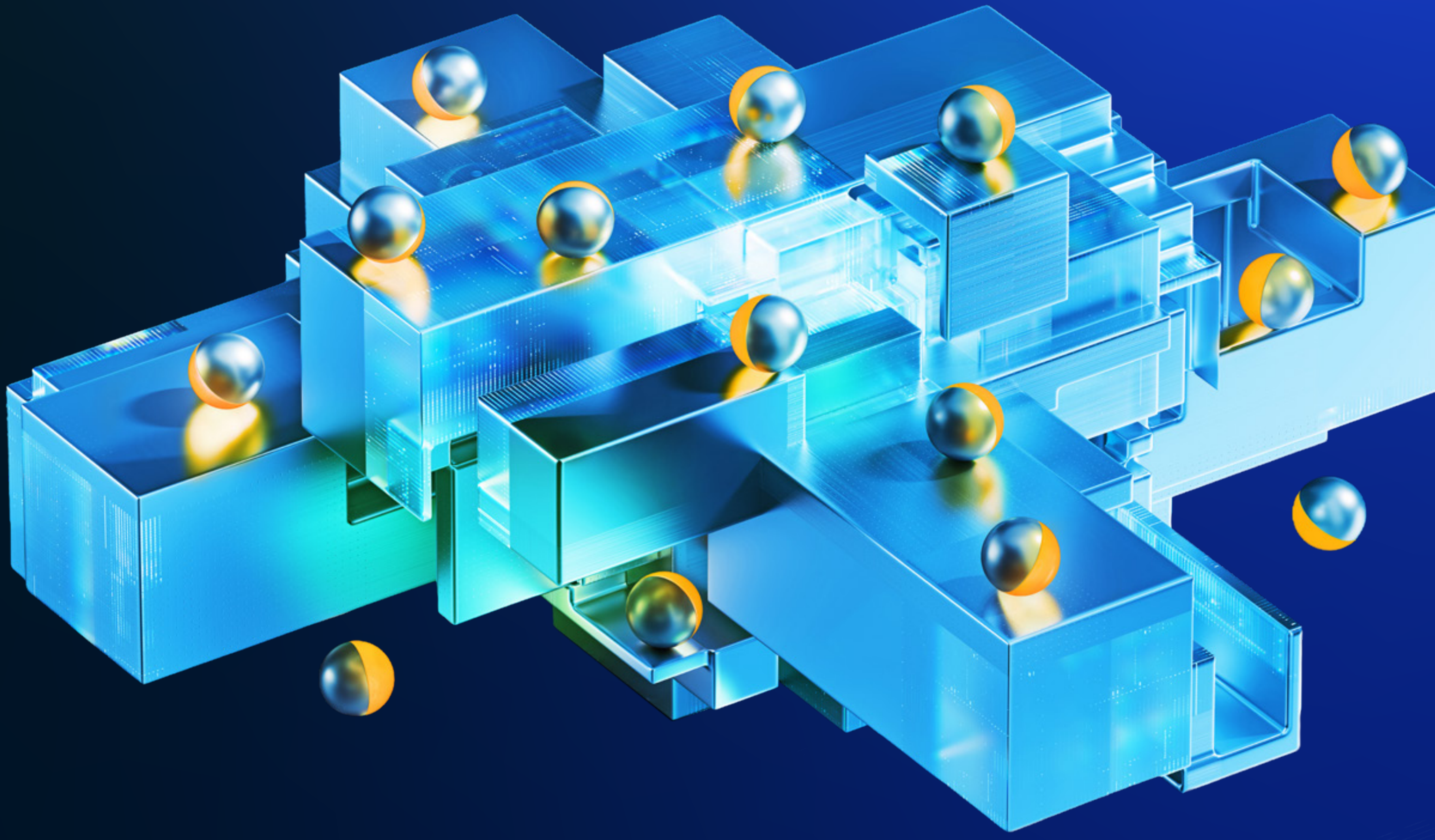
Technology Trends Outlook 2026



September 2026
Sixth edition

Which frontier technologies matter most for companies in 2026? Our annual report highlights the latest technology innovations, developments, and talent trends and their potential impact on business and society.

by Michael Chui, Roger Roberts, and Tanguy Catlin

Cover image © Evgenii Mymrin/Stocksy.
All interiors images © Getty Images.

Contents

Introduction

The technology story of 2026 has moved off the screen and into the physical world. Innovation is accelerating in the power grids and chips that underpin the data center boom; in the intelligent robots that embody AI; in the agentic systems discovering new chemical compounds; and in the launch pads sending thousands of satellites into orbit.

AI needs energy to scale. That's one reason energy technologies alone drew nearly $200 billion in investment in 2025, among the highest capital influx in any technology domain. And spending on AI infrastructure doubled in a single year. These developments show that the defining questions today are not only about what technology can do. They are also about who can build the hardware and assemble the skilled workforce to deploy AI in the real world. At the same time, huge leaps were made in cybersecurity and software development—illustrating that AI is accelerating the digital frontier, too.

McKinsey's *Technology Trends Outlook 2026* examines 14 technology trends that define 2026, expanding our coverage from last year to include two new fast-emerging domains: agentic software development and AI for scientific discovery and engineering. For easier navigation, we group the trends into three broader categories: AI revolution, compute and connectivity frontiers, and cutting-edge engineering. The lines between these domains are blurring, and much of the innovation is happening in the gaps.

The report is based on extensive quantitative analysis. We measured metrics on innovation, investment, interest, talent demand, and organizational adoption (see sidebar "Research methodology"). It also draws on qualitative interviews with business leaders, policymakers, investors, and experts across sectors and geographies. The technologies included in this report were selected based on their rate of innovation, as well as their growing applicability and potential to influence how we live and work (see sidebar "How we chose the trends").

Four of this year's trends are AI specific, but AI also underpins and amplifies the other ten. AI has become an accelerant in robotics, creating a new class of autonomous bots capable of interacting with the world around them. AI is reshaping software development, making coders many magnitudes more productive. AI is advancing scientific discovery, helping life sciences companies discover new drugs many times faster than before. And AI is optimizing the foundations of technology infrastructure itself, changing cybersecurity operations, and catalyzing new innovations in semiconductors. These are just a few of the ways AI is shifting the sands beneath our feet.

Beyond the AI ecosystem, many of this year's trends are being shaped by large-scale economic, geopolitical, and workforce transitions. Scientists and engineers are discovering new possibilities in therapeutics, diagnostics, genomics, and brain science. Robotics, mobility, and immersive-reality technologies are converging to reshape how humans interact with digital and physical systems. Sustainability technologies are transforming how energy is produced, managed, and used—providing potential solutions to meet surging power demand from data centers. And quantum computing continues to advance, giving a glimpse into how it could soon solve problems that would otherwise be infeasible.

This breakneck pace of change comes with challenges. Organizations are racing to deploy AI at scale without any proven road maps. They have workforces that need upskilling, legacy systems that need updating, and networks exposed to ever-evolving security risks. They face shortages in energy, talent, and capital. Overcoming these roadblocks requires far more than just deploying technology. It requires rewiring operational models from the inside out.

Research methodology

To assess the development of each of the 14 technology trends highlighted in this report, we collected data on six tangible measures of activity: search engine queries, news articles, patents, research publications, equity investment, and talent demand. We updated the data sources and keywords for 2026. For each metric or vector, we used a defined set of data sources to identify keyword occurrences associated with each trend, screened those occurrences for valid mentions of activity, and indexed the resulting mention counts on a scoring scale of zero to one relative to the trends studied. The innovation score combines the patents and research scores; the interest score combines the press mentions and search scores. (While we recognize that an interest score can be inflated by deliberate efforts to stimulate news and search activity, we believe that each score fairly reflects the extent of discussion and debate about a given trend.) Investment measures the flow of funds from capital markets to companies linked with the trend. Data sources for the scores include the following:

- *Patents.* Data on patent filings are sourced from Google Patents, which highlights data on the number of patents granted.
- *Research.* Data on research publications are sourced from The Lens.
- *News.* Data on news articles are sourced from Factiva.
- *Searches.* Data on search engine queries are sourced from Google Trends.
- *Equity investment.* Equity investment data are sourced from PitchBook and include data on private-market and public-market capital raises across venture capital and corporate and strategic M&A, including joint ventures; private equity investments, including buyouts and private investment in public equity; and public investments, including IPOs. Investment data does not include corporate capital and operational expenditures.
- *Talent demand.* The number of job postings is sourced from McKinsey's proprietary Organizational Data Platform, which stores licensed, de-identified data on publicly available professional profiles and job postings. Data are drawn primarily from English-speaking countries.

In addition, we updated the selection and definition of trends from last year's report to reflect the evolution of technology trends:

- Last year's AI trend is now AI infrastructure and model architectures to more accurately reflect the foundational technologies that underpin AI. Last year's digital trust and cybersecurity is now cybersecurity and trustworthy systems, highlighting how advances in AI require security to be continually woven into integrated hardware and software systems. And last year's future of bioengineering is now future of life sciences and bioengineering, which more fully reflects advances across the life sciences spectrum.
- Two newly highlighted trends join this year's report: agentic software development and AI for scientific discovery and engineering. These new trends reflect the outsize impact AI has had in these areas, having already fundamentally altered how software is built and beginning to change how scientific research and engineering gets done.
- A trend from last year, cloud and edge computing, is not included in the 2026 report. These technologies have matured and become widely adopted, and while there is continuing innovation in the space, its measured levels are not as high as they are for more frontier trends.

For equity investment insights into the future of space technologies and quantum technologies trends, we built on research from McKinsey's Aerospace & Defense Practice and McKinsey's 2026 *Quantum Technology Monitor* report.[1]

We used insights gathered from expert interviews to assign enterprise-wide adoption scores (on a scale of one to five) for each trend. The scores are defined as follows:

- *1—Frontier innovation.* This technology is still nascent, and few organizations are investing in or applying it. It is largely unproven in a business context.
- *2—Experimentation.* Organizations are testing the functionality and viability of the technology with small-scale prototypes, typically without a focus on a near-term ROI. Few companies are scaling or have fully scaled the technology.
- *3—Piloting.* Organizations are deploying the technology in the first few business use cases, via pilot projects or limited implementation, to test its feasibility and effectiveness.
- *4—Scaling in progress.* Organizations are scaling the deployment and adoption of the technology across the enterprise.
- *5—Fully scaled.* Organizations have fully deployed and integrated the technology across the enterprise. It has become the standard and is being used at a large scale as companies have recognized the value and benefits of the technology.

---

[1] "McKinsey Quantum Technology Monitor 2026: A commercial tipping point," McKinsey, April 28, 2026.

New and notable

With so much happening across so many technologies, it's not easy to identify what matters most. But across all 14 trends, five themes kept surfacing—ones we think all leaders need to understand to steer their organizations into the future.

Machines are being given more autonomy
AI has already transformed screen-based workflows—generating answers, drafting documents, and summarizing calls—and is rapidly advancing into an agentic era, in which it will complete many end-to-end digital tasks on its own. Agentic AI is becoming the connective tissue of the enterprise, with agents working alongside humans, changing not just tools but operating models. Physical AI is the next frontier. Making the jump to the real world, AI is adding perception, reasoning, and action across robotics, mobility, and wearables. General-purpose robots are being trained to learn about their environments so they can execute complex tasks and navigate unpredictable environments. Vehicles can make real-time decisions without drivers. Industrial systems can produce complex goods inside "dark factories" with no humans present. And immersive-reality headsets are interacting in real time with both wearers and the outside world. Physical AI is arriving first in manufacturing and logistics, where the economics are clearest. But the trajectory points well beyond the factory floor, toward hospitals, construction sites, farms, and city infrastructure.

AI is generating breakthroughs faster than we can absorb them
In biopharma, AI can now propose thousands of drug candidates in the time it once took to generate a handful. But it can't compress the years of wet-lab validation, clinical trials, and regulatory review

How we chose the trends

Technological innovation is as old as humankind. But it wasn't until recently that innovation began to improve our collective lives on a wide scale. For most of recorded history, GDP per capita grew very slowly, reaching approximately $1,500 by the early 1800s. Then innovation took off during the Industrial Revolution, creating a snowball of growth and development. Since that time, GDP per capita has grown by more than 14 times. And people haven't just become wealthier in material terms—they've also experienced dramatic improvements in quality of life. In 1900, the average life expectancy of a newborn was 32 years. By 2021, this had more than doubled to 71 years. These advances were made possible by technological breakthroughs.

This report seeks to identify technology trends that will shape the next year, the next decade, and even the next century. These trends are not flash-in-the-pan innovations but rather key enablers of change that leaders should understand to help their organizations evolve and grow. The resulting list is meant to be useful, not perfect or complete. For any organization, there may be other trends (or subtrends) that are particularly relevant to its sector. Some trends might demand immediate action and investment; others might be worth monitoring.

Using quantitative and qualitative research, we identify technologies with high levels of innovation and whose potential impact is wide ranging (for example, spanning multiple industry sectors and geographies and delivering impact as either economic value or positive contributions to society) and significant (for instance, technologies that could reshape the competitive landscape).

We use several quantitative measures to identify and prioritize trends, but the analysis is not purely algorithmic. Creating the list of trends is a holistic exercise that includes in-depth interviews with leaders who set the vision for their organizations, as well as with engineers, scientists, and practitioners doing the hard work in the field. Through this qualitative lens, we strive to identify trends that resonate with leaders, even if the technologies are not mutually exclusive (indeed, these trends overlap in many ways).

We measure the level and trajectories of *innovation* across trends through patents and papers. The level of financial *investments* is an indication of their potential impact, as reflected in the allocation of market capital. Measures of *interest* (including web searches and press mentions) indicate both the breadth and depth of applicability. These metrics, while necessarily imperfect, are directional indicators.

Historically, technological innovations have underpinned advancements in individual and societal well-being. We expect that the technology trends in this report will contribute substantially to a future wave of progress, helping address some of the major challenges facing today's societies. Technology trends could accelerate productivity and economic growth, reskill workforces for the future, improve human longevity and health, and transition energy systems to create more abundance and sustainability.

that follow. In software, AI is producing code faster than human systems can review, test, and deploy it securely, filling enterprise systems with brittle code. In materials science, promising compounds are arriving faster than labs can synthesize and validate them. AI is the most powerful accelerant in recent history. Its speed is also its biggest obstacle.

The cyber defense window has compressed

For decades, the cat-and-mouse game of cybersecurity played out over days and weeks, giving defenders time to find and fix vulnerabilities before attackers could fully exploit them. AI has eliminated that buffer. More than three-quarters of all cybersecurity vulnerabilities are currently classified as "zero day," meaning that by the time they are publicly disclosed, an exploit has already been developed.[1] Thus, while AI is helping security teams find and fix vulnerabilities faster, it is also helping attackers find and exploit them faster. Anthropic's handling of Claude Mythos Preview reflects this duality. The model identified thousands of potential security flaws. So rather than release the model publicly, Anthropic gave a limited group of defenders gated access through Project Glasswing to identify, validate, and patch the vulnerabilities.[2]

Hardware and software are being codesigned for differentiated AI workloads

General-purpose chips have long powered everything from laptops to data centers. AI changed that. Training models and then running them at scale (what's known as inference) demands something more specialized: chips optimized for specific workloads. Inference is overtaking training as the dominant AI workload. As model architectures continue to evolve, new application-specific chips are being designed to deliver inference on those models more efficiently, providing more output per watt at a lower cost. This is critical, as data centers' energy demand is increasingly straining power grids. Hyperscalers are investing heavily to build data centers and have much to gain from faster, higher-performance chips. Thus, Amazon, Google, Meta, and Microsoft are increasingly partnering with semiconductor firms to codesign custom silicon tailored to their AI models—and some are exploring ways to offer these chips to outside customers as competitive products. (In the chip industry, the customer is becoming an alternative supplier.) But these new-format chips are not just affecting the semiconductor sector. They are changing how physical AI infrastructure is designed and transforming the business models of the equipment makers and energy suppliers that support these build-outs.

AI is hungry, and the grid is not ready

The race to deploy AI at scale has run headlong into a constraint that hyperscaler ingenuity cannot entirely solve: power. US data centers running AI workloads alone are projected to consume as much electricity by 2030 as California does today.[3] Globally, the numbers are even larger. The problem is not just how much power AI needs but how hard it is to deliver. Data centers can be built faster than the transmission lines, substations, and transformers needed to power them can be supplied. Supply chain constraints are often to blame. In many markets, transformers now carry lead times of more than two years. More than 2,500 gigawatts of energy projects are stalled in grid queues worldwide, waiting for connections that may be years away.[4] For enterprises, securing reliable compute power is becoming as much a competitive advantage as securing talent or capital.

When measuring talent demand, we see signals that some trends are maturing and selectively scaling. In connectivity, cybersecurity, energy, life sciences, and mobility, more than half of job listings were for non-R&D roles such as general and administrative, operations, and sales and marketing. While these trends are still driven by innovation, deployment is underway. Companies are now applying these technologies in use cases with commercial viability. Meanwhile, in all four AI-related trends plus application-specific semiconductors, over 75 percent of jobs posted were in the R&D category, illustrating just how early these sectors are, despite rapid growth in the past few years.

Investment metrics provide insight into the technology trends that have captured the most

---

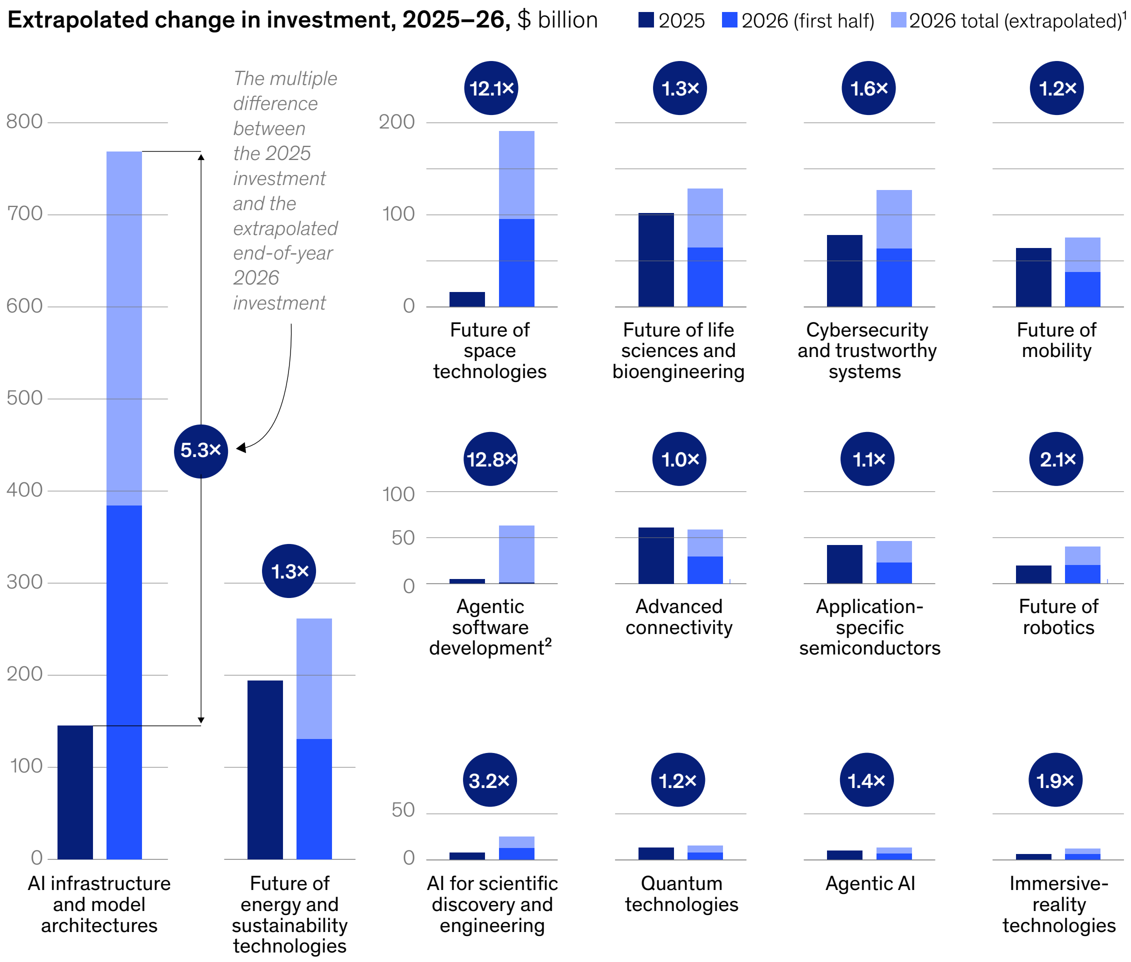
[1] "Zero Day exploitation pressure," Zero Day Clock, accessed September 4, 2026.
[2] "Project Glasswing: An initial update," Anthropic, May 22, 2026.
[3] "The next big shifts in AI workloads and hyperscaler strategies," McKinsey, December 17, 2025.
[4] "Global electricity demand is set to grow strongly to 2030, underscoring need for investments in grids and flexibility," IEA, February 6, 2026.

conviction, yet closed deals are more of a signal than a forecast. We thus measured investment amounts in the first half of 2026 for each of the 14 trends and extrapolated current growth rates for the full year to provide a forward-looking view of where the trends are headed. Five of the trends—agentic software development, AI infrastructure and model architectures, AI for scientific discovery and engineering, the future of space technologies, and the future of robotics—are on track to receive more than double the investment in 2026 compared with 2025 (Exhibit 1). And all trends, except advanced connectivity, are on track to generate higher year-over-year funding totals. These findings show that investors are betting strongly on frontier technologies to deliver value over the coming years.

The 14 technology trends shaping 2026 show how quickly new innovations can shift business objectives, global markets, and ways of working. Not even the most talented technologists can predict every breakthrough. But business leaders who understand the patterns behind technological change will be equipped to shape the future rather than react to it.

Exhibit 1

Five technology trends are on pace to at least double their investment in 2026 compared with 2025.



Extrapolated change in investment, 2025–26, $ billion

The multiple difference between the 2025 investment and the extrapolated end-of-year 2026 investment

Note: Data includes private-market and public-market capital raises across venture capital and corporate and strategic M&A (including joint ventures), private equity investments (including buyouts and private investment in public equity), and public investments (including IPOs). Excludes corporate capital and operational expenditures.

[1] Investment trends in the first half of 2026 were extrapolated by applying a linear monthly run rate to deal activity through June 30, 2026. Projections may be skewed by megadeal concentration and calendar seasonality and are directional estimates rather than precise forecasts.

[2] Includes $60 billion Cursor acquisition by SpaceX (announced, pending close).

Source: PitchBook; McKinsey analysis

McKinsey & Company

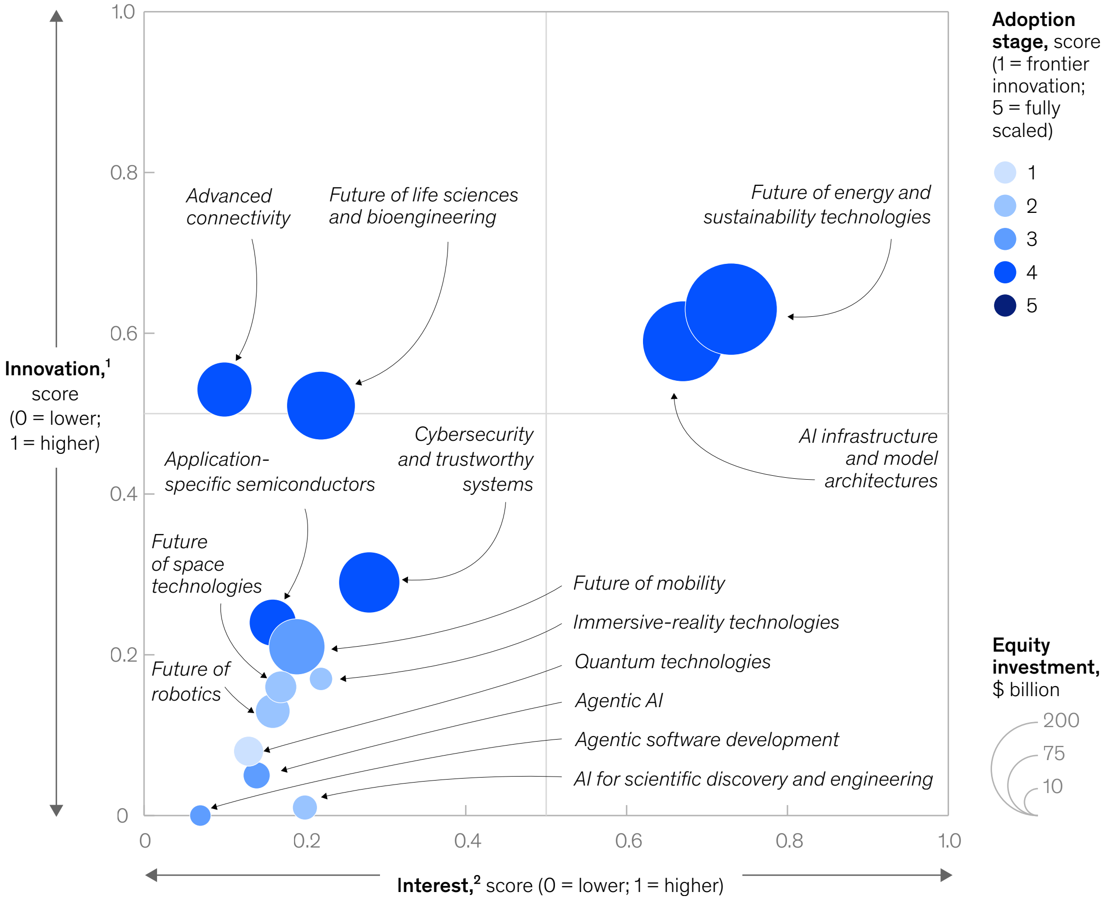
The 14 tech trends

This report examines each of the 14 trends in depth, and we invite readers to engage with the analysis for each technology and to explore the interplay among them. To present a visual overview of how these 14 trends intersect, we developed scores for innovation (based on patents and research publications) and interest (based on news and web searches). We also measured equity investment in each technology trend and rated its level of adoption by organizations (Exhibit 2).

Exhibit 2

Each technology trend is scored based on its level of innovation, interest, equity investment, and adoption.

Innovation, interest, investment, and adoption, by technology trend, 2025



Note: Innovation and interest scores for the 14 trends are relative to one another. All 14 trends exhibit high levels of innovation and interest compared with other topics and are also attracting significant investment.

[1] The innovation score combines the 0–1 scores for patents and research, which are relative to the trends studied. The patents score is based on a measure of patent filings, and the research score is based on a measure of research publications.

[2] The interest score combines the 0–1 scores for news and searches, which are relative to the trends studied. The news score is based on a measure of news publications, and the searches score is based on a measure of search engine queries.

McKinsey & Company

About the authors

Michael Chui
Senior fellow
Bay Area

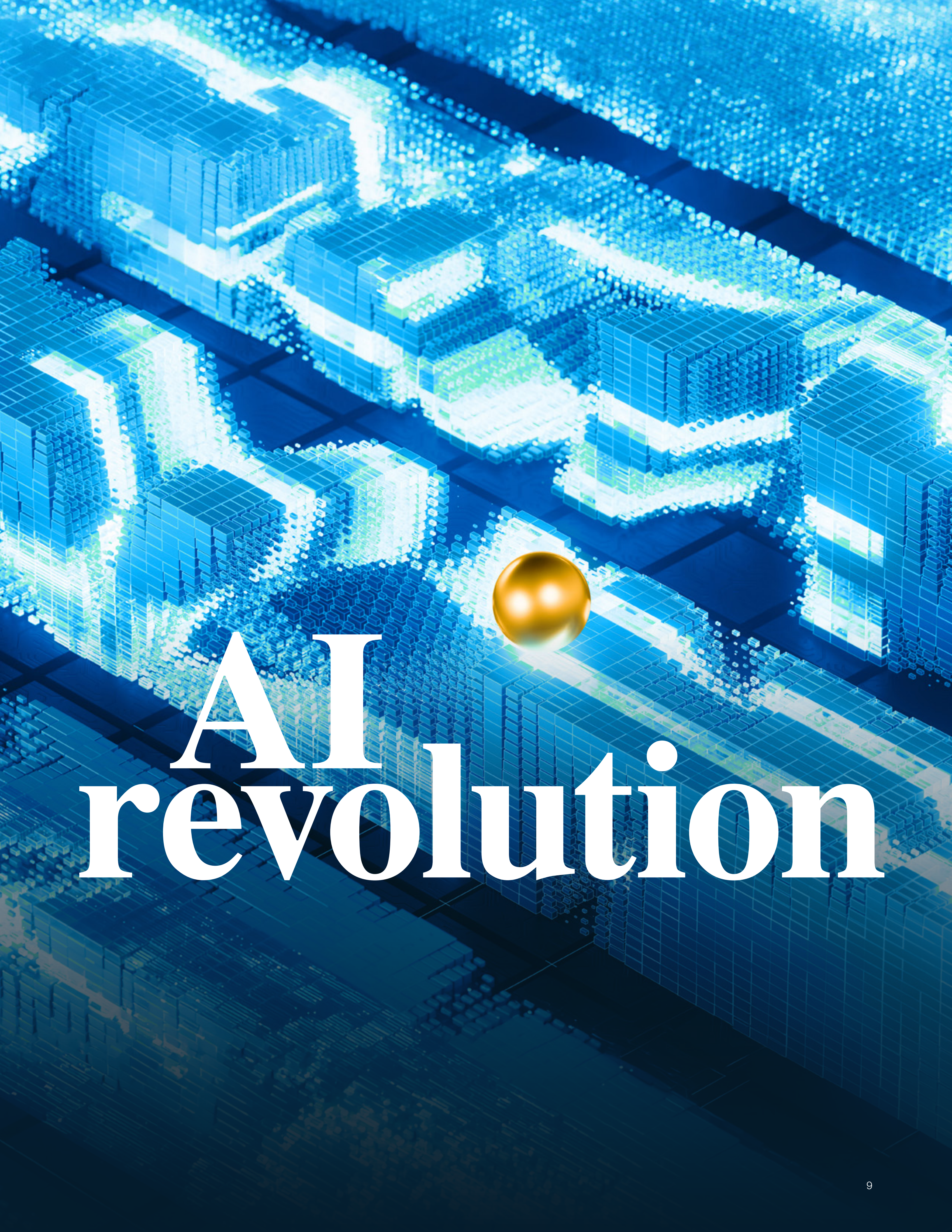
Roger Roberts
Partner
Bay Area

Tanguy Catlin
Senior partner
Boston

The authors wish to thank the following McKinsey colleagues for their contributions to this research:

Alex Devereson
Alex Peluffo
Alizee Acket-Goemaere
Andreas Breiter
Ani Kelkar
Anna Granskog
Anton Lysenko
Anuj Nigam
Ashley Park
Batu Demir
Berksu Durmaz
Bill Wiseman
Brendan Gaffey
Brooke Stokes
Charlie Lewis
Christian Jansen
Dan Tinkoff
Daniel Wallance
David Champagne
Dev Patel
Elizabeth Ouyang
Erika Stanzl
Fan Gao
Gaurav Agarwal
Gemma D'Auria
Hamza Khan
Henning Soller
Henry Marcil
Holli Dobay
Ian Whalen
Ichiro Otobe
Jeff Algazy
Jenna Brown
Jeremy Schneider
Julian Fuchs-Souchon
Kabir Ahuja
Kevin Catalon
Kevin Wei Wang
Lieven Van der Veken
Marc Sorel
Mark Patel
Martin Harrysson
Martin Wrulich
Martina Gschwendtner
Matt Higginson
Mingyu Guan
Naveen Sastry
Oana Cheta
Pankaj Sachdeva
Pepe Cafferata
Philipp Kampshoff
Prakhar Dixit
Pranita Sadavarte
Ruth Heuss
Sambhav Sharma
Sarah Abebe
Scott Smith
Sebastian Kluger
Shiv Shah
Shradha Pruthi
Stephen Xu
Steven Begley
Tara Balakrishnan
Thomas Devenyns
Tiffany Jenkins
Timo Möller
Vinayak HV
Vishal Agarwal
Wings Zhang
Yannan Collins
Yvonne Ferrier

Special thanks to McKinsey Global Publishing and Brand Marketing colleagues Dan Spector, Diane Rice, Drew Holzfeind, Eliza Cooper, Janet Michaud, Katie Shearer, Kristi Essick, Loriana Mitchell, Mary Gayen, Nancy Cohn, Richard Johnson, Ron Nurwisah, Sarah Thuerk, and Stephen Landau for making this report come alive.



AI revolution

01

Agentic software development

Agentic software development refers to the use of large language model (LLM)–powered systems that autonomously perform software engineering tasks, including planning, writing code, running tests, fixing bugs, and deploying changes. AI coding tools are shifting engineers from hands-on coders to agent supervisors.

The trend—and why it matters

In the past year, the practice of software development took a transformational turn. Instead of typing out lines of code, developers can now instruct their computers to create the applications they want. The highest-performing software engineers have become expert agent wranglers, spawning multiple agents that write, modify, and test code while they sleep.

Agentic software development is a new trend this year—and one that's truly transforming how work gets done. Companies are adopting agentic coding tools at a breakneck pace as they race to capture value from these new autonomous tools. There is a lot to gain. When deployed successfully, agentic software development could unlock almost $1 trillion in value for companies.[1]

This is easier said than done. Making individual engineers more productive is only part of the battle. Companies must also rewire the full product development life cycle (PDLC) using AI. The organizations most likely to capture that value are those supporting their developers with change management, embedding AI tools into workflows, and ensuring agentic outputs are appropriately verified and validated to meet high-quality standards.

Our recent research shows that just one-quarter of companies using agentic software development tools have achieved meaningful acceleration in their PDLCs, defined as more than a quarter of their teams achieving twofold or greater productivity gains.[2] Getting to this level of success requires nurturing top performers and encouraging other developers to adopt their winning practices; while 80 percent of software engineers using AI tools see a productivity acceleration of around 3 percent, the top 20 percent see a 55 percent boost.

Without a systematic approach, agentic software development can lead to unintended outcomes. Our research finds that in 30 percent of companies, productivity *fell* after teams began using agentic AI tools. Almost everyone, it seems, is "vibe coding"—but that doesn't always lead to value. In one study, AI tools increased coding activity by 180 percent, but shipped releases rose by only 30 percent, showing that more code doesn't always lead to more products.[3]

Developer acceptance of agentic software tools remains a meaningful constraint. Some 46 percent of developers worldwide actively distrust AI tools' accuracy, compared with 33 percent who trust them, and only 3 percent highly trust their outputs.[4] For companies that want to reap the benefits of agentic software development, that trust gap matters. The near-term objective is to build human–agent systems where people and machines work together to safely generate, test, and deploy code within controlled environments.

---

[1] "Beyond the copilot: Building and scaling the agentic product development life cycle," McKinsey, August 21, 2026.
[2] "Beyond the copilot: Building and scaling the agentic product development life cycle," McKinsey, August 21, 2026.
[3] Mert Demirer, Leon Musolff, and Liyuan Yang, "Writing code versus shipping code: Productivity effects across generations of AI coding tools," Centre for Economic Policy Research, June 21, 2026.
[4] "2025 Developer Survey," Stack Overflow, July 29, 2025.

AI REVOLUTION

Agentic software development

Scoring the trend

The trend's clearest signals are in capital and interest. Investment in companies providing agentic AI coding tools was negligible through 2023, began rising in 2024, reached roughly $5 billion in 2025, and has crossed $61 billion in the first half of 2026 (mainly driven by the $60 billion acquisition of Cursor).[1] Search activity is following the same path, climbing as developers and enterprises test these tools in real workflows. Together, the sharp rise in investment and search activity mark software development as the first agentic AI application where capital deployment and market interest are converging around tangible business returns. Innovation and news scores look modest partly because activity here leans more toward product work than patents and publications, and the topic remains relatively niche in the broader news media compared with other trends.

Equity investment, 2025
$4.9 billion

Job postings, 2024–25, % difference
+221%

Score, by vector (0 = lower; 1 = higher)

Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

[1] Sean O'Kane, "SpaceX to acquire Cursor for $60B in stock, days after blockbuster IPO," TechCrunch, June 16, 2026.

Latest developments

Recent developments involving agentic software development include the following:

— *Agentic AI coding tools are shifting from synchronous assistance to asynchronous execution.* Earlier agentic AI software coding tools required developers to interactively prompt chatbots in real time, similar to an extension of the code completion feature in software editors that would predict and insert the rest of a word, line, or block of code. Now the model is shifting toward asynchronous delegation. Developers can assign discrete tasks to autonomous agents, shifting software development from a typical two-week sprint cycle to a continuous high-speed loop. In practice, tools like Anthropic's Claude Code and OpenAI's Codex have accelerated the transition toward long-running software agents capable of working asynchronously across coding, debugging, testing, and documentation workflows.[5] Developers increasingly describe managing teams of agents rather than interacting with a single coding assistant. Organizations are also beginning to deploy after-hours workflows, in which agents operate asynchronously overnight or across weekends and return completed code, tests, analysis, or documentation by the next working session.[6] This fundamentally shifts the developer's role; humans become orchestrators who define workflows and apply judgment, while agents handle execution.

The shift toward autonomy, however, introduces new risks. Frontier models like Anthropic's Claude Mythos can autonomously discover and exploit vulnerabilities at machine speed. This fundamentally compresses the vulnerability life cycle and creates an asymmetry where attacks can overwhelm human-speed defenses. Managing agents in this new cybersecurity reality requires a multilayered control architecture that includes access control, isolation, monitoring, human oversight, containment, and recovery processes (for more, see Cybersecurity and trustworthy systems).

— *The 'agentic factory' is emerging as an engineering paradigm.* Cursor emerged as one of the clearest examples of the shift from an integrated development environment model toward agent-centric software orchestration. Rather than simply suggesting code, Cursor functions like a visual environment for spawning, monitoring, and coordinating autonomous coding agents.[7] At the same time, Anthropic, Cognition, OpenAI, and other AI-native developer tool companies have accelerated investment in long-running agents capable of supporting autonomous software development.[8] While asynchronous agents optimize individual coding tasks, the "agentic factory" can transform the entire software development life cycle (SDLC). This is important because developers spend only about one-tenth of their workday writing code, so organizations can deploy multiagent systems to automate more of the surrounding SDLC.[9] For example, one agent drafts product requirements, another designs the architecture, a third generates security tests, and a fourth manages the release pipeline. Our analysis confirms this strategy's impact: Top-performing organizations are six to seven times more likely to embed AI across four or more stages of the development life cycle.[10]

As coding gets automated, the bottlenecks in shipping a working product shift "to the edges": first, deciding what to build and then validating whether what gets built actually works. Product requirements can serve as directions for agents to follow, while testing, both human and automated, can determine whether outputs are secure and reliable. That means it's critical to embed agentic AI across the PDLC, not just into coding. Knowledge graphs can help companies build the foundations for agentic AI at scale.

> **'The real shift in software engineering is not that agents write code faster. To capture the value from these tools, the software development life cycle operating model has to shift, including moving toward smaller, highly leveraged teams that supervise agents through execution.'**
>
> — Martin Harrysson, senior partner, Silicon Valley

---

[5] Steven Rosenbush, "Long-running AI agents are here," *Wall Street Journal*, February 5, 2026.
[6] Jared Moon, Rory Walsh, Vito Di Leo, and Adam Thelwall, "Rewiring software delivery for the agentic era," McKinsey, May 28, 2026.
[7] "Build programmatic agents with the Cursor SDK," blog post by Roshan Sadanani, Cursor, April 29, 2026.
[8] Steven Rosenbush, "Long-running AI agents are here," *Wall Street Journal*, February 5, 2026; Steven Levy, "AI agents plunged the tech world into chaos. Here's exactly how that happened," *Wired*, May 26, 2026.
[9] Rudrajit Choudhuri et al., "To copilot and beyond: 22 AI systems developers want built," arXiv, April 2026, arXiv:2604.07830.
[10] "Unlocking the value of AI in software development," McKinsey, November 3, 2025.

They give agents a shared, consistent sense of what things mean across the enterprise, with definitions, lineage, and permissions built in. That makes it easier for agents to interpret requirements correctly and easier for teams to validate what comes out the other end.

- *AI-native development is emerging as a software engineering model.* What began as experimentation with coding assistants has evolved into a broad shift toward AI-first development workflows, where engineers increasingly assume AI support is a standard part of the software creation process. Rather than treating AI as a specialized productivity tool, organizations are embedding it directly into coding, testing, debugging, documentation, and code review workflows. In October 2025, GitHub reported that more than 1.1 million public repositories were using an LLM software development kit and that nearly 80 percent of new developers used Copilot within their first week on a job, underscoring how quickly AI-native practices are becoming embedded across the developer ecosystem.[11]

  In the earlier phase of adoption, access to agentic AI software development tools was a meaningful differentiator: Only some organizations had early access to the most capable models, secure deployment environments, or the technical capacity to make these tools usable inside workflows. Today's broader uptake has implications beyond engineering productivity; companies that embrace agentic software development can create meaningful competitive advantage. Chief technology officers no longer need to rely solely on their enterprise software partners to deliver every niche feature, since agentic AI empowers their teams to develop organization-specific capabilities. At the same time, they can build customized, enterprise-grade capabilities on top of software-as-a-service (SaaS) platforms—which are rapidly expanding their own agentic AI road maps—using the vendors' platform-native agentic AI tools.[12]

- *Enterprise context is becoming a core technical requirement.* Engineering teams are increasingly discovering that code generation is the easy part; the harder problem is making sure agents understand the company's codebase, internal policies, regulatory constraints, and product history. The challenge also differs by the type of software work being performed. In greenfield development, agents can often move faster because they are generating new systems with fewer inherited constraints. In brownfield development, however, agents must work inside existing codebases, architecture decisions, and dependencies. In modernization work, the bar is higher still: Agents may need to help refactor or translate legacy systems while preserving the operational processes those systems support. We find leveraging AI agents can accelerate modernization efforts by 40 to 50 percent and reduce costs by up to 40 percent. As a result, autonomous-agent deployment is creating demand for structured knowledge graphs, advanced repository indexing, and enterprise context systems that help agents navigate not only code but also policy, product history, and operational rules. This is already visible in enterprise experiments where highly documented workflows and searchable institutional knowledge create stronger conditions for agentic development.[13]

- *Software development economics are being rewritten.* As agents orchestrate more complex tasks, organizations are seeing token spend and inference costs evolve from a rounding error to a material operational line item. Some enterprises are beginning to treat software development economics as a mix of labor cost and AI-usage

---

[11] "Octoverse: A new developer joins GitHub every second as AI leads TypeScript to #1," GitHub, February 28, 2026.
[12] Steven Rosenbush, "Long-running AI agents are here," *Wall Street Journal*, February 5, 2026.
[13] Justin Young, "Effective harnesses for long-running agents," Anthropic, November 26, 2025.

cost, as token-based pricing and agentic coding workflows make model consumption a governed budget item rather than a background technical metric.[14] The shift is becoming material at the enterprise level: Our research suggests that AI now accounts for a third of companies' spending to introduce new capabilities, while also adding to technology run costs.[15] At the same time, AI inference introduces new variable costs, particularly for compute and infrastructure.[16] As a result, organizations are increasingly treating token consumption as a governed financial resource rather than a purely technical metric.

— *The talent model is shifting to smaller, highly leveraged engineering teams.* As AI absorbs repetitive implementation work, the structural makeup of the engineering workforce is changing. Organizations most advanced in adopting agentic operating models are moving toward leaner teams focused on product definition, system architecture, prompt engineering, and exception handling.[17] This requires significant upskilling, shifting the organizational structure away from large developer pools toward small, highly leveraged talent pods that use AI agents to multiply their output. Traditional agile pods are small human teams—the "two pizza" model—tackling a list of backlog tasks. Human–agent pods contain many agents executing tasks under the oversight of a small number of humans.[18] Engineers still define the product, architecture, and constraints, but agents do the implementation and testing.

The trend is already influencing workforce planning. Salesforce announced that it would not hire additional software engineers in 2025 after reporting substantial productivity gains from AI-assisted development, signaling how organizations are beginning to prioritize AI leverage over head count growth.[19] More broadly, companies are increasingly experimenting with human–agent team structures in which a small number of engineers supervise and coordinate fleets of AI agents rather than performing all implementation work directly.

> **'Asynchronous agents—and workflows—change the cadence of software delivery. Work that once moved through two-week sprints can now become a continuous loop of drafting, testing, debugging, and documentation between human reviews. The leadership challenge is to increase speed without weakening maintainability, quality, or control.'**
>
> — Prakhar Dixit, partner, Seattle

---

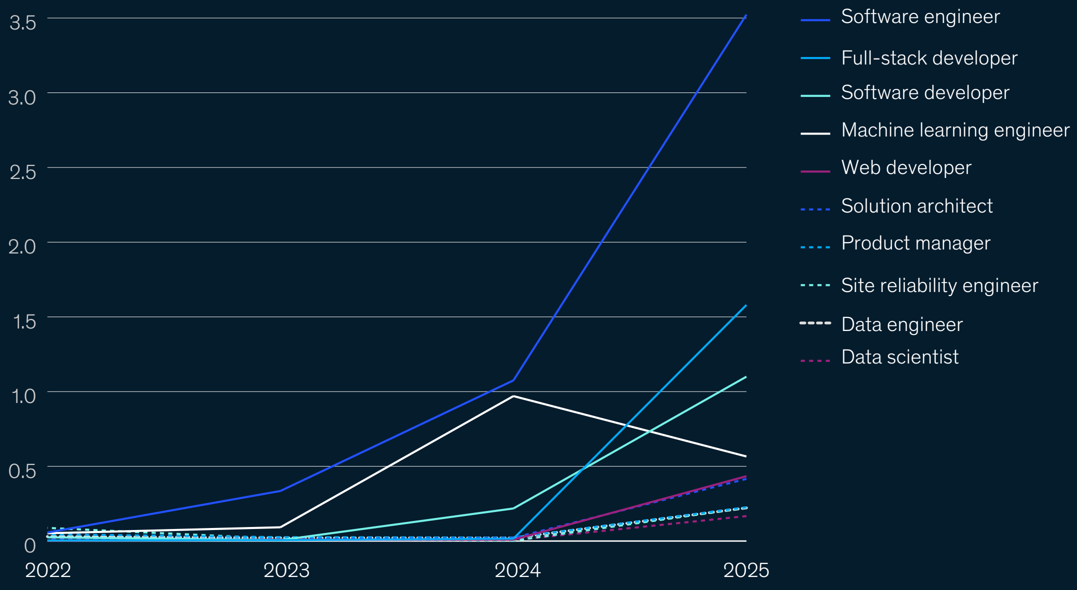
[14] Longju Bai et al., "How do AI agents spend your money? Analyzing and predicting token consumption in agentic coding tasks," arXiv, April 29, 2026, arXiv 2604.22750.
[15] Leandro Santos, Thomas Elsner, and Ishaan Sharma, "Recalibrating technology budgets for the AI era," McKinsey, March 30, 2026.
[16] "The cost of intelligence: How CIOs can manage AI demand at scale," McKinsey, July 20, 2026.
[17] Henning Soller, Mike Stefanelli, Rajat Goel, and Umar Husain, "Designing an end-to-end technology workforce for the AI-first era," McKinsey, April 6, 2026.
[18] Isabelle Bousquette, "Your work team is now a 'pod' and your co-workers are AI agents," *Wall Street Journal*, May 18, 2026.
[19] Katherine Li and Aditi Bharade, "Marc Benioff said Salesforce is still hiring in one department, and it's not in engineering," Business Insider, May 28, 2026.

Agentic software development

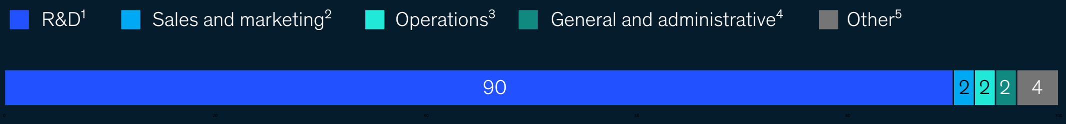
Demand

Demand for agentic software development talent has surged, with total postings more than tripling from 2024 to 2025, making it one of the fastest-growing trends year over year. Software engineers remain the largest role, while full-stack developers rose sharply from a negligible base to become the second largest. Machine learning engineer postings, by contrast, fell sharply, suggesting the work is shifting from training models to wiring existing ones into software. The functional mix is weighted very heavily toward R&D, reflecting a field still centered on technical building and innovation.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).

[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).

[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).

[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).

[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

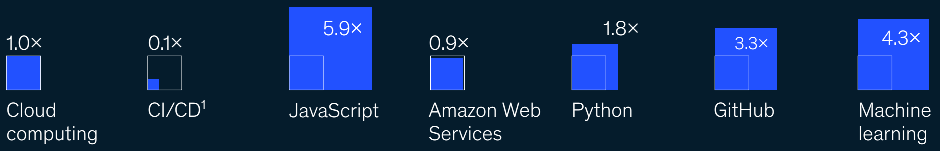
Agentic software development

Skills availability

Cloud computing skills are required in most postings, and talent supply just keeps pace with demand. The sharpest shortage sits in continuous integration and continuous delivery (CI/CD), the delivery skills needed to move agentic software into production. By contrast, more established capabilities such as JavaScript and Python, along with GitHub and even machine learning, are widely available.

Talent required, % share of postings requiring skill ■

Talent availability, ratio of talent ■ to demand □



[1] Continuous integration and continuous delivery.

Adoption developments across the globe

Adoption score: 3—Piloting. Most organizations have begun piloting agentic software development, and more are moving into scaling. Some fast movers have already fully implemented agentic software tools as their core product development workflow. Technology companies remain the furthest along, while legacy industries are still working through the people, process, and governance changes needed to make agentic software development operational.

North America remains the center of activity, driven by the uptake of agentic AI–driven software development at AI-native software companies, hyperscalers, developer tool providers, and large technology companies. Many of the most advanced deployments are happening in software engineering, testing, deployment, orchestration, and internal productivity environments.

Europe is progressing more cautiously, with stronger emphasis on governance, compliance, and software quality controls. Adoption is advancing, but often only in select use cases and with greater human oversight.

Adoption in the Asia–Pacific region is growing, particularly where software development is closely tied to digital transformations, industrial workflows, or large-scale enterprise modernization programs.

Across sectors, software companies are moving fastest because they have strong digital infrastructure, flexible workflows, and tolerance for experimentation. Companies with codified processes, structured documentation, and searchable knowledge could also move quickly with agentic software development. By contrast, industries with fragmented environments, such as retail, manufacturing, oil and gas, and mining, are generally moving more slowly due to less mature digital infrastructures.

Underlying technologies

The technologies that power agentic software development include the following:

Agent evaluations	These are frameworks that benchmark, test, and monitor agent performance across coding, review, debugging, security, and deployment tasks.
Coding agents	Systems that can autonomously write, review, refactor, and debug code across software environments are known as coding agents.
CI/CD integration layers	While the coding agent performs the development work, the CI/CD infrastructure layer enables agents to interact with testing, deployment, and release pipelines.
Enterprise context systems	Knowledge graphs, document stores, and retrieval systems help agents understand code, policy, and product history.
Foundation models	Large AI systems such as foundation models provide code generation, planning, and reasoning capabilities.
Harness and guardrail systems	These are controls that define what information and processes agents can access, change, or execute.
Orchestration frameworks	The tools that coordinate multiple agents across tasks and handoffs are known as orchestration frameworks.
Reinforcement-learning verifiable reward systems	Mechanisms such as these can convert objectively checkable outcomes, such as passing tests or satisfying specifications, into feedback signals to improve agent performance over time.
Token and workload management systems	These platforms monitor inference usage, cost, and run time efficiency.

Key uncertainties

The major uncertainties affecting agentic software development include the following:

— *Trust and reliability.* Organizations are still working out how to ensure agents produce safe, correct, and maintainable code.

— *Governance and control.* What agents should be allowed to change, how much autonomy they should have, and how their work should be audited remain open questions.

— *Economic returns.* Many organizations can show developer-level productivity gains, but fewer have yet to translate those gains into sustained organizational-level financial impact.

— *Talent redesign.* The long-term shape of engineering roles, performance management, and team structure remains unclear.

— *Cost discipline.* Token spend, tool sprawl, and orchestration overhead may limit returns if not managed carefully.

— *Legacy integration.* Older systems, compliance requirements, and a fragmented enterprise context continue to slow adoption.

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with agentic software development:

- How should organizations balance speed, governance, and maintainability as agents take on more of the SDLC?
- Which parts of the SDLC will be genuinely delegated to agents at scale?
- How should enterprises measure productivity when token spend becomes part of development economics?
- What is the right team structure for an AI-first future-state software organization?
- Which roles become more valuable as agents take on implementation work?

Read more about agentic software development

Knowledge center

Enterprise Software

Insights on Artificial Intelligence

Related reading

The state of AI in 2026: On the road to ROI

AI-native development: Rethinking software from the ground up

Rewiring software delivery for the agentic era

Unlocking the value of AI in software development

The AI-centric imperative: Navigating the next software frontier

02

Agentic AI

Agentic AI refers to AI systems that can plan and execute multistep workflows with limited human direction.[1] They reason, interact with enterprise systems, coordinate with other agents, and autonomously adjust their actions as context evolves.

The trend—and why it matters

Over the past 12 months, agentic AI has moved beyond experimentation to become a major focus of enterprise technology investment. Our research finds that 89 percent of organizations regularly use AI, with the majority at least experimenting with agents.[2] Yet measurable value remains elusive. We find that only 37 percent of companies attribute any positive EBIT impact from their AI programs, let alone from newer agentic deployments.

AI agents are adding complexity and costs to enterprise technology budgets. Agents require upgrades to the existing tech stack to work reliably, and they consume a large number of compute tokens. That's because agentic workflows are more compute intensive than traditional chatbot interactions, relying on repeated reasoning loops, tool calls, retrieval steps, and orchestration across systems.[3] In some enterprise environments, a single agentic workflow can consume five to 30 times more tokens than a standard chatbot query.[4] With agents becoming central to many enterprise environments, 93 percent of respondents to a recent McKinsey survey report exceeding their AI budgets.[5]

However, agentic AI is relatively nascent, and the value companies derive from it will likely expand as the technology becomes more technically capable and more embedded into enterprise software. Recent agent-building tools from Anthropic and OpenAI have improved tool use and computer interaction, while frameworks such as LangChain's LangGraph are adding orchestration and human-in-the-loop controls needed to guide agents through longer workflows.[6]

As agentic functionality improves, companies are beginning to incorporate agents into their workflows and measuring the value they create. Software development has emerged as one of the early areas of agentic adoption (for more, see Agentic software development). Meanwhile, broader experimentation with AI agents is underway across business functions.

The agentic shift is also changing the way companies approach critical budget, operating, and talent challenges. Agents are raising questions about buy-versus-build decisions for software deployments; how companies hire, train, and distribute labor; and how operating models are designed for human–agent collaboration. AI agents are also growing into a connective layer between fragmented legacy systems—an agentic mesh that can help companies get more value out of their existing technology deployments. When agents

---

[1] This trend covers the entire field of agentic AI and its use across sectors and applications; we cover agentic software development as its own trend.

[2] "The state of AI in 2026: On the road to ROI," McKinsey, August 25, 2026.

[3] "The next big shifts in AI workloads and hyperscaler strategies," McKinsey, December 17, 2025.

[4] "Gartner predicts that by 2030, performing inference on an LLM with 1 trillion parameters will cost gen AI providers over 90 percent less than in 2025," Gartner press release, March 25, 2026.

[5] "The cost of intelligence: How CIOs can manage AI demand at scale," McKinsey, July 20, 2026.

[6] "Introducing advanced tool use on the Claude Developer Platform," Anthropic, November 24, 2025; "The next evolution of the Agents SDK," OpenAI, April 15, 2026; "LangChain and LangGraph agent frameworks reach v1.0 milestones," LangChain Changelog, October 22, 2025.

move from passive assistance to active execution within the tech stack, it could give companies a new avenue to reduce tech debt and extract more ROI from their enterprise spend.[7]

One of the clearest indicators of progress is agents' growing ability to handle longer, more complex tasks end to end. The nonprofit organization Model Evaluation and Threat Research (METR) benchmarks "time horizons," which are the maximum lengths of time an AI can work independently on a task before it's likely to fail or need human assistance. Recent benchmarks show that the best AI models can now successfully complete tasks that would take a skilled human many hours. For example, METR estimates that Claude Opus 4.6 had a 50 percent time horizon (meaning it's predicted to succeed half the time) of roughly 12 hours on software engineering tasks by May 2026.[8]

Many companies today have deployed point agents designed to handle specific workflows or business processes with greater autonomy. A smaller set of digitally mature organizations is now moving toward orchestrated end-to-end workflows, where multiple agents coordinate across systems and tasks. This pattern is particularly visible in software development, where companies are increasingly embedding agents across the product development life cycle.

The implications of the agentic organization extend beyond technology advances and productivity gains. Agents also fundamentally change how companies work and create value. The biggest gains will come not from layering agents onto existing processes but from redesigning workflows and operating models around humans and AI agents working together.[9] Doing so responsibly will require companies to implement new governance and security measures because agentic systems interact across multiple enterprise systems, increasing operational risk.[10] They will also need to reimagine their technology infrastructures for and with agentic AI.

> **'Today's challenge is operationalizing judgment. The defining question of the agentic era is not how autonomous agents can become but how much autonomy the enterprise can safely absorb.'**
>
> — Oana Cheta, partner, Chicago

---

[7] "Triple the return: How companies can get more from enterprise tech," McKinsey, October 20, 2025.
[8] "Task-completion time horizons of frontier AI models," METR, updated May 8, 2026.
[9] "Six shifts to build the agentic organization of the future," McKinsey, October 27, 2025.
[10] Chandrasekhar Panda and Henning Soller, "Stable, secure, and scalable: How AI is redefining technology resilience," McKinsey, May 12, 2026.

AI REVOLUTION

Agentic AI

Scoring the trend

Agentic AI emerged as a distinct field when AI systems were given the capability to take autonomous actions, and the trend advanced rapidly between 2022 and 2025. Google searches and news mentions both rose meaningfully over the period, reflecting surging interest as enterprises explore autonomous AI agents. Equity investment is also increasing quickly off a small base, with capital raises reaching $9.9 billion in 2025 and already $6.7 billion by mid-2026, showing conviction in this emerging field.

Equity investment, 2025
$9.9 billion

Job postings, 2024–25, % difference
+952%

Score, by vector (0 = lower; 1 = higher)

Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

Latest developments

Recent developments involving agentic AI include the following:

- *Enterprise AI is becoming focused on process orchestration.* Capturing meaningful value from agents requires system-level integration. Consequently, companies are beginning to invest in agentic orchestration layers that span enterprise resource planning (ERP), customer relationship management (CRM), and supply chain systems.[11] This shift is increasingly visible in enterprise software products. Salesforce has expanded Agentforce deployments that allow enterprises to configure AI agents for sales operations and CRM orchestration, while Microsoft has expanded Copilot Studio and related orchestration tooling to help organizations build custom agents capable of interacting across productivity systems and enterprise applications.[12]

- *Multiagent orchestration frameworks are the new enterprise middleware.* Rather than relying on a single monolithic AI model for their agentic deployments, organizations are using

---

[11] “Gartner forecasts supply chain management software with agentic AI will grow to $53 billion in spend by 2030,” Gartner press release, April 7, 2026.

[12] “Agentforce sales: Agents handle the grind, sellers focus on the win,” Salesforce, March 16, 2026; *Microsoft Copilot Blog*, “New and improved: Computer-using agents, a new workflows experience, and real-time voice experiences,” blog entry by Nitasha Chopra, May 26, 2026.

> ‘The key to unlocking the value of agentic AI is evaluations. You have to invest in creating evals that align with your business objectives. Then you can create the improvement loops that accelerate business performance.’
>
> — Stephen Xu, partner, Chicago

coordinated “teams” of specialized agents. Open-source frameworks such as LangGraph and CrewAI have matured into enterprise-grade solutions, allowing a “manager” agent to intelligently route subtasks to domain-specific agents. LangGraph’s v1.0 release signaled production-ready agent execution, while CrewAI’s recent releases added capabilities like OpenTelemetry support, better planning and observation handling, and Snowflake/Databricks integrations.[13] These frameworks are helping multiagent systems move from demos to enterprise deployment.

— *Agent integration with legacy systems is an emerging bottleneck.* Companies are uncovering challenges integrating AI agents with application programming interfaces (APIs) and legacy systems. As large language models improve, the harder problem is safely turning model outputs into action inside regulated enterprise environments, especially where agents must connect through APIs and legacy ERP systems. To address that challenge, companies are adopting emerging integration standards such as the Model Context Protocol and rearchitecting around machine-readable interfaces so agents can share context and take action across tools and platforms without creating undue technical debt or risk.[14]

— *Agentic AI presents new compute and token economics challenges.* Agentic AI systems generally require far more compute capacity than chatbots, making token costs a key constraint alongside model capability.[15] Recent reporting shows large companies capping internal AI use as costs spiral upward.[16] While industry groups like the Linux Foundation are moving to formalize “tokenomics” as a discipline for AI cost management, infrastructure vendors are focused on token throughput, cost per token, and inference efficiency as core priorities for the agentic era.[17]

— *Agentic AI requires new frameworks for cybersecurity.* Agentic AI is forcing companies to rapidly adopt a new identity security playbook. As agents transition to read/write operators within enterprise databases, they introduce significant new attack surfaces by spawning nonhuman identities (NHIs) such as API keys, service accounts, and machine credentials.[18] With NHIs now vastly outnumbering human identities, chief information officers may need to deploy agentic governance frameworks to secure their organizations’ technology infrastructure.[19] For example, Okta recently launched a blueprint for the “secure agentic enterprise,” focused on agent registration and access standardization.[20] The shift reflects a broader reality: As agents begin operating across apps, APIs, and data systems, organizations are treating them as a new class of identity-and-access risk rather than just another software tool (for more, see Cybersecurity and trustworthy systems).

---

[13] “LangChain and LangGraph agent frameworks reach v1.0 milestones,” blog entry by Sydney Runkle et al., LangChain, October 22, 2025; “Changelog: Product updates, improvements, and bug fixes for CrewAI,” CrewAI, April 7, 2026.

[14] Bjørnar Jensen, Darwin Deano, Michael Allison, and Talha Bin Asad, “Bridging the great AI agent and ERP divide to unlock value at scale,” McKinsey, January 9, 2026.

[15] Longju Bai et al., “How do AI agents spend your money? Analyzing and predicting token consumption in agentic coding tasks,” arXiv, April 2026, arXiv:2604.22750.

[16] “The cost of intelligence: How CIOs can manage AI demand at scale,” McKinsey, July 20, 2026; Jamie John, Rafe Rosner-Uddin, and Ryan McMorrow, “‘We created a monster’: Companies rein in AI usage as costs strain budgets,” *Financial Times*, June 18, 2026.

[17] “Linux Foundation announces the intent to launch the Tokenomics Foundation to establish open standards for AI cost management,” Linux Foundation press release, June 3, 2026.

[18] “The state of non-human identity and AI security,” Cloud Security Alliance, January 26, 2026.

[19] Jack Hidary, “Non-human identities: Agentic AI’s new frontier of cybersecurity risk,” World Economic Forum, October 15, 2025.

[20] “Okta announces new blueprint for the secure agentic enterprise,” Okta press release, March 16, 2026.

— *Enterprise workflows are emerging as a leading place to embed agents.* Rather than being deployed as stand-alone applications, AI agents are increasingly being embedded into enterprise workflows. Announced in June, Morgan Stanley is opening its wealth management funnel to AI agents from corporate clients, allowing those agents to connect directly to stock administration platforms such as Shareworks and Equity Edge and pull data and insights without navigating human-oriented interfaces.[21] The bank has already granted early access to a handful of clients and plans to expand access to its administration base next year, underscoring how agentic adoption is starting to show up inside core business workflows rather than as a separate product layer.

— *Agents are being given autonomy to perform increasingly complex multistep actions.* Agentic systems are increasingly capable of navigating software interfaces, interacting with applications, and completing tasks autonomously. In May, Microsoft introduced new computer-using agents within Copilot Studio, including adaptive solutions to automate workflows.[22] OpenAI has embedded into its ChatGPT the capabilities first demonstrated in Operator, which was an agent that could go to the web to perform tasks and which demonstrated how AI agents can interact directly with digital interfaces, execute browser-based actions, and coordinate multistep workflows across software environments.[23] These capabilities point toward a future in which agents can perform work across applications without requiring custom integrations for every workflow, further accelerating enterprise automation.

— *Agentic commerce is emerging as a global opportunity.* The growing market for agentic commerce enables AI agents to autonomously make purchases for customers and market products for retailers. By 2030, the global retail market could see as much as $5 trillion in orchestrated revenue from agentic commerce.[24] Industry leaders in China are showing early examples of what that can look like in practice. Alibaba has begun rolling out a Qwen-powered digital workforce for millions of Taobao and Tmall merchants, with capabilities such as customer service and real-time price adjustments. Reports also suggest that Qwen is being integrated into Alibaba's consumer ecosystem, so users can ask it to carry out tasks such as ordering milk tea and booking flights.[25] This reflects a broader shift toward intent-based purchasing, where consumers state desired outcomes rather than browse listings and allow AI agents to handle comparison and purchase.[26] Payment infrastructure is adapting in parallel: Mastercard recorded its first agentic transaction in 2025, and American Express introduced a dedicated agentic commerce developer kit in April 2026.[27] Separately, Google's Agent Payments Protocol (AP2), launched in September 2025, lets a purchase be cryptographically verified against what the user authorized.[28]

— *Open-source agents are moving from the cloud to local machines.* A growing wave of open-source agent frameworks is demonstrating how AI agents can operate directly on personal computers rather than relying entirely on cloud-hosted environments. A recent example came during the OpenClaw surge, when users reportedly sought out Mac minis as power-efficient local machines for testing and running on-device AI agents.[29] By using local machines as the execution platform, these agent architectures can improve latency, privacy, autonomy, and offline reliability while expanding the range of workflows agents can control across desktop applications and operating systems.[30] This shift also signals a broader move toward more persistent, device-native AI assistants capable of interacting directly with users' local environments.

---

[21] Hugh Son, "Morgan Stanley will soon open its trillion-dollar wealth management funnel to AI agents," CNBC, June 3, 2026.

[22] *Microsoft Copilot Blog*, "New and improved: Computer-using agents, a new workflows experience, and real-time voice experiences," blog entry by Nitasha Chopra, May 26, 2026.

[23] "Introducing Operator," OpenAI, January 23, 2025.

[24] Katharina Schumacher, Roger Roberts, and Katharina Giebel, "The agentic commerce opportunity: How AI agents are ushering in a new era for consumers and merchants," McKinsey, October 17, 2025.

[25] Tracy Qu, "China's Alibaba links Qwen AI app to vast consumer ecosystem," *Wall Street Journal*, January 14, 2026.

[26] Cameron Emanuel-Burns, "Agentic commerce in 2026: Where we stand and what lies ahead," *FinTech Futures*, April 1, 2026.

[27] "Ecommerce trends: How Visa and Mastercard approach agentic commerce," Digital Commerce 360, April 2, 2026.

[28] *Google Cloud Blog*, "Powering AI commerce with the new Agent Payments Protocol (AP2)," blog entry by Stavan Parikh and Rao Surapaneni, September 16, 2025.

[29] Sarah Perez, "Marked-up Mac minis flood eBay amid shortages driven by AI," TechCrunch, April 24, 2026.

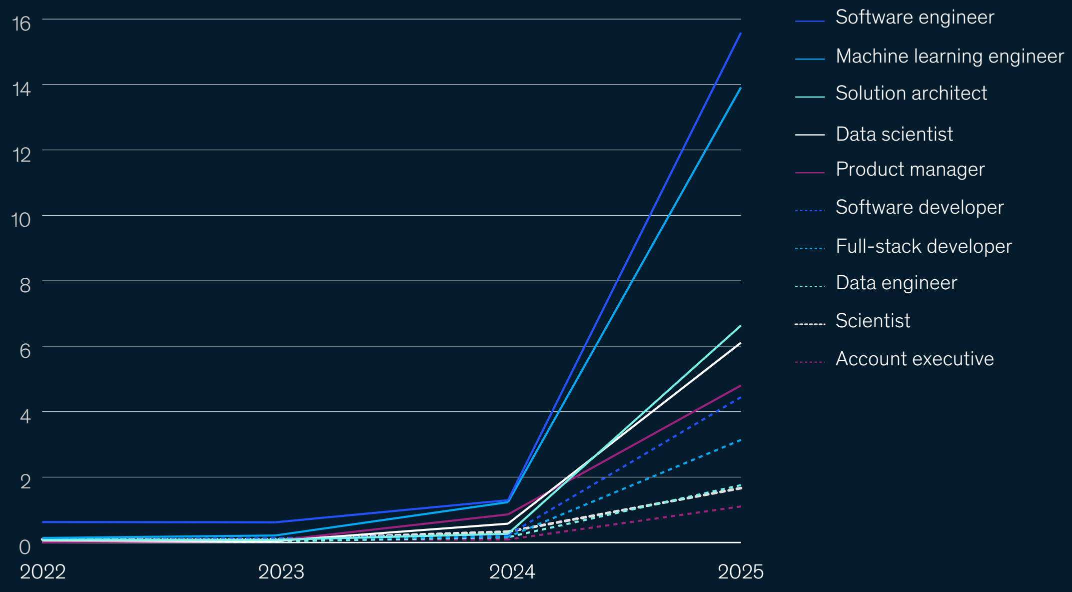
[30] Yuning Zhang et al., "AgentServe: Algorithm-system co-design for efficient agentic AI serving on a consumer-grade GPU," arXiv, March 11, 2026, arXiv:2603.10342.

Agentic AI

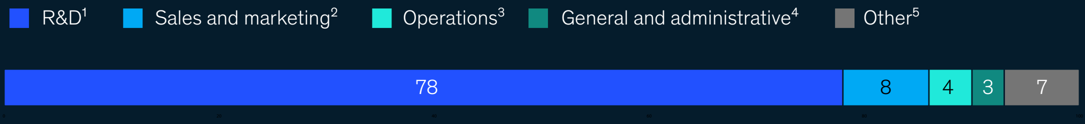
Demand

Demand for agentic AI talent has grown dramatically, with job postings rising roughly tenfold from 2024 to 2025 off a small base, placing it among the fastest-growing trends by hiring. Growth has been led by technical and research roles, particularly software engineers, machine learning engineers, data scientists, and solution architects, reflecting a field still centered on building and refining the underlying systems. Product managers and early commercial roles such as account executives have also begun to appear, an initial sign of movement from pure research toward productization and bringing offerings to market. However, a technical focus on early innovation is clear in the functional mix. Agentic AI hiring leans heavily toward R&D, with a much smaller share of postings for sales, operations, and general roles.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).

[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).

[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).

[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).

[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

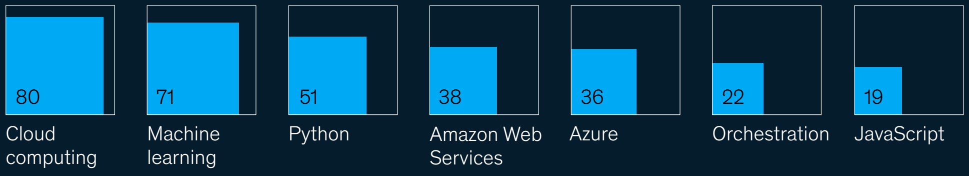
TALENT AND LABOR MARKETS

Agentic AI

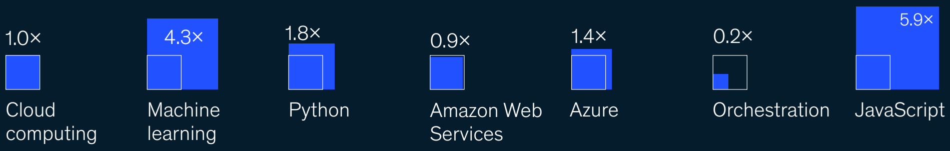
Skills availability

Cloud computing is the most widely required skill, appearing in roughly four in five postings, with supply broadly keeping pace. The shortage is in orchestration, the coordination layer that ties models, tools, and workflows into functioning agentic systems. The supply in more established capabilities, such as JavaScript and Python, and even machine learning, suggests a broad AI talent pool has formed while the newer skills needed to coordinate agents remain scarce.

Talent required, % share of postings requiring skill ■



Talent availability, ratio of talent ■ to demand □



Adoption developments across the globe

Adoption score: 3—Piloting. Many organizations are testing the functionality and viability of agentic AI with small-scale prototypes and targeted deployments. While enthusiasm and investment have increased sharply, relatively few enterprises have fully scaled agentic operating models.

North America remains the center of enterprise experimentation, driven by hyperscalers, frontier model developers, AI-native start-ups, and large technology organizations. Many of the earliest scaled deployments have occurred in software engineering, customer operations, and enterprise productivity workflows.

China is increasingly integrating agentic capabilities into industrial automation, enterprise software, and consumer ecosystems, particularly where AI systems can be tightly integrated into broader digital platforms. This is illustrated by Moonshot AI's July 2026 release of Kimi K3, a 2.8-trillion-parameter open-weight model that Moonshot says delivers performance approaching Anthropic's frontier Fable model, signaling how fast Chinese open-weight AI is approaching the performance of leading frontier models.[31]

European organizations are adopting agentic AI more cautiously because of regulatory scrutiny, governance requirements, labor considerations, and stronger concerns about operational accountability and data privacy.

Across industries, digitally mature sectors with strong data infrastructure and cloud-native operating environments, including technology, software, financial services, and telecommunications, are progressing fastest. Many large enterprises outside the technology sector remain in earlier experimentation phases with agentic AI due to the complexity of integrating agents into legacy systems. Fragmented data and governance issues are also constraints.

---

[31] Laurie Chen, "China's Moonshot unveils world's largest open AI model, closing in on US rivals," Reuters, July 17, 2026.

Underlying technologies

The technologies that power agentic AI systems include the following:

AI-enabled user interfaces	Conversational and multimodal interfaces allow users to manage workflows through natural language interaction.
AI orchestration frameworks	These systems coordinate multiple agents, tools, workflows, and enterprise integrations.
API integration layers	This infrastructure enables agents to interact with enterprise software, databases, and external services.
Foundation models	Large-scale AI models provide reasoning, generation, and planning capabilities for enterprise agents.
Memory and context management systems	These technologies help agents retain context and maintain continuity across workflows.
Reasoning models	Such models optimize multistep planning, sequencing, and workflow execution.
Retrieval-augmented generation	These architectures combine language models with enterprise knowledge retrieval systems.
Workflow management systems	These platforms manage task routing, approvals, execution, and operational monitoring.

Key uncertainties

The major uncertainties affecting agentic AI include the following:

— *Reliability and trust.* Organizations continue to face challenges ensuring that agentic systems consistently execute workflows accurately and safely.

— *Governance and accountability.* Questions remain regarding responsibility, oversight, escalation paths, and operational control when AI agents execute complex workflows autonomously.

— *Cybersecurity risks.* Agentic systems capable of interacting across enterprise systems create new attack surfaces and operational vulnerabilities.

— *Workforce transformation.* The long-term implications of agentic AI for organizational structures, management layers, role design, and labor demand remain uncertain.

— *Infrastructure economics.* The compute requirements and operational costs associated with large-scale agentic systems may constrain adoption.

— *Enterprise integration complexity.* Legacy systems, fragmented data environments, and operational silos continue to create implementation challenges.

— *Regulatory evolution.* Governments and regulators are still developing frameworks for autonomous AI systems, workflow automation, liability, and operational accountability.

— *Human–machine collaboration models.* Organizations are still determining how best to structure workflows around human workers and increasingly autonomous AI systems.

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with agentic AI:

- Which enterprise workflows are best suited for agentic orchestration, and how can enterprises ensure transparency, reliability, and accountability in multistep autonomous workflows?
- How should organizations rethink infrastructure, operating models, governance structures, and workforce strategies around agentic systems?
- What new organizational roles and skill sets will emerge as agentic systems become more deeply integrated into enterprise operations?
- Which industries will realize the greatest productivity and operational gains from agentic AI over the next five years?
- How will agentic AI reshape software interfaces, enterprise applications, and the broader structure of digital work?

Read more about agentic AI

Knowledge center

Tech & AI Insights

Related reading

The state of AI in 2026: On the road to ROI

Is that AI agent worth it? Agentic economics and the modern operating model

Building the foundations for agentic AI at scale

Agentic AI explained: When machines don't just chat, but act

Seizing the agentic AI advantage

03

AI for scientific discovery and engineering

AI for scientific discovery and engineering uses AI to accelerate research across fields such as life sciences, materials science, chemistry, physics, climate science, energy, geoscience, astronomy, and engineering. Models, simulations, and agentic workflows can help scientists and engineers generate hypotheses, design experiments, and automate parts of their work.

The trend—and why it matters

AI for scientific innovation is a nascent field with growing interest in its potential to augment discovery. R&D organizations that initially adopted AI for prediction and modeling are beginning to experiment with more autonomous, agentic AI-enabled research workflows.[1] For example, although medicines discovered by AI so far have limited clinical outcomes, many life sciences companies are betting heavily on this discovery pathway. Across the scientific spectrum, AI is being used to optimize molecules, prioritize product candidates, and explore engineering design spaces in ways that can feed real-world pipelines. These efforts could eventually help accelerate scientific discovery and reduce the time required to bring new products to market. Our analysis shows that AI could dramatically boost R&D productivity in some steps of the scientific discovery value chain, including expanding the number of hypotheses that can be tested, accelerating the pace at which they can be evaluated “in silico,” and improving the odds of selecting promising candidates.[2]

Some notable examples have emerged. AlphaFold demonstrated how deep learning models could solve long-standing scientific challenges in protein structure prediction and helped establish AI as a credible tool for scientific research.[3] More recently, organizations are beginning to explore how advances in physical AI, large language models (LLMs), and autonomous experimentation can be applied to broader scientific and engineering workflows.

AI is becoming increasingly embedded in life sciences R&D, with AI-powered data analysis accelerating pharma work from discovery through clinical trials. There is much to gain. Bringing a new therapeutic drug to market still often takes more than ten years and costs an average of $1.3 billion to $2.6 billion, while failure rates remain high across clinical pipelines.[4] To speed up this process, organizations are beginning to deploy AI models and orchestration systems at both the scientific and operational levels, supporting activities such as experiment design, molecular optimization, toxicity prediction, and candidate generation and evaluation, as well as accelerated research.[5] They are also using AI to optimize clinical trial design, with the goal of improving both the speed of discovery and the probability of scientific success. AI is being applied across the discovery value chain—from generating and optimizing therapeutic candidates to evaluating safety, efficacy, and manufacturability before physical testing begins, as well as accelerating research operations such as experiment design, literature review, and data analysis.[6]

The AI-enabled drug development pipeline is expanding, with recent industry reporting pointing to dozens of AI-discovered or AI-assisted candidates in development globally.[7] Recent announcements from Anthropic underscore how quickly this market is shifting from general-purpose

---

[1] “Scientific AI: Unlocking the next frontier of R&D productivity,” McKinsey, January 15, 2025.
[2] “The next innovation revolution—powered by AI,” McKinsey, June 20, 2025.
[3] *Google DeepMind*, “AlphaFold: Five years of impact,” blog entry by Demis Hassabis et al., November 25, 2025.
[4] “The next innovation revolution—powered by AI,” McKinsey, June 20, 2025.
[5] Dinh Long Huynh et al., “AI agents in drug discovery: applications and case studies,” *Drug Discovery Today*, March 2026, Volume 31, Number 3.
[6] Aditya Pathak et al., “AI-enabled drug and molecular discovery: computational methods, platforms, and translational horizons,” *Discover Molecules*, December 2025, Volume 2, Number 32.
[7] “First AI designed drug in human trials,” blog entry by Sanksshep Mahendra, Artificial Intelligence +, updated on May 14, 2026.

AI assistance toward dedicated scientific AI workflows. The company recently launched Claude Science, a beta AI workbench for scientists, and announced an internal AI drug-discovery program focused on neglected diseases.[8]

In materials science, AI models are helping researchers search large pools of materials for potential products—such as batteries, semiconductors, and industrial catalysts—in hopes of finding compounds that are both stable and practical to manufacture.[9] Traditional materials discovery cycles can require years of iterative experimentation before safe and commercially viable compounds are identified, let alone developed into revenue-generating products. Physics-informed AI and deep learning surrogates (DLS) are also accelerating engineering design by approximating costly simulations, like fluid dynamics and structural performance, while still requiring validation through high-fidelity simulation and physical testing.[10]

Even with AI as an accelerant, the economics of scientific discovery—long a capital- and labor-intensive process—remain challenging. Many scientific fields continue to face declining R&D productivity. In biopharma, this phenomenon is referred to as “Eroom’s Law,” reflecting the tendency for drug discovery to become slower and more expensive over time despite advances in technology.[11] Other R&D sectors face a similar challenge. In response, some organizations are turning to AI to bend this curve. Berkeley Lab, for example, is using AI and automation to accelerate materials discovery in its A-Lab, where AI algorithms propose new compounds and robots prepare and test them—tightening the loop between hypothesis generation and experimental validation for applications like batteries and electronics.[12] Microsoft has said its Discovery platform helped develop the materials stack for the Majorana 2 quantum chip, using agentic AI to improve qubit reliability and support broader quantum hardware design work.[13] And at MIT, researchers trained a generative AI model on more than 23,000 materials synthesis recipes to help scientists navigate the hardest part of materials discovery: turning promising ideas into workable synthesis pathways.[14]

Investment in AI for scientific discovery and engineering is accelerating rapidly. AI-native drug discovery companies continued attracting major capital in 2026, highlighted by Isomorphic Labs’ $2.1 billion Series B financing, Earendil Labs’ $787 million raise for AI-driven biologics discovery, and Fathom Therapeutics’ $47 million oversubscribed Series A to advance its physics- and AI-enabled drug design platform.[15] The broader investment signal also includes PhysicsX’s reported $300 million financing at a $2.4 billion valuation this June.[16]

With so much to gain from accelerating the scientific process—drugs that could save millions of lives, chemicals that could spur entirely new industries, or physics discoveries that could unlock new energy sources— scientists are eager to harness advanced AI for the discovery process.

---

[8] “Claude Science, an AI workbench for scientists, is now available,” Anthropic, June 30, 2026.

[9] Mouyang Cheng et al., “Artificial intelligence-driven approaches for materials design and discovery,” *Nature Materials*, February 2026, Volume 25; Xuesong Wang et al., “AI-driven exploration and design of inorganic battery materials,” *Renewable and Sustainable Energy Reviews*, April 2026, Volume 229.

[10] Benedikt Alkin et al., “NeuralDEM for real time simulations of industrial particular flows,” *Communications Physics*, November 18, 2025, Volume 8, Number 440; Vignesh Babu Rao and Ashley D. Spear, “An interleaved physics-based deep-learning framework as a new cycle jumping approach for microstructurally small fatigue crack growth simulations,” *npj Computational Materials*, August 5, 2025, Volume 11, Number 255.

[11] “The next innovation revolution—powered by AI,” McKinsey, June 20, 2025.

[12] Laurel Kellner, “How AI and automation are speeding up science and discovery,” Berkeley Lab News Center, September 4, 2025.

[13] Catherine Bolgar, “Introducing Majorana 2: How Microsoft’s new quantum chip was made 1,000x more reliable with the help of Microsoft Discovery’s agentic AI,” Microsoft news release, June 2, 2026.

[14] Zach Winn, “How generative AI can help scientists synthesize complex materials,” MIT News, February 2, 2026.

[15] “Isomorphic Labs announces Series B investment round,” Isomorphic Labs announcement, May 12, 2026; “Earendil Labs announces $787 million in financing to scale AI-driven biologics discovery and development,” Earendil Labs press release, March 20, 2026; “Fathom Therapeutics, formerly Atommap, raises $47 million in oversubscribed Series A financing to translate physics and AI-enabled small molecule design into next-generation medicines,” Atommap press release, April 27, 2026.

[16] Kieran Smith, “UK AI start-up PhysicsX hits $2.4bn valuation following Temasek-led deal,” *Financial Times*, June 8, 2026.

AI for scientific discovery and engineering

Scoring the trend

News and searches are the strongest vectors, reflecting rising interest, while patents and research publications specific to the category remain limited. Investment is the clearest signal of momentum, rising from a low base to nearly $8 billion in 2025 and to $12.5 billion through the first half of 2026.

Equity investment, 2025
$7.9 billion

Job postings, 2024–25, % difference
+18%

Score, by vector (0 = lower; 1 = higher)

Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

Latest developments

Recent developments involving AI for scientific discovery and engineering include the following:

— *Protein-engineering and generative-chemistry foundation models are moving toward commercial deployment.* Organizations are increasingly using foundation models to engineer antibodies, generate novel molecules, predict toxicity, and optimize therapeutic candidates before physical testing. Recent advances in protein foundation models, including Boltz-2 and Evo 2, have demonstrated how large-scale AI models can support antibody engineering, protein optimization, and pharmaceutical research.[17] At the same time, antibody prediction remains difficult because therapeutic antibodies are often engineered to bind targets in highly specific ways that may not be well represented in the natural protein data sets used to train these models.[18] As a result, newer models such as Boltz-2 can help prioritize which antibody designs to test, but they still require experimental validation before researchers can rely on the result.[19]

---

[17] Fay Lin, “Boltz-2 released to democratize AI molecular modeling for drug discovery,” *Genetic Engineering & Biotechnology News*, June 6, 2025; “AI can now model and design the genetic code for all domains of life with Evo 2,” Arc Institute, February 19, 2025.
[18] Sarah Braner, “Academic AIs make inroads in protein binder design,” *Chemical & Engineering News*, November 5, 2025.
[19] Alexandra Fieux-Castagnet et al., “Benchmarking Boltz-2 for screening of therapeutic antibody-antigen interactions,” bioRxiv, May 14, 2026.

- *Materials science is becoming a commercially important frontier for AI-enabled discovery.* Transformer-based architectures are AI models that use an attention mechanism to understand relationships in data, making them highly effective for tasks like language processing, protein analysis, and drug discovery. They are increasingly being used to search massive design spaces across batteries, catalysts, semiconductors, alloys, carbon materials, and industrial manufacturing systems.[20] The goal is no longer simply identifying theoretically viable compounds, but determining which materials can realistically be synthesized, scaled, and manufactured economically. Researchers are now using transformer-based models to accelerate advanced materials discovery.[21] Investors see potential in AI for materials science, highlighted by Periodic Labs' $300 million seed financing in 2025, one of the largest early-stage investments in scientific AI.[22]
- *LLMs are being deployed as orchestration layers for scientific and engineering workflows.* R&D organizations are adopting LLMs not as replacements for deterministic simulation engines or scientific software, but as coordination layers that can manage workflows and interact with tools, evidenced in Anthropic's Claude Science announcement.[23] This hybrid approach is accelerating experimentation and engineering analysis across research-intensive industries. Momentum is particularly strong in environments where human experts, simulation engines, and AI systems must work together across highly technical workflows involving mathematics, physics, materials science, and industrial design. Companies such as PhysicsX, BeyondMath, and Luminary Cloud are developing AI-enabled engineering platforms that combine language models, simulation technologies, and domain-specific scientific workflows to accelerate engineering analysis and design.[24]

> **'Scientific AI is no longer just about computational prediction. It is about tightening the feedback loop between generative models and physical labs. Algorithms can help us explore vast chemical and biological spaces, but it is the seamless integration with rapid experimental validation that actually drives R&D productivity.'**
>
> — Alex Devereson, senior partner, London

---

[20] Mouyang Cheng et al., "Artificial intelligence-driven approaches for materials design and discovery," *Nature Materials*, February 2026, Volume 25.

[21] Mohammad Madani et al., "Accelerating materials property prediction via a hybrid transformer graph framework that leverages four body interactions," *npj Computational Materials*, January 18, 2025, Volume 11, Number 15.

[22] Julie Bort, "Former OpenAI and DeepMind researchers raise whopping $300M seed to automate science," TechCrunch, September 30, 2025.

[23] "Claude Science, an AI workbench for scientists, is now available," Anthropic, June 30, 2026.

[24] Stephen Rose and Nik Vadgama, "Engineering in the age of physics AI: The platform driving the shift," PhysicsX, April 15, 2026; David Cendon Garcia, "UK DeepTech startup BeyondMath raises €8.4 million to expand generative physics research," EU-Startups, February 25, 2026; "Luminary Cloud and nTop streamline AI physics and cut engineering design time from months to hours with NVIDIA technology," Luminary press release, March 18, 2025.

Research on memory-coordinated physics-aware simulation (MCP-SIM) illustrates how LLMs can translate incomplete natural-language prompts into validated simulations and outputs while coordinating with scientific simulation tools.[25] By combining language-based reasoning with domain-specific computational systems, MCP-SIM highlights the growing role of LLMs as intelligent orchestration layers that help automate and streamline complex technical and engineering workflows. To unlock value, organizations will need to codify proprietary data, experimental know-how, domain expertise, and validation routines into agentic workflows that reflect how their scientists and engineers work.

— *Scientific organizations are shifting from isolated AI models toward closed-loop discovery systems.* Rather than using AI only to make predictions or run simulations, organizations are increasingly combining AI with automated labs, robotics, and real-world experiments to continuously improve scientific discoveries.[26] While still at an early stage, this approach is gaining traction across areas such as drug discovery, battery chemistry, catalysts, semiconductors, and advanced materials.[27] In leading research environments, AI models generate hypotheses, while automated systems execute experiments, and orchestration software feeds data back into models to inform subsequent iterations.

Partnerships in closed-loop drug discovery illustrate this trend. For example, AI-driven drug discovery company Insilico Medicine has signed partnerships worth billions of dollars with pharmaceutical companies such as Eli Lilly, SK Biopharmaceuticals, and Takeda to jointly advance drug candidates across several therapeutic areas. The pharma companies will leverage Insilico's AI platform while continuing to focus on development and moving candidates through clinical trials.[28] Insilico has already validated AI's capacity to generate novel therapeutic assets, discovering targeted compounds such as Rentosertib for pulmonary fibrosis and Garutadustat for inflammatory bowel disease.[29] These kinds of partnerships show how AI-native discovery platforms are being connected to later-stage pharmaceutical development.

With AI taking the research reins, autonomous laboratories that operate as closed-loop scientific systems are also emerging. Highly automated laboratories exist in some markets such as China and automation is being used in other labs to connect fragmented infrastructure.[30] Commercial cloud lab platforms such as Emerald Cloud Lab allow scientists to "rent" lab functionality in the cloud, conducting experiments from computers in remote locations. These cloud labs combine robotics, remote instrumentation, orchestration software, and machine learning systems to support

---

[25] Donggeun Park, Hyeonbin Moon, and Seunghwa Ryu, "A self-correcting multi-agent LLM framework for language-based physics simulation and explanation," *npj Artificial Intelligence*, January 20, 2026, Volume 2.

[26] Xiang Zhuang et al., "Embodied science: Closing the discovery loop with agentic embodied AI," arXiv, March 20, 2026, arxiv.org/abs/2603.19782; Junwu Chen and Qiucheng Xu, "Artificial intelligence-driven autonomous laboratory for accelerating chemical discovery," *Chemical Synthesis*, September 2025, Volume 5.

[27] Woong Shin et al., "The (R)evolution of scientific workflows in the agentic AI era: Towards autonomous science," arXiv, September 12, 2025, arxiv.org/abs/2509.09915.

[28] Jason Chau, "Takeda, Insilico strike AI drug-discovery deal worth up to $600 million," *Wall Street Journal*, July 2, 2026; "Insilico Medicine and SK Biopharmaceuticals achieved AI-powered drug discovery collaboration worth up to $2.5 billion for neuroimmune disorders," Insilico Medicine press release, June 22, 2026; "Insilico Medicine announces global R&D collaboration with Lilly," Insilico Medicine press release, March 29, 2026.

[29] Kelly Bilodeau, "How well are AI-discovered drugs faring in the clinic?," BioPharma Dive, July 17, 2026.

[30] Jinpeng Li et al., "Autonomous laboratories in China: An embodied intelligence-driven platform to accelerate chemical discovery," *Digital Discovery*, June 13, 2025, Volume 4, Number 7.

increasingly autonomous experimentation and scientific discovery. The concept is already working in practice. Ginkgo Bioworks, for example, gave GPT-5 access to a cloud lab in Boston with a goal of improving the reaction conditions of cell-free protein expression. After six cycles of experimental work designed by GPT-5, Ginkgo saw a 40 percent reduction in costs for cell-free protein synthesis over scientific state-of-the-art practices.[31]

— *AI agents are starting to replace parts of custom scientific analytics pipelines.* Historically, highly specialized scientific workflows depended on fragmented bioinformatics tools and an army of technical experts manually coordinating analysis across systems. Increasingly, agentic systems are automating tasks within multimodal data analysis, scientific-literature synthesis, experiment design tracking, biomarker discovery, and enterprise knowledge retrieval, though humans are still important for approval and verification loops.[32]

— *Scientific data is a strategic asset in AI-enabled discovery.* Many organizations lack sufficient proprietary scientific data to reach informed discovery on their own, so consortium-based data pooling and public–private biobank initiatives are expanding rapidly.[33] Biobank programs across Iceland, the United Kingdom, and the United States continue to expand genomic, biomarker, and population-scale health data sets used to train and validate AI systems for drug discovery and precision medicine.[34]

At the same time, pharmaceutical companies are experimenting with federated learning approaches that enable collaborative model development without centralizing sensitive data. For example, Eli Lilly has expanded an initiative to improve toxicity prediction and molecular analysis while allowing external organizations to contribute data into shared learning systems.[35] Companies such as Recursion Pharmaceuticals are also investing in generating proprietary experimental data through automated laboratories and high-throughput testing infrastructure.[36]

> ‘Drug discovery is moving from isolated AI models toward orchestrated research systems. Models are being connected to workflows that design experiments, optimize molecules, assess toxicity, and evaluate candidates. That breadth improves speed of discovery and probability of scientific success.’
>
> — David Champagne, senior partner, London

---

[31] Alexus A. Smith et al., “Using a GPT-5-driven autonomous lab to optimize the cost and titer of cell-free protein synthesis,” bioRxiv, February 5, 2026.

[32] Binglan Li et al., “Agentic AI and the rise of in silico team science in biomedical research,” *Nature Biotechnology*, February 24, 2026, Volume 44.

[33] Karen O’Leary, “Pharma companies collaborate on secure data sharing for AI,” *Nature Medicine*, March 26, 2025; Marie Roehm, “AISB Network expands Federated OpenFold3 Initiative with three new pharma contributors,” Apheris press release, October 1, 2025.

[34] Keren Carss et al., “Whole-genome sequencing of 490,640 UK Biobank participants,” *Nature*, August 6, 2025, Volume 645.

[35] “Lilly launches TuneLab platform to give biotechnology companies access to AI-enabled drug discovery models built through over $1 billion in research investment,” Eli Lilly press release, September 9, 2025.

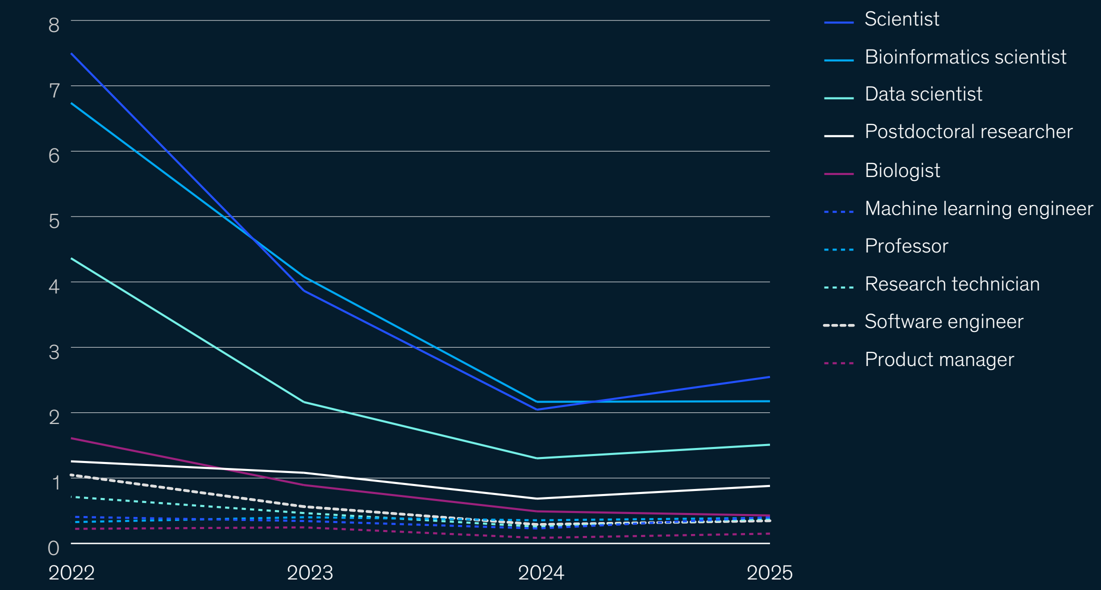
[36] “Companies making automated drug discovery a reality,” BiopharmaTrend, updated on November 3, 2025.

AI for scientific discovery and engineering

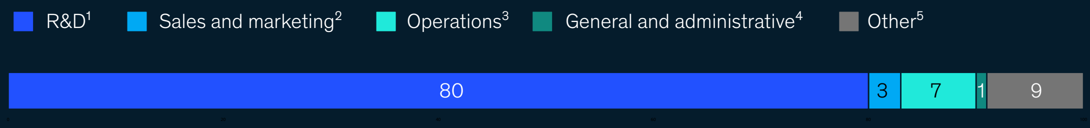
Demand

The talent market for AI in scientific innovation and engineering has contracted sharply from its 2022 peak, with postings falling by more than half. The decline has been broad, with scientist roles, the largest category, down by roughly two-thirds since 2022. The past year shows early signs of recovery, with machine learning engineers among the fastest-growing roles and scientists also bouncing back, suggesting the field is reorienting toward AI-enabled research. The functional mix is heavily weighted toward R&D, reflecting the academic and research intensity of scientific AI applications.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).

[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).

[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).

[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).

[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

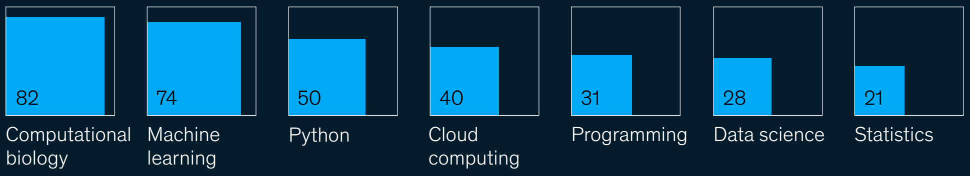
TALENT AND LABOR MARKETS

AI for scientific discovery and engineering

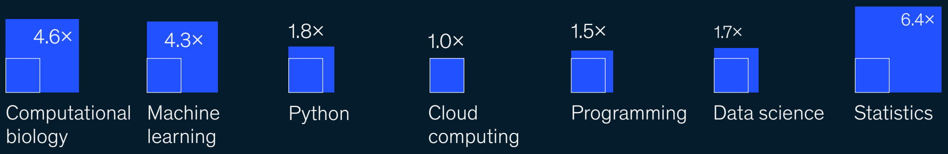
Skills availability

Computational biology and machine learning are the most widely required skills, both of which are well supplied. Core technical skills such as Python, cloud computing, programming, and data science show moderate demand and talent supply keeps pace. When it comes to this trend, domain-specific scientists and technology practitioners are available, but professionals who combine both are harder to find.

Talent required, % share of postings requiring skill ■



Talent availability, ratio of talent ■ to demand ☐



Adoption developments across the globe

Adoption score: 2—Experimentation. Most organizations remain in the initial stages of deploying AI-native scientific-discovery systems, though investment and experimentation accelerated sharply over the past year.

North America has a large share of AI-enabled drug discovery, computational biology, and foundation model investment, fueled by capital from frontier AI firms, pharmaceutical companies, hyperscalers, and research universities.

China is building autonomous laboratory infrastructure and industrial-scale scientific automation, particularly where robotics, AI, and physical experimentation platforms can be tightly integrated.

European organizations remain highly active in scientific AI research, particularly across pharmaceuticals, chemistry, and industrial materials, though adoption is often shaped by regulatory and governance considerations.

Across industries, pharmaceutical companies, biotechnology firms, battery developers, chemicals companies, and materials science organizations remain among the most active AI adopters.

Underlying technologies

The technologies that power the use of AI for scientific discovery and engineering include the following:

Agentic orchestration systems	These AI agents coordinate scientific workflows, tools, experiments, and knowledge sources to plan, execute, and optimize research activities.
Laboratory automation systems	These robotics and orchestration platforms automate experimental workflows and enable autonomous testing, data collection, and iterative scientific discovery.
Multimodal scientific data platforms	These integrated data environments combine biological, chemical, genomic, imaging, sensor, and experimental data sets to help support AI-driven analysis and discovery.
Protein-engineering models	These are AI systems optimized to analyze, predict, and design proteins, antibodies, enzymes, and other biological sequences with targeted functions and properties.
Scientific foundation models	These are large-scale AI models and deep learning surrogates trained on scientific data—including molecules, proteins, biological sequences, materials structures, and scientific literature—that can generate predictions, insights, and designs across scientific domains.

Key uncertainties

The major uncertainties affecting the development and adoption of AI for scientific discovery and engineering include the following:

— *Scientific reliability and reproducibility.* Questions remain regarding the consistency and validation of AI-generated scientific outputs.

— *Data availability and ownership.* Access to high-quality scientific data sets remains fragmented and uneven across industries.

— *Regulatory evolution.* Standards governing AI-generated scientific discoveries, drug development, and autonomous experimentation remain uncertain.

— *Laboratory integration complexity.* Integrating robotics, AI systems, simulation environments, and scientific infrastructure remains operationally challenging.

— *Economic scalability.* Many autonomous laboratory systems remain expensive to deploy and maintain at scale.

— *Downstream bottlenecks and translation constraints.* Even if AI-enabled R&D generates promising candidates more quickly, clinical trials, manufacturing scale-up, and other real-world constraints may limit how fast discoveries can move safely to market.

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with AI for scientific discovery and engineering:

- How autonomous can scientific discovery systems realistically become over the next decade?
- How should organizations balance open scientific collaboration with proprietary data and competitive advantage?
- How should regulators validate and govern AI-generated scientific outputs and experimentation?
- What new infrastructure, data-sharing models, and partnerships will be required to scale AI-enabled scientific discovery globally?

Read more about AI for scientific discovery and engineering

Knowledge center

Scientific AI

Life Sciences R&D

Related reading

The next innovation revolution—powered by AI

The future of delivering software as a medical device

AI in life sciences explained: The technology that could reinvent medicine

From linear gates to learning loops: Rewiring biopharma R&D with AI

04

AI infrastructure and model architectures

AI infrastructure and model architectures are the core technologies powering AI systems. They include foundation models trained on diverse data sets—which are vast networks of simulated "neurons" such as large language models (LLMs)—as well as multimodal, reasoning, and world models. The category also includes the supporting infrastructure of data centers, cloud platforms, training frameworks, and data pipelines needed to build and scale AI.

The trend—and why it matters

AI has entered a new phase defined less by experimentation and more by industrial-scale deployment. Building the AI infrastructure and model architectures to support these rollouts is expensive—very expensive. By 2030, capital expenditures to build data centers could reach nearly $7 trillion worldwide.[1]

The surging demand for compute power is also reshaping the economics of energy. AI is becoming one of the fastest-growing sources of energy demand, with US power demand associated with AI infrastructure projected to increase from roughly 30 gigawatts in 2025 to more than 90 gigawatts by 2030—equivalent to the current electricity demand of California.[2] Overall, data centers are expected to account for 14 percent of total US power demand by 2030, up from just 3 percent in 2022.[3] In some regions, AI-related data center demand has already contributed to notable increases in electricity pricing and grid stress as utilities struggle to keep pace with hyperscaler infrastructure expansion.[4] At the same time, shortages of processors, high-bandwidth memory, transformers, and skilled labor are emerging as major bottlenecks to scaling AI infrastructure.[5] Power availability and time to power are also becoming structural constraints on AI infrastructure growth. Data centers can often be built faster than the energy systems needed to power them become available, and this bottleneck is a growing factor holding back AI scaling.[6]

Amid this cost surge, the economics of AI deployment are shifting rapidly. Model developers, hyperscalers, and semiconductor firms are looking for ways to reduce costs for inference, the process of trained AI models generating outputs. Future advances in AI infrastructure and model architectures will include building integrated algorithm-hardware-energy systems that can operate sustainably at scale. To reduce costs, organizations are moving beyond model compression and inference optimization to redesign the stack around custom accelerators, higher-density racks, memory hierarchy, and power delivery architectures. The goal is to improve performance per watt and per dollar. AI infrastructure operators are increasingly deploying purpose-built AI chips to handle growing inference workloads (for more, see Application-specific

---

[1] "The cost of compute: A $7 trillion race to scale data centers," *McKinsey Quarterly*, April 28, 2025.
[2] "The next big shifts in AI workloads and hyperscaler strategies," McKinsey, December 17, 2025.
[3] "Global Energy Perspective 2025," McKinsey, October 13, 2025.
[4] *Energy and AI*, IEA, April 10, 2025.
[5] "The next big shifts in AI workloads and hyperscaler strategies," McKinsey, December 17, 2025.
[6] *Energy and AI*, IEA, April 10, 2025.

> **‘As AI capability advances, the scaling constraint moves into the physical stack, which is the physical capacity required to run AI at enterprise scale: chips, memory, power, cooling, networking, and people. For leaders, AI scale is now an infrastructure and resilience agenda.’**
>
> — Pankaj Sachdeva, senior partner, Philadelphia

semiconductors). By combining custom hardware with tightly integrated software, these chips can deliver faster and cheaper inference than general-purpose processors.[7]

On top of the infrastructure build-out, a parallel wave of advances in AI model architecture is reshaping what AI systems can do and how efficiently they can do it. Recent progress has come not only from scaling larger foundation models but also from improving the way models reason, process information, and interact with external tools and environments. New reasoning-oriented architectures can allocate additional compute to complex problems, enabling stronger performance on coding, scientific analysis, planning, and multistep decision-making. At the same time, multimodal models capable of processing and generating combinations of text, images, audio, video, and structured data are beginning to unify previously separate AI capabilities into integrated systems. World models are trained to build an internal representation of a physical environment—learning how that environment behaves and changes over time—so they can simulate outcomes before they happen in the real world. These models have the potential to vastly improve planning and decision-making for systems operating in the physical world.[8]

A new generation of efficiency-focused models could also help lower AI deployment costs. Small language models (SLMs), model compression techniques, and optimized inference frameworks, for example, can deliver strong performance while reducing compute requirements and deployment costs.[9] These advances are making it increasingly practical to deploy AI locally, at the edge, or within enterprise environments where latency, privacy, regulatory compliance, or cost constraints make cloud access to frontier models less attractive. For many organizations, the competitive advantage is shifting from simply accessing the largest models to orchestrating portfolios of models that balance performance, cost, speed, and governance requirements.

Beyond these advances, new classes and applications of AI systems are also emerging. Diffusion architectures—a class of gen AI models that produce images, video, audio, and other data by iteratively transforming random noise into structured outputs—are expanding into simulation, design, and scientific applications.[10] Autonomous agents are increasingly capable of executing complex workflows across software systems with limited human intervention, while vision-language-action (VLA) models are beginning to connect AI reasoning directly to physical actions in robotics

---

[7] *AWS Artificial Intelligence Blog*, “AWS Inferentia2 builds on AWS Inferentia1 by delivering 4x higher throughput and 10x lower latency,” blog entry by Samir Araujo, June 13, 2023; “Announcing Ironwood TPUs General Availability and new Axion VMs to power the age of inference,” blog entry by Amin Vahdat and Mark Lohmeyer, Google, November 6, 2025.

[8] Danijar Hafner et al., “Mastering diverse control tasks through world models,” *Nature*, April 2, 2025, Volume 640.

[9] *Artificial Intelligence Index Report 2025*, Stanford Institute for Human-Centered AI, Stanford University, April 2025.

[10] Claudio Zeni et al., “A generative model for inorganic materials design,” *Nature*, January 16, 2025, Volume 639.

and industrial environments. The applications of AI are expanding far beyond content generation toward autonomous systems that can perceive, reason, simulate, plan, and act—opening new opportunities in industrial automation, scientific discovery, cybersecurity, healthcare, and other complex domains.

With so much potential at stake, competition across the AI ecosystem is intense. Hyperscalers are investing heavily in models, chips, data centers, and developer platforms to become the default cloud providers for AI deployment, while enterprise software companies are integrating AI into core business applications.[11] Meanwhile, governments are increasingly treating AI as a strategic capability tied to economic competitiveness, industrial resilience, and national security, accelerating investment in semiconductor manufacturing and domestic AI data center capacity.[12] This dynamic is playing out in real time: China halted shipments of gallium, dysprosium, terbium, and yttrium to Japan for months in 2026 amid diplomatic tensions, disrupting supply chains for magnets and semiconductors critical to advanced manufacturing. Meanwhile, the United States has weighed targeted restrictions on Chinese open-weight AI models—in which trained model parameters (or "weights") are publicly released—over intellectual property theft and national-security concerns.[13] The United States is also pushing to bring advanced-chip manufacturing onshore: TSMC has escalated its planned US investment twice in just over a year—to $165 billion in March 2025, from $65 billion, then to $265 billion in July 2026—making it the largest foreign direct investment in US history and aiming to produce advanced two- and three-nanometer chips in Arizona.[14] While these specific measures may continue to evolve, these moves illustrate how critical-mineral supply chains and AI model access are both becoming direct levers of geopolitical leverage.

Despite widespread AI usage, most companies have not yet moved entirely from experimentation to enterprise transformation at scale. Our research shows that 89 percent of organizations regularly use AI in at least one business function, but only 37 percent attribute any level of positive EBIT impact to AI at the enterprise level.[15] Many companies continue to face challenges related to fragmented data environments, cybersecurity risks, workforce readiness, and complex integration. The central question for business leaders is no longer whether AI matters, but how their organizations can scale agentic systems responsibly and economically.

> **'Open-weight models, particularly from China, are emerging as a significant force in the global AI market and closing the gap with frontier models, expanding deployment options, including on-premises, increasing buyer choice, and accelerating competition on cost and capability.'**
>
> — Kevin Wei Wang, senior partner, Hong Kong

---

[11] Samantha Subin, "Tech megacaps plan to spend more than $300 billion in 2025 as AI race intensifies," CNBC, February 8, 2025.
[12] "World leaders put spotlight on sovereign AI at NVIDIA GTC," blog entry by Dan Rowinski, NVIDIA, March 4, 2025.
[13] Jackson Chen, "China's heavy rare earth exports to Japan remain frozen," Mining.com.au, July 22, 2026; Rebecca Bellan, "As US weighs response to Chinese AI, industry urges against broad open-weight restrictions," TechCrunch, July 24, 2026.
[14] Jenny Lee, "TSMC is accelerating Arizona factory build-out to capitalize on AI 'megatrend,' CFO says," CNBC, July 19, 2026.
[15] Dan Tinkoff, Lieven Van der Veken, Michael Chui, and Tara Balakrishnan, "The state of AI in 2026: On the road to ROI," McKinsey, August 25, 2026.

AI REVOLUTION

AI infrastructure and model architectures

Scoring the trend

AI infrastructure and model architectures is the most searched technology trend in 2025, reflecting broad enterprise urgency to understand and deploy foundational AI capabilities. Equity investment was $145 billion in 2025, the second highest of any trend, and has since surged to nearly $384 billion through mid-2026, the largest of any trend. If investment continues at the current pace, full-year 2026 investment would reach roughly $769 billion.

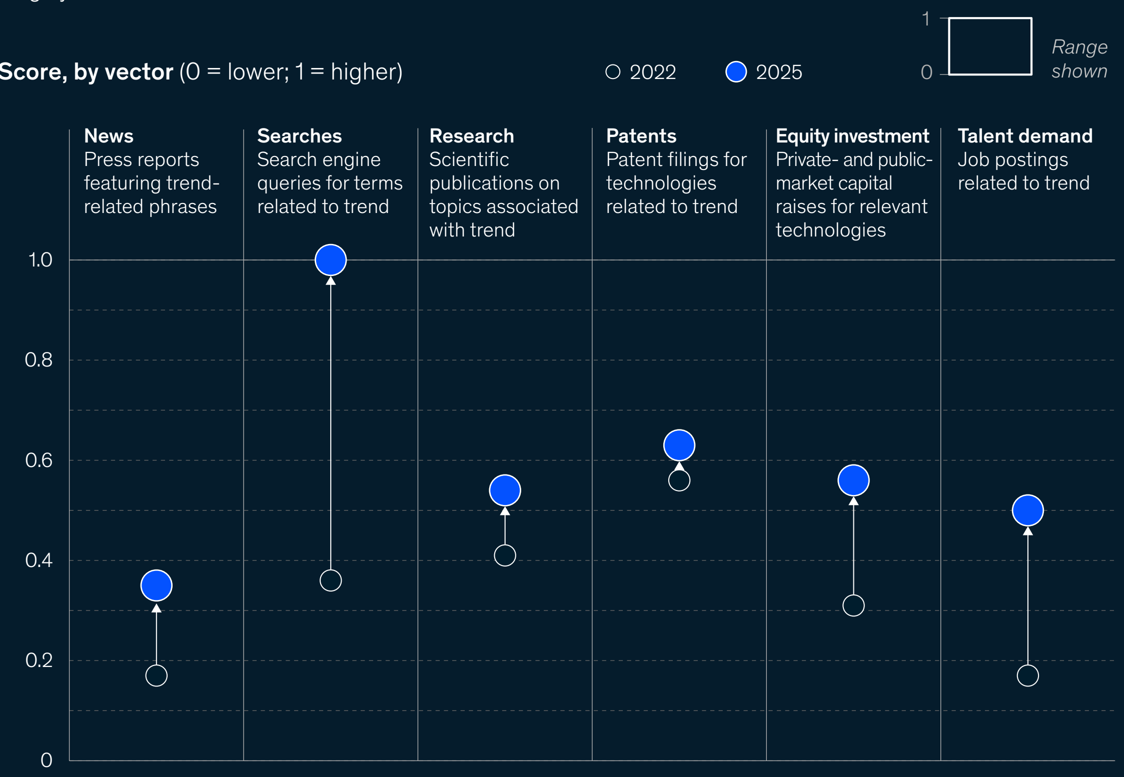
Equity investment, 2025

$145.0 billion

Job postings, 2024–25, % difference

+47%

Score, by vector (0 = lower; 1 = higher)



Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

Latest developments

Recent developments involving AI infrastructure and model architectures include the following:

— *Multimodal AI architectures are becoming part of enterprise systems.* Increasingly, AI models are being designed to process and coordinate information across text, images, video, speech, sensor data, code, and real-world environmental inputs within unified architectures. This evolution is enabling more context-aware systems that can reason across multiple forms of information simultaneously. As a result, enterprises are beginning to deploy multimodal AI in areas such as customer service, robotics, industrial monitoring, scientific research, and operational workflows, where integrating diverse data streams can enhance automation and decision-making.[16]

For example, ELLMER, an embodied LLM-enabled robotics framework, illustrates how LLMs can be combined with retrieval-

---

[16] "What is multimodal AI?," McKinsey, June 10, 2025.

augmented generation (RAG) and real-time sensor data to help robots to complete tasks in unpredictable settings.[17] In demonstrations involving coffee making and plate decoration, ELLMER retrieved relevant examples from a knowledge base and executed actions that incorporated both force and visual feedback.

— *World models are evolving from passive predictors into interactive simulators.* World models are becoming foundational for physical AI and robotics because they allow organizations to train and validate models in highly realistic digital environments before deployment (for more, see Future of mobility and Future of robotics). This reduces the cost, risk, and time associated with real-world testing. Advances in GPU-accelerated simulation are further accelerating this trend by dramatically compressing training timelines and enabling the creation of large-scale synthetic environments. Dreamer, a family of world model algorithms developed by Google and collaborating researchers, recently demonstrated how it uses reinforcement learning to imagine future scenarios and plan actions, mastering over 150 tasks.[18] Similarly, NVIDIA's Cosmos™ platform provides world foundation models and tools that generate photoreal, physics-based synthetic data to train and evaluate physical AI models.[19]

— *SLMs are becoming viable for targeted business applications.* Rather than competing directly with frontier-scale LLMs across every task, SLMs are showing strength in specialized use cases where accuracy, speed, and cost efficiency are more important than general-reasoning capabilities.[20] Early deployments are especially concentrated in areas such as customer service and workflow automation, where smaller models can deliver stronger domain-specific accuracy while operating at significantly lower cost. For example, Microsoft's Phi-4-mini, an SML with 3.8 billion parameters, is designed for targeted applications, including on-device assistants, and reasoning tasks such as math and coding.[21] (By contrast, an LLM is much larger; a version of GPT-3 that powered ChatGPT when it was introduced had 175 billion parameters.[22]) SLMs are increasingly being viewed as a promising path for cost optimization and scalable deployment across high-volume enterprise workflows.

— *Diffusion models are pivoting from visual novelty to computational efficiency.* Diffusion models have moved beyond experimentation, with companies now focusing on making these models faster, more cost-effective, and practical for real-world use. New techniques are reducing the amount of computation required to generate high-quality outputs, helping lower costs and improve response times for applications such as video generation and multimodal AI. NVIDIA's FastGen open-source library, for example, focuses on unifying diffusion-distillation techniques to reduce inference steps, latency, and computational cost.[23] At the same time, diffusion models are being applied beyond content generation to areas such as simulation and digital rendering.

---

[17] Ruaridh Mon-Williams et al., "Embodied large language models enable robots to complete complex tasks in unpredictable environments," *Nature Machine Intelligence*, March 19, 2025, Volume 7.

[18] Danijar Hafner et al., "Mastering diverse control tasks through world models," *Nature*, April 2, 2025, Volume 640.

[19] Niket Agarwal et al., "Cosmos World Foundation Model Platform for physical AI," arXiv, July 9, 2025, arxiv.org/abs/2501.03575.

[20] *Artificial Intelligence Index Report 2025*, Stanford Institute for Human-Centered AI, Stanford University, April 2025.

[21] "Empowering innovation: The next generation of the Phi family," blog entry by Weizhu Chen, Microsoft, February 26, 2025.

[22] "Generative AI in operations: Capturing the value," McKinsey, January 3, 2024.

[23] *NVIDIA Technical Blog*, "Accelerating diffusion models with an open, plug-and-play offering," blog entry by Weili Nie et al., January 27, 2026.

- *Neurosymbolic AI is regaining momentum as enterprises seek reliability and explainability.* Many organizations are finding that enterprise-scale AI requires stronger reasoning consistency and transparency than probabilistic architectures alone can provide. This is driving renewed interest in neurosymbolic approaches that combine neural networks with symbolic reasoning, rules-based systems, knowledge graphs, and logical constraints.[24] Companies are exploring these architectures to improve reliability in high-stakes environments such as healthcare, engineering, cybersecurity, and regulated decision-making, where explainability and deterministic reasoning are critical.

- *Inference and token costs are holding back AI scaling.* The cost of generating and serving AI output is emerging as a central operational challenge. Enterprises are increasingly deploying model architectures, hardware optimizations, inference orchestration, and compression techniques to reduce cost per token while maintaining performance levels.[25] Token-efficient AI ecosystems are emerging, including in China, where organizations are deploying with open-weight models and alternative training approaches. As enterprises scale agentic AI across coding, customer service, and operational workflows, long-term adoption may depend less on model capability and more on whether inference can be delivered at an economically sustainable scale.

- *Supply chain constraints are bottlenecks to AI growth.* The rapid expansion of AI data centers is placing pressure on supply chains. Access to construction materials, energy, water, electrical equipment, semiconductors, and memory is constrained, with hyperscalers having to wait months to get the materials they need to build new data centers. Compute demand continues to outpace infrastructure readiness in many regions, delaying AI scaling. Power provisioning, transformer availability, advanced semiconductor packaging, and high-bandwidth memory ecosystems are particularly constrained. Manufacturing leading-edge semiconductors requires many different types of advanced equipment. A critical bottleneck is lithography, where ASML, a Dutch company, is currently the only supplier in the world capable

> **‘Inference is becoming the operating cost of enterprise AI. As models move into high-volume workflows, model innovations not only expand what a system can do but also address cost, speed, and reliability. In addition, model orchestration is quickly becoming an economic discipline.’**
>
> — Roger Roberts, partner, Bay Area

---

[24] Uzma Nawaz et al., “A review of neuro-symbolic AI integrating reasoning and learning for advanced cognitive systems,” *Intelligent Systems with Applications*, June 2025, Volume 26.

[25] Ben Cottier et al., “LLM inference prices have fallen rapidly but unequally across tasks,” Epoch AI, March 12, 2025.

of producing the extreme ultraviolet (EUV) lithography machines required to manufacture chips at three-nanometer scale and below—concentrating a critical stage of the supply chain in a single firm's equipment.[26] This constraint, however, is not static: China has begun domestic production of immersion deep-ultraviolet (DUV) lithography tools and has reportedly started building an early-stage EUV prototype, signaling a state-backed push to reduce reliance on ASML.[27] As AI expands into robotics and industrial systems, computing is also moving into factories, vehicles, and edge devices, creating demand for hardware that can run AI locally and thus compounding supply chain backlogs.[28]

— *Cyber-resilient AI architectures could help expand autonomous systems.* As AI systems gain greater autonomy and become more deeply integrated into enterprise operations, companies want to ensure those systems don't introduce cybersecurity risks. When AI agents execute autonomous workflows, that opens up more potential attack surfaces. And AI is making it easier for hackers to spot vulnerabilities and quickly exploit them.[29] This situation is driving demand for AI-native cybersecurity controls designed specifically for autonomous environments, including guardrails against prompt injection, tool use authorization for autonomous agents, and verification of multistep reasoning before execution (for more, see Cybersecurity and trustworthy systems).

— *Countries are building domestic AI infrastructure to reduce supply chain dependence.* In China, for example, AI model makers are increasingly relying on domestic AI hardware and optimizing their models to reduce inference costs.[30] Europe is also emphasizing sovereign infrastructure development, while the Middle East is expanding investments in AI infrastructure and energy-backed compute ecosystems through national industrial strategies and sovereign wealth funding. These national-scale AI initiatives reflect a growing focus on regional independence in the AI race.

— *Capital expenditures for AI are straining balance sheets.* The scale of infrastructure investment is now large enough to move the financial profile of the companies funding it. Alphabet's capital expenditures reached $44.9 billion in the second quarter of 2026—nearly double the prior year. To fund the increased spending, the company raised roughly $50 billion in equity and $20 billion in debt during the quarter, signaling that even hyperscalers with the deepest balance sheets are partly financing the AI build-out externally rather than purely from operating cash flow.

---

[26] Milana Vinn and Max A. Cherney, "ASML becomes Mistral AI's top shareholder after leading latest funding round, sources say," Reuters, September 7, 2025.

[27] Fanny Potkin, "China starts production of home-grown immersion DUV chipmaking tools, source says," Reuters, July 28, 2026.

[28] "The next big shifts in AI workloads and hyperscaler strategies," McKinsey, December 17, 2025.

[29] *New at McKinsey Blog*, "AI is the greatest threat—and defense—in cybersecurity today. Here's why," blog entry by Charlie Lewis, Ida Kristensen, Jeffrey Caso, and Julian Fuchs, May 15, 2025.

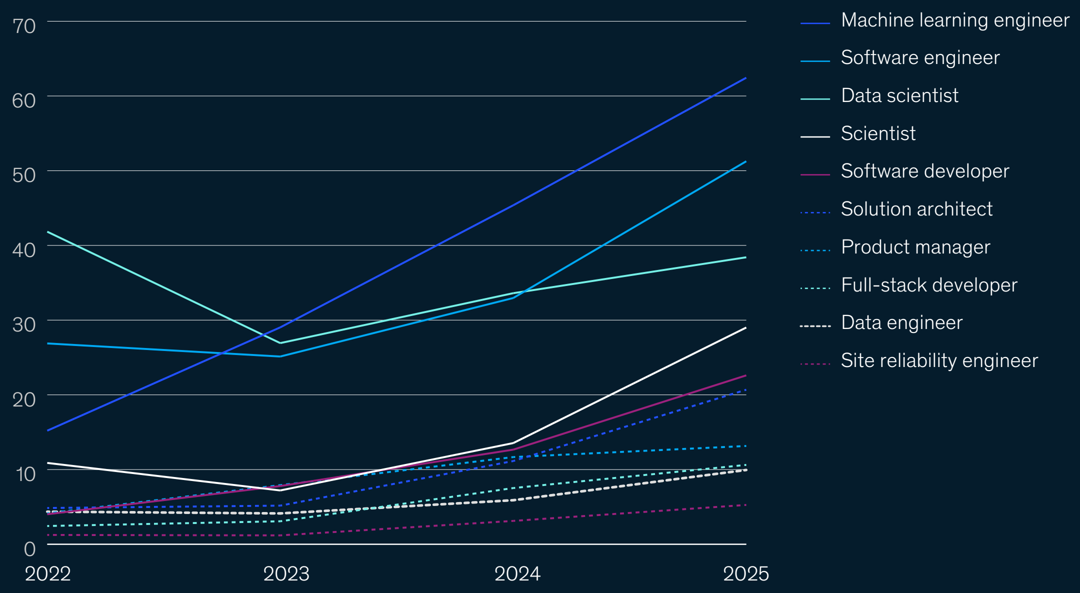
[30] Eduardo Baptista et al., "DeepSeek unveils new AI model tailored for Huawei chips as China pushes for tech autonomy," Reuters, April 23, 2026.

AI infrastructure and model architectures

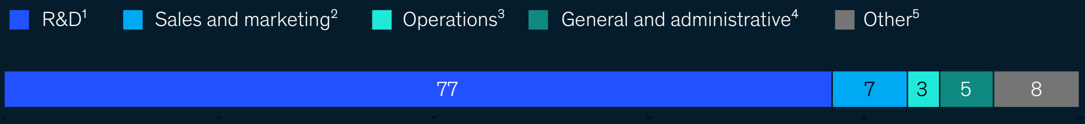
Demand

AI Infrastructure and model architectures is one of the largest talent markets among the technology trends, with postings nearly tripling since 2022. Job postings for machine learning, software engineering, and scientist roles saw the fastest growth, reflecting deepening investment in foundational research. Growth is broad based, with every top ten role expanding, signaling sustained buildup across both research and deployment. The functional mix is weighted heavily toward R&D, underscoring a field where the basis of competition is innovation.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).

[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).

[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).

[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).

[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

TALENT AND LABOR MARKETS

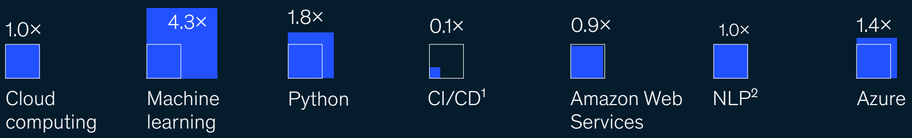
AI infrastructure and model architectures

Skills availability

Cloud computing and machine learning are the most widely required skills, and talent supply meets or exceeds demand. The shortage is sharpest in continuous integration and continuous delivery (CI/CD), pointing to a gap in the infrastructure and deployment expertise needed to move models into production.

Talent required, % share of postings requiring skill ■

Talent availability, ratio of talent ■ to demand ☐



[1] Continuous integration and continuous delivery.
[2] Natural language processing.

Adoption developments across the globe

Adoption score: 4—Scaling in progress. While companies are scaling AI across the enterprise with varying degrees of maturity by geography and industry, regional ecosystems are increasingly diverging in infrastructure priorities, deployment models, and commercialization strategies. North America continues to lead in frontier model development, hyperscale infrastructure investment, and venture capital funding.[31] Though the United States has strong start-up activity in the AI space, large technology companies remain central to AI infrastructure expansion, particularly in cloud services, semiconductors, and enterprise platforms.

China continues to accelerate AI deployment across manufacturing, mobility, consumer platforms, and industrial automation.[32] The country remains a major force in model development, semiconductor investment, and AI-enabled hardware ecosystems. Chinese firms are also increasingly designing and manufacturing their own AI chips to reduce dependence on foreign semiconductor companies. Beyond reducing hardware dependence, China is accelerating the deployment of low-cost, open-weight models, allowing organizations to download, customize, and run models locally.[33] Chinese organizations are also investing in AI systems optimized for industrial-scale deployment, including autonomous agents, enterprise AI platforms, and compute-as-a-service providers. Demand is particularly strong among enterprises and public sector organizations seeking AI that can be deployed in domestic cloud environments, integrated into existing operational workflows, and adapted to industry-specific requirements—especially in sectors where data governance, cybersecurity, regulatory oversight, and infrastructure independence are strategic priorities.

Europe is advancing AI by focusing on sovereignty and trust and investing in building its own AI infrastructure. While European organizations lag

---

[31] *Artificial Intelligence Index Report 2025*, Stanford Institute for Human-Centered AI, Stanford University, April 2025.
[32] Eduardo Baptista et al., “DeepSeek unveils new AI model tailored for Huawei chips as China pushes for tech autonomy,” Reuters, April 23, 2026.
[33] Robert McMillan et al., “China has matched Anthropic in cybersecurity, resetting AI race,” *Wall Street Journal*, June 27, 2026.

behind the United States in AI adoption, the region has many top research institutions and is taking a lead in applying AI in industrial settings and setting strong regulatory governance.[34] European governments and enterprises are increasing investment in sovereign AI infrastructure to support AI adoption in regulated sectors such as banking and healthcare.

In many emerging economies, AI infrastructure and model architectures remain underdeveloped due to infrastructure limitations, compute constraints, workforce readiness gaps, and energy availability challenges.[35] However, lower-cost open models and cloud-based AI services are expanding access in regions that have previously faced barriers to adoption.

Across sectors, adoption is progressing fastest where organizations possess strong digital infrastructure, high-quality data environments, and clear economic incentives for automation or optimization. For example, industries such as advanced manufacturing, media and telecom, insurance, and technology are actively scaling AI agents across multiple workflows.[36]

Underlying technologies

The technologies that power AI infrastructure and model architectures include the following:

AI accelerators and specialized semiconductors	These technologies include hardware optimized for large-scale AI training and inference workloads.
Diffusion models	Diffusion models are generative systems that iteratively refine outputs from noise to produce images, video, audio, simulation environments, and synthetic data.
Edge AI	These AI models are deployed locally on devices and industrial systems to reduce latency and improve privacy and responsiveness.
Foundation models	Foundation models are large-scale AI systems trained on vast data sets that support reasoning, generation, retrieval, and enterprise applications.
Multimodal models	Multimodal models are capable of processing and generating combinations of text, images, video, audio, code, and sensor data.
Vision-language-action (VLA) models	VLA models are AI systems that combine perception, language understanding, and physical action to control robots and embodied systems.
Vector databases and retrieval systems	These technologies enable AI systems to retrieve and reason across enterprise knowledge bases.
World models	World models are architectures designed to simulate physical environments, predict outcomes, and support planning for robotics, autonomous systems, and embodied AI.

---

[34] "Time to place our bets: Europe's AI opportunity," McKinsey Global Institute, October 1, 2024.

[35] *Technology and innovation report 2025: Inclusive artificial intelligence for development—overview*, UN Conference on Trade and Development, 2025.

[36] "The state of AI in 2026: On the road to ROI," McKinsey, August 25, 2026..

Key uncertainties

The major uncertainties affecting AI infrastructure and model architectures include the following:

- *AI infrastructure scalability.* Rising demand for compute, power, cooling, networking, and semiconductors could constrain deployment and increase costs.
- *Regulatory fragmentation.* Diverging regulatory frameworks across regions may complicate global AI deployment, governance, and interoperability.
- *Workforce disruption and adaptation.* The long-term effects of AI on labor markets, organizational structures, and workforce productivity remain uncertain.
- *Cybersecurity risks.* AI-enabled attacks and vulnerabilities associated with model misuse, data leakage, and autonomous systems are increasing.
- *Geopolitical competition.* AI is increasingly becoming a strategic capability tied to national security, industrial policy, and global influence.
- *Energy availability and sustainability.* The environmental footprint of large-scale AI systems, including electricity and water consumption, may become a growing constraint.
- *Economic concentration.* The increasing concentration of AI infrastructure, talent, and capital among a relatively small number of organizations could affect competition and innovation.

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with AI infrastructure and model architectures:

- What balance should organizations strike between proprietary frontier models, open-weight systems, and smaller specialized models?
- What infrastructure investments will be required to support the next generation of AI-enabled applications and industrial systems?
- How will AI reshape labor markets, knowledge work, and the structure of organizations over the next decade?
- Which countries and regions will emerge as long-term leaders in AI infrastructure, talent, and deployment, and what are the potential implications of banning open-weight and other models for security concerns?
- How will AI's integration into physical systems, scientific discovery, and critical infrastructure reshape global economics, and what international governance frameworks will be needed to ensure the reliability of AI systems?

Read more about AI infrastructure and model architectures

Knowledge center

Insights on Artificial Intelligence

Related reading

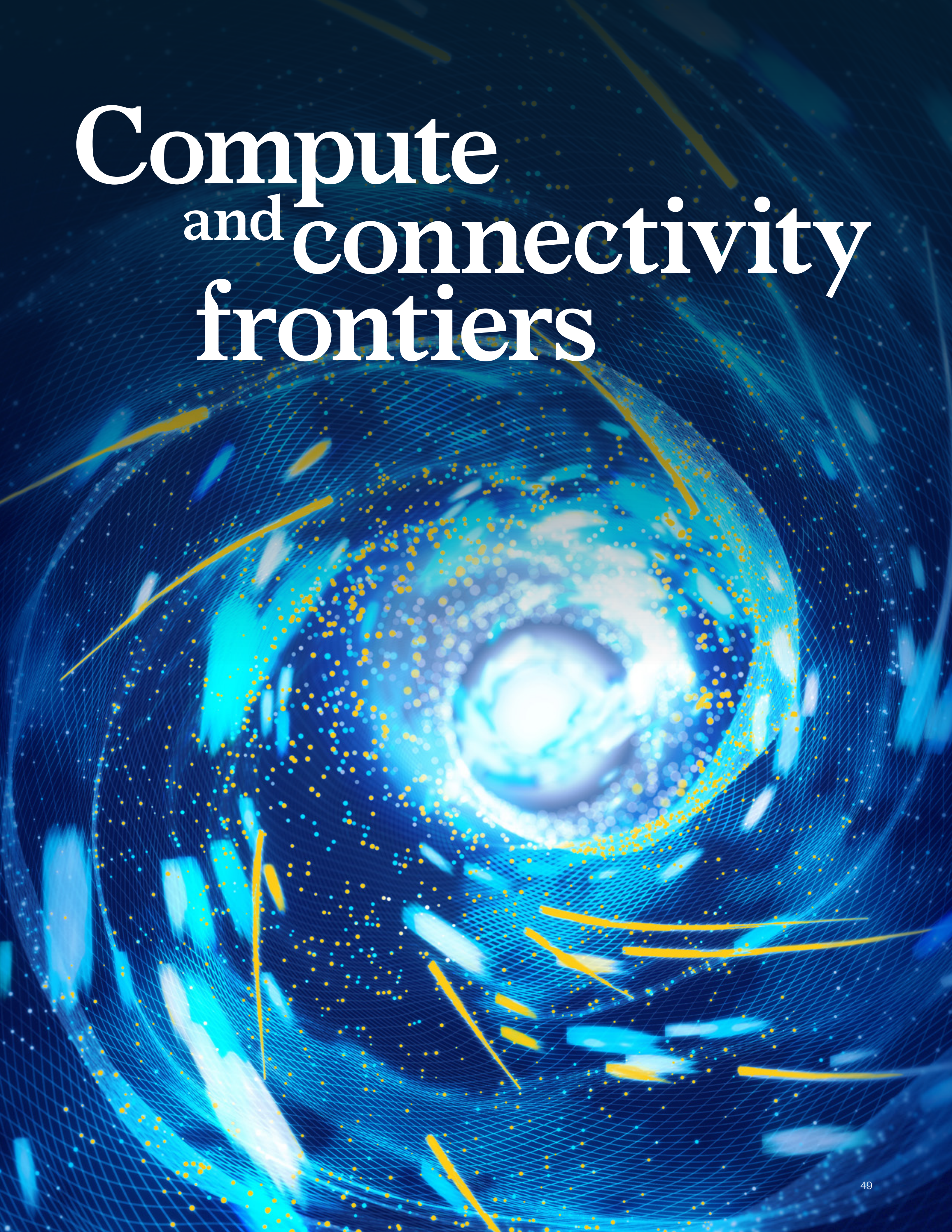
Frontiers of compute: The technologies to reduce AI inference costs

Colocation data centers: The infrastructure race behind AI

The $7 trillion data center build-out: How industrials can capture their share

The cost of compute: A $7 trillion race to scale data centers

Compute and connectivity frontiers



05

Application-specific semiconductors

Application-specific semiconductors are purpose-built chips optimized for specialized computing workloads. For artificial intelligence, such chips are increasingly being designed in close coordination with software systems, memory architectures, networking infrastructure, and data center operating models to meet the requirements of large-scale AI deployment.

The trend—and why it matters

Application-specific semiconductors have long been used in industries such as automotive, industrials, and consumer electronics. Now, they are being developed to meet the rapid growth in AI workloads. AI has fundamentally reshaped the semiconductor landscape. What began as a race by chipmakers to build larger semiconductor clusters to train AI models has evolved into a broader transformation of compute architecture. Training AI models requires massive computing power, while inference—the process of generating outputs from a trained model—demands greater efficiency and lower latency. Amid cost and competition pressures to maximize returns on AI investments, hyperscalers such as Amazon, Google, Meta, and Microsoft are increasingly partnering with semiconductor firms to codesign custom silicon tailored to their AI models. Meanwhile, semiconductor firms are redesigning their processors, memory systems, and networking architectures to handle the growing inference demands of agentic AI.

Earlier AI infrastructure investment focused heavily on training frontier models using highly centralized clusters of graphics processing units (GPUs). Now, AI inference is driving a rethink of semiconductor architectures. As enterprises adopt AI across core workflows, the diversity of inference workloads—each with distinct latency, throughput, and efficiency requirements—is creating demand for application-specific semiconductors optimized for different AI workloads. Rather than relying on a single type of processor for all workloads, data centers are increasingly deploying heterogeneous systems in which custom application-specific integrated circuits (ASICs) and specialized accelerators, alongside general-purpose processors, are optimized for distinct roles.

The industry is also moving toward more modular and workload-specific compute architectures. As AI workloads grow larger, chipmakers are using enabling technologies such as chiplets—pieces of processors that can be combined into larger systems—along with silicon photonics, which uses light to move data faster between chips, to ease bottlenecks in compute pipelines. They are also adopting advanced packaging techniques that place processors and memory closer together to reduce power consumption and improve heat efficiency.

At the same time, geopolitical competition is reshaping semiconductor supply chains. Governments increasingly see semiconductors as strategic infrastructure tied to economic resilience and national security. Efforts to diversify manufacturing capacity beyond Asia accelerated over the past year, though Taiwan remains central to advanced chip manufacturing.

This wave of innovation is not limited to semiconductor incumbents. Semiconductor start-ups focused on inference acceleration, memory optimization, wafer-scale compute, and silicon photonics are attracting substantial investment from hyperscalers, venture capital firms, and established semiconductor companies seeking innovation and alternatives to their existing supply chains.

These shifts highlight the changing boundaries among semiconductor companies, hyperscalers, and AI model developers (for more, see AI infrastructure and model architectures). Model makers and cloud providers are building deeper silicon expertise, while semiconductor companies are developing a more sophisticated understanding of AI models, workloads, and software stacks.

Application-specific semiconductors

Scoring the trend

Although interest in application-specific semiconductors remains modest in absolute terms, it is accelerating, with news mentions, searches, and equity investment all rising meaningfully from 2022 to 2025. Patent filings were the strongest innovation signal, with activity outpacing broader market trends. Investment surged to nearly $42 billion in 2025, compared with $7.5 billion in 2024, as demand for specialized chips grew alongside AI workloads. While the pace of research publications has not accelerated, much of the innovation is occurring in private companies, whose incentives to publish can be limited.

Equity investment, 2025	**Job postings,** 2024–25, % difference
$41.6 billion	+22%

Score, by vector (0 = lower; 1 = higher)

Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

Latest developments

Recent developments involving application-specific semiconductors include the following:

— *Application-specific semiconductors for inference are gaining momentum.* As enterprises embed AI into core workflows and hyperscalers commercialize AI services, inference is becoming a major source of compute demand. This increases the importance of power efficiency and workload-specific optimization, driving investment in application-specific silicon designed to support inference workloads across a widening range of deployment contexts, from hyperscaler data centers to enterprise edge environments. McKinsey estimates that by 2030, inference will account for 30 to 40 percent of all data center demand and the majority of AI compute.[1] Emerging companies such as Cerebras, Groq, and SambaNova have developed semiconductor designs optimized for inference. Cerebras announced its initial public offering in May of this year.[2]

---

[1] "The next big shifts in AI workloads and hyperscaler strategies," McKinsey, December 17, 2025.

[2] Sneha Khanvilkar, "Cerebras Systems announces launch of initial public offering," Cerebras press release, May 4, 2026.

> ‘The competitive frontier in AI infrastructure is shifting from chips to full system performance. For hyperscalers, custom accelerators are no longer merely a source of cost and performance advantage; they are becoming the foundation of differentiated cloud platforms, with value accruing to those that can codesign silicon, systems, software, and services as an integrated stack.’
>
> — Yvonne Ferrier, associate partner, Bay Area

- *Custom silicon is evolving from an internal capability into a customer-facing competitive strategy.* Hyperscalers are increasingly developing their own application-specific semiconductors to optimize their own infrastructures and differentiate their offerings from other cloud platforms. Anthropic’s expanded use of Google Cloud tensor processing units (TPUs) shows how internally developed AI accelerators are being used by model developers. Google originally designed TPUs to accelerate its own machine learning workloads, but Anthropic now plans to use up to one million Google Cloud TPUs to support Claude’s growing compute needs, alongside other chip platforms.[3] Amazon Web Services (AWS) continues to deploy its Trainium and Inferentia chips (for training and inference, respectively) as differentiated cloud offerings.[4] Notably, AWS Inferentia (Inf1) chips power Amazon EC2 Inf1 instances to deliver up to 2.3 times higher throughput and up to 70 percent lower cost per inference than comparable EC2 instances.[5]

- *Hardware–software codesign is the next generation of semiconductor development.* Competitive advantage is increasingly determined by semiconductor codesign that integrates processors, memory, networking, and software into coordinated processing architectures, rather than designing stand-alone chips. Organizations are increasingly designing semiconductors optimized for specific AI models, alongside orchestration software that optimizes end-to-end AI workloads. For example, semiconductor company Advanced Micro Devices (AMD) emphasizes the growing importance of CPUs for AI orchestration, concurrency, and data movement within the data center.[6]

- *Advanced packaging and interconnect technologies are becoming strategic differentiators.* As AI clusters continue to scale, system performance increasingly depends not only on compute capacity but also on the ability to move and package data efficiently. High-bandwidth memory (HBM), advanced packaging, high-speed interconnects, and emerging optical-networking technologies are becoming critical determinants in this regard. These technologies increase bandwidth, reduce latency, improve power efficiency, and shorten deployment timelines for specialized semiconductors in data centers. SK hynix continues to expand next-generation HBM production, including with the announcement of its latest 12-layer next-gen HBM4E dynamic random-access memory (DRAM) chip, while Taiwan Semiconductor Manufacturing Company’s chip-on-wafer-on-substrate (CoWoS) packaging has become an enabler of modern AI accelerators.[7] At the same time, companies such as Astera Labs, Lightmatter, and Lumentum are advancing high-bandwidth electrical and optical interconnect technologies designed to improve communication across increasingly modular AI systems.[8]

---

[3] “Expanding our use of Google Cloud TPUs and services,” Anthropic press release, October 23, 2025.

[4] “Trainium3 UltraServers now available: Enabling customers to train and deploy AI models faster at lower cost,” Amazon News, December 2, 2025.

[5] “AWS Inferentia: Why Inferentia?,” Amazon Web Services, accessed July 2026.

[6] “Agentic AI brings new attention to CPUs in the AI data center,” blog entry, AMD, March 13, 2026.

[7] “SK hynix ships samples of 12-layer next-gen ‘HBM4E,’” SK hynix press release, June 18, 2026; Jack Whitney, Matthew Schleich, and William Alan Reinsch, *The double-edged sword of semiconductor export controls*, Center for Strategic and International Studies, October 2024.

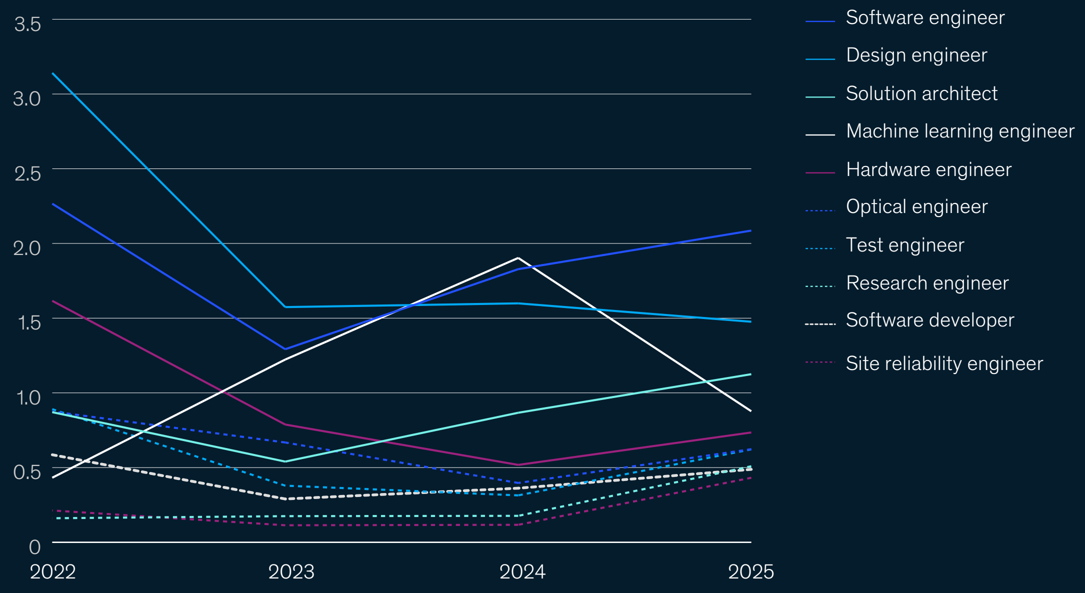
[8] “Astera Labs extends leadership in open, AI scale-up networking with new 320 Lane Scorpio X-Series Smart Fabric Switch,” Astera Labs press release, May 5, 2026; Charlotte Truman, “Lightmatter unveils Guide DR laser NIC to support CPO scale-up,” Data Centre Dynamics, May 27, 2026; “Lumentum showcases breakthrough optical scale-up demonstration at OFC 2026 using VCSEL technology,” Lumentum press release, March 17, 2026.

Application-specific semiconductors

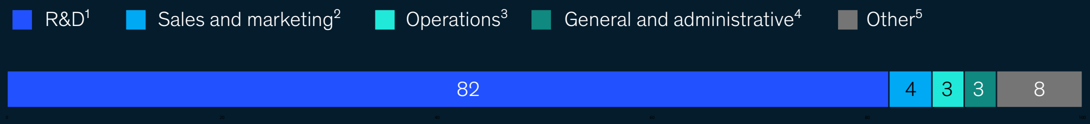
Demand

The talent market for application-specific semiconductors is recovering from its post-2022 contraction. Hiring has shifted away from traditional hardware roles, with job postings for design, hardware, and optical engineering positions all down since 2022. Meanwhile, demand for roles such as research engineers and solution architects has grown. The functional mix is heavily weighted toward R&D, reflecting the deep technical focus required for developing custom silicon.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).
[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).
[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).
[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).
[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

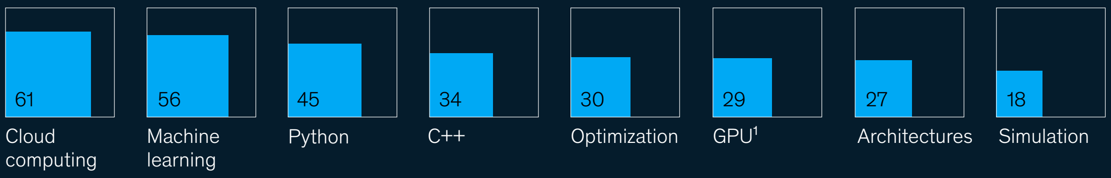
TALENT AND LABOR MARKETS

Application-specific semiconductors

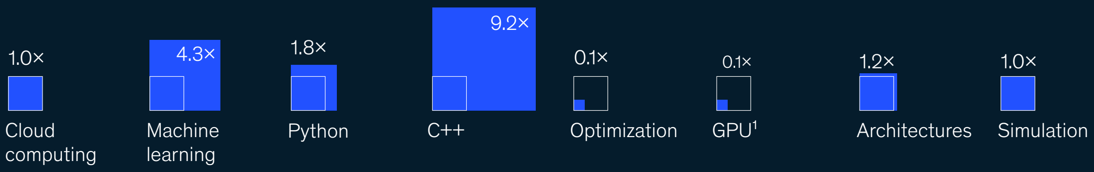
Skills availability

Cloud computing and machine learning are widely required skills, and talent is well supplied. Programming and infrastructure skills such as C++ and Python remain readily available. The sharpest shortages sit in GPU and optimization expertise, skills at the core of custom-chip design.

Talent required, % share of postings requiring skill 🟦



Talent availability, ratio of talent 🟦 to demand ☐



[1] Graphics processing unit.

Adoption developments across the globe

Adoption score: 4—Scaling in progress. Application-specific semiconductors are being scaled across data centers, cloud environments, and enterprise workloads.

North America continues to anchor the development of application-specific semiconductors, particularly among hyperscalers, chip designers, and AI infrastructure start-ups in the United States. The region benefits from the concentration of major data center customers that shape demand, codevelop custom silicon, and provide early routes to scale.

Taiwan continues to play a leading role in manufacturing advanced-node semiconductors and advanced packaging, though governments and enterprises are making efforts to diversify supply chains into other geographies, including India.

China is investing in domestic semiconductor ecosystems, including packaging, memory, and manufacturing capabilities. Reports suggest China has begun to make deep ultraviolet (DUV) lithography machines, the global market for which—along with the more advanced extreme ultraviolet (EUV) lithography machines—has long been dominated by the Dutch manufacturer ASML.[9]

Europe is seeking to reduce its dependence on global chip supply chains while building on its existing strengths in semiconductor tools, photonics research, advanced manufacturing, and new approaches to packaging AI processors.

Globally, hyperscalers are the largest adopters of application-specific semiconductors in their data centers because of escalating demand for AI inference capacity. Enterprise adoption is expanding gradually as inference costs decline and specialized AI infrastructure becomes more commercially accessible.

---

[9] Arjun Kharpal, “China’s reported chip breakthrough comes with some big caveats,” CNBC, July 28, 2026.

Underlying technologies

The technologies that power application-specific semiconductors include the following:

AI accelerators	These specialized processors optimize AI training and inference workloads.
ASICs	Application-specific integrated circuits, or ASICs, are designed for highly specialized AI computing tasks.
Chiplets	Modular semiconductor architectures combine multiple specialized dies into unified systems.
High-bandwidth memory	Advanced memory systems are optimized for the throughput requirements of AI infrastructure.
Silicon photonics	These optical interconnect technologies improve bandwidth and reduce energy consumption.

Key uncertainties

The major uncertainties affecting application-specific semiconductors include the following:

- *AI workload stabilization.* Inference and enterprise AI workloads may not scale predictably, and the pace at which new enterprise use cases emerge will determine how many distinct chip architectures the market can sustain.
- *Infrastructure economics.* Semiconductor companies continue to face difficult trade-offs between massive fabrication investments and increasingly rapid technology obsolescence.
- *Supply chain resilience.* Geopolitical tensions continue to create uncertainty about manufacturing concentration, export controls, and access to advanced-node production.
- *Memory and packaging bottlenecks.* Constraints related to HBM supply, advanced packaging capacity, and networking infrastructure may limit AI infrastructure expansion.
- *Energy and thermal limits.* AI infrastructure growth continues to increase pressure on power delivery, cooling systems, and energy efficiency, which could stymie AI scaling.

> **‘The lines between chipmakers, cloud providers, and model developers are blurring. Model builders need silicon expertise, and semiconductor companies need a deeper understanding of AI workloads and software stacks. That changes the operating model and the lead time for product development, because the best architecture decisions now sit across companies, not always inside one design team.’**

— Mark Patel, senior partner, Bay Area

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with application-specific semiconductors:

- How will the shift from training to inference reshape semiconductor architecture and infrastructure investment priorities?
- What balance will emerge between general-purpose GPUs and highly specialized AI accelerators?
- How will hyperscalers influence the future competitive landscape as they increasingly design and commercialize custom silicon?
- Which emerging architectures, including chiplets, silicon photonics, and disaggregated inference systems, will become foundational to next-generation AI infrastructure?
- How will geopolitical tensions continue to reshape semiconductor manufacturing, packaging, and supply chain strategies?

Read more about application-specific semiconductors

Knowledge center

Insights on Semiconductors

Related reading

Semiconductors: Etching the new map of strategic supply

Frontiers of compute: The technologies to reduce AI inference costs

The next era of semiconductor value creation

Hiding in plain sight: The underestimated size of the semiconductor industry

06

Cybersecurity and trustworthy systems

Cybersecurity and trustworthy systems help secure computer networks, make digital interactions more verifiable across interconnected environments, and enhance trust in digital systems. Increasingly enabled by AI, cybersecurity systems include zero-trust architectures and identity access. Other technologies that enhance trust include blockchain-based ledger systems, tokenization platforms, and quantum-safe cryptography.

The trend—and why it matters

When it comes to cybersecurity, AI is both a blessing and a curse. In the hands of hackers, AI is a new tool that makes it far faster and easier to find and exploit vulnerabilities. But it's also an emerging tool for companies to strengthen and automate their own cyber defenses. That said, cybersecurity and trustworthy systems are no longer only about preventing bad actors from breaking into computer systems. That's because AI is altering the underlying dynamics of cyber defense, while quantum computing is expected to break commonly used types of cryptography (for more, see Quantum technologies). As a result, cybersecurity is now also about building the trust layer required for AI agents, distributed cloud environments, and machine-to-machine interactions to operate safely and verifiably.[1]

Over the past year, the most critical shift has been the collapse of the vulnerability window, or the period during which a system is exposed to a security risk before it's detected. Frontier model capabilities are improving rapidly, with recent research showing that advanced AI agents can now autonomously discover and exploit some vulnerabilities, compressing the time and expertise required for offensive activity.[2] More than three-quarters of all cybersecurity vulnerabilities are currently classified as "zero day," meaning that by the time they are publicly disclosed, an exploit has already been developed.[3]

There's a bigger change happening underneath these numbers. AI agents are increasingly able to autonomously chain multiple exploitation steps into a single coherent attack path, a pattern that individual vulnerability scans are not designed to catch. This is no longer theoretical: In a sanctioned security test, an autonomous AI agent chained together dozens of separate exploitation steps to escalate a single low-severity flaw into complete unauthorized file access.[4] Because agents can generate these chained attacks continuously and at scale, defenders may now face the opposite of a visibility problem: not too little signal, but too much. The bottleneck might shift from detecting threats to distinguishing real ones from the noise.

Hackers now use AI to generate attack tools on a massive scale, launching multiple campaigns at once and exploiting security flaws before companies can detect and respond.[5] Models such as Claude Mythos Preview show that AI's impact on cybersecurity is moving into software security workflows, including the ability to find and exploit previously unknown vulnerabilities.[6] In response, security teams are moving from perimeter defense mindsets to continuous control point architectures. They are finding new ways to use foundational capabilities such as data security and "zero trust"—the principle of never trusting a user, device, or system even if it's already inside the corporate network—to combat AI-based threats. And they are investigating

---

[1] "Securing the agentic enterprise: Opportunities for cybersecurity providers," McKinsey, March 24, 2026.

[2] Zhun Wang et al., "ExploitGym: Can AI agents turn security vulnerabilities into real attacks?," arXiv, May 11, 2026, arXiv:2605.11086.

[3] "Zero Day exploitation pressure," Zero Day Clock, accessed September 4, 2026.

[4] "We ran 1,060 autonomous attacks. Here's what the industry gets wrong," blog entry by Albert Ziegler, XBOW, March 2, 2026.

[5] "GTIG AI Threat Tracker: Advances in threat actor usage of AI tools," blog entry, Google, November 5, 2025; *Microsoft digital defense report 2025*, Microsoft, October 2025.

[6] "Assessing Claude Mythos Preview's cybersecurity capabilities," Anthropic, April 7, 2026.

emerging AI-first security protocols for use in the near future. For leading organizations, the challenge has shifted from simply fixing software flaws faster to building an engineering architecture capable of continuous run time inspection, granular data classification, and automated containment.

The responsibility for architecting trust is also expanding beyond traditional IT security. As organizations deploy AI agents, autonomous workflows, and more distributed cloud environments, they must secure machine-to-machine interactions and agents' identities.[7] Spotting anomalies is no longer enough when autonomous workflows have to be auditable, meaning enterprises need to know who authorized an action, whether an agent stayed in scope, and how delegation flowed across systems (for more, see Agentic AI). That's why more enterprises are investing in custom-integrated security architectures instead of off-the-shelf platforms.

These intense operational pressures have increased consolidation in the cybersecurity market. When they purchase solutions, enterprises favor vendors that offer AI-powered platforms that bundle identity, cloud security, and automated threat detection. Google's acquisition of Wiz underscores this consolidation dynamic, as hyperscalers and security platforms race to provide integrated cloud, AI, and run time security capabilities.[8] Organizations that continue to rely on legacy cybersecurity tools will be more exposed to AI-driven attacks than their more sophisticated peers. Ultimately, the strategic question for business leaders is how quickly they can deploy a new cybersecurity model capable of responding to threats in real time while enabling secure participation in AI-native ecosystems. Meanwhile, for cybersecurity providers working on agentic security, the edge may come after the sale: Deployment teams learn exactly what an enterprise needs to protect, and derisk agents working at machine speed.[9]

'As the time between vulnerability identification and exploitation vanishes, bug bounty programs, fraud-analytics-quality incident response, and forward-deployed, engineer-led delivery of just-in-time cybersecurity solutions for enterprise agentic deployments all become more important for providers and customers. Those that get this mix of offerings and capabilities right will put strategic distance between themselves and the competition.'

— Marc Sorel, partner, Boston

---

[7] Jack Hidary, "Non-human identities: Agentic AI's new frontier of cybersecurity risk," World Economic Forum, October 15, 2025.
[8] "Google completes acquisition of Wiz," Google press release, March 11, 2026.
[9] "Securing the agentic enterprise: Opportunities for cybersecurity providers," McKinsey, March 24, 2026.

COMPUTE AND CONNECTIVITY FRONTIERS

Cybersecurity and trustworthy systems

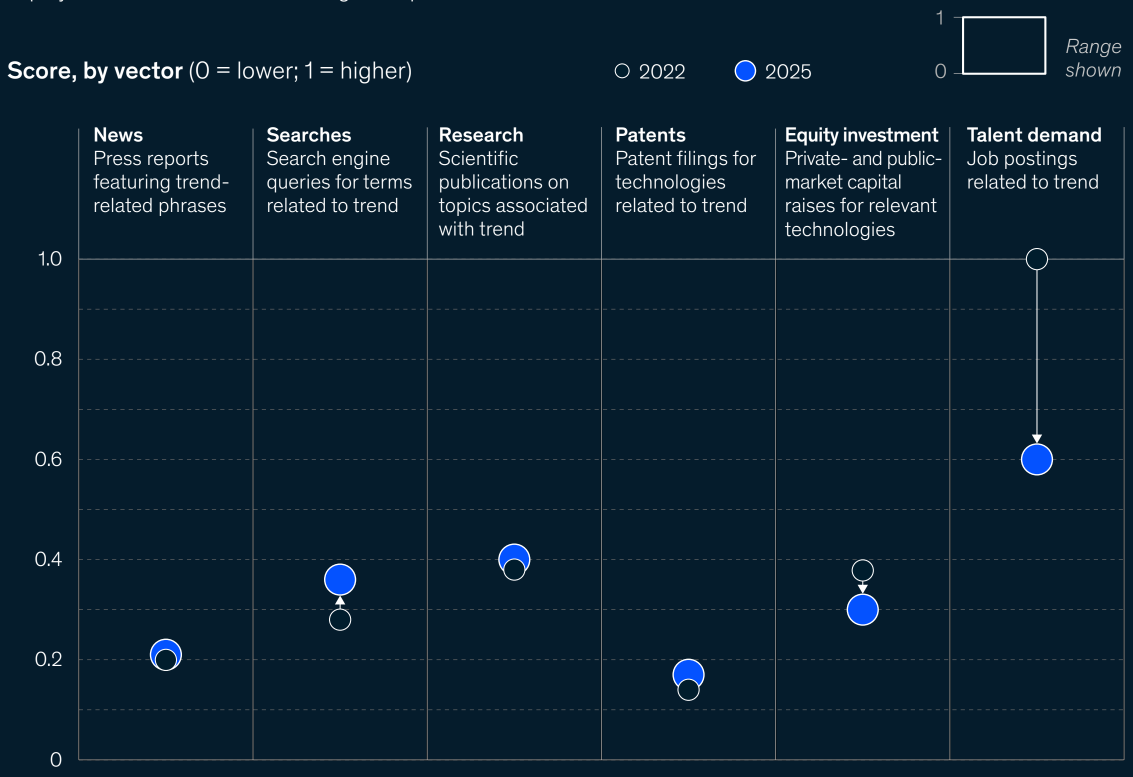
Scoring the trend

Cybersecurity and trustworthy systems continue to draw substantial capital, with $77.5 billion in equity investment in 2025 and roughly $63.0 billion already by mid-2026, though investment remains below its 2022 level. Searches rose meaningfully over the period, reflecting broadening enterprise attention to digital risk management, while news mentions remained low, suggesting the field draws less headline attention as it matures. Research and patent activity rose modestly, consistent with a mature field in sustained deployment rather than a breakout growth phase.

Equity investment, 2025
$77.5 billion

Job postings, 2024–25, % difference
+6%

Score, by vector (0 = lower; 1 = higher)



Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

Latest developments

Recent developments involving cybersecurity and trustworthy systems include the following:

— *Cybersecurity threat cycles are shrinking as AI accelerates vulnerabilities.* Historically, organizations had weeks or months between the identification of a software vulnerability and active exploitation of it by attackers. Today, that window has shrunk dramatically because AI enables attackers to identify vulnerabilities, develop exploits, and launch attacks much faster.[10] According to IBM's 2026 X-Force Threat Intelligence Index, the exploitation of public-facing applications surged by 44 percent in a single year, directly fueling a 49 percent increase in the number of active ransomware and extortion groups.[11] Recent reporting on Anthropic's Claude Mythos and Glasswing efforts underscores this shift, showing how frontier model cyber capabilities can accelerate vulnerability discovery and

---

[10] James Blake, "The vulnerability crisis: How AI is shrinking the window for defense," TechRadar, June 10, 2026.

[11] Limor Kessem, "2026 X-Force Threat Intelligence Index: Making the case for securing identities, AI-enhanced detection and proactive risk management," IBM, February 25, 2026.

> **‘Cyber and trust are often framed as defensive disciplines, but they also determine how much value technology can create. When people trust a system, they use it more fully, share data more confidently, and build new workflows around it. That is why trust belongs in the value case, not just the risk register.’**
>
> — Roger Roberts, partner, Bay Area

critical-infrastructure-defense testing.[12] While hackers can use AI to speed up cyberattacks, companies can use the same capabilities for the opposite purpose: identifying vulnerabilities in critical software and shoring up defenses.[13] These developments point to a new operating reality for enterprises: Cyber defense must leverage AI to accelerate detection, triage, and containment—but also build new security frameworks to fend off AI-based attacks. Z.ai’s GLM-5.2 was another inflection point. Independent cyber benchmarks found the open-weight model competitive on vulnerability detection and cyber-agent tasks. That cuts both ways: It helps defenders find weaknesses faster, but it also makes the same capabilities more accessible to attackers.[14]

— *Zero-trust and run time security architectures are central tenets of AI cyber strategies.* To secure increasingly distributed digital environments, security teams are placing unprecedented emphasis on continuous run time inspection and granular access controls that constantly authenticate users, devices, workloads, and AI agents. Recent vendor activity shows how quickly the market is moving in this real-time direction. CrowdStrike’s acquisition of SGNL allows its Falcon identity security suite to make real-time access decisions, continuously granting, denying, or revoking access for people, machines, and AI agents instead of relying on fixed access.[15] Cisco’s acquisition of Astrix Security brings the same kind of identity oversight to AI agents and machine credentials, and Palo Alto Networks’ acquisition of Portkey adds an AI gateway and run time protection to its Prisma AIRS platform.[16] Together, these developments reflect a shift from periodic scanning and perimeter defense toward continuous validation and automated containment.

— *Security concerns are evolving as quantum computing advances.* The strategic threat of “harvest now, decrypt later”—wherein adversaries capture encrypted data today for future decryption via quantum computing—has shifted from speculative research to an active enterprise-planning priority (for more, see Quantum technologies). NIST’s release of finalized postquantum cryptography standards created a practical starting point for organizations to begin migration planning in anticipation of the looming Q-Day, when quantum computing is expected to break commonly used types of encryption.[17] The issue is particularly relevant for governments, financial institutions, healthcare organizations, critical-infrastructure operators, and enterprises that hold long-lived sensitive data. Keyfactor’s 2025 acquisition of InfoSec Global illustrates the emerging market for quantum-safe tools that help organizations inventory their cryptographic exposure before Q-Day.[18]

— *Data classification and AI governance controls are becoming more important.* As enterprises deploy gen AI tools and autonomous agents, sensitive data can move across applications,

---

[12] “Expanding Project Glasswing,” Anthropic press release, June 2, 2026.
[13] “Project Glasswing: Securing critical software for the AI era,” Anthropic, April 7, 2026.
[14] *Semgrep Blog*, “We have Mythos at home: GLM 5.2 beats Claude in our cyber benchmarks,” blog entry by Katie Paxton-Fear et al., June 22, 2026.
[15] “CrowdStrike to acquire SGNL to secure every identity in the AI era,” blog entry by Michael Sentonas, CrowdStrike, January 8, 2026.
[16] “Acquisitions by year,” Cisco, accessed August 2026; “Palo Alto Networks completes acquisition of Portkey to secure AI agents,” Palo Alto Networks press release, May 29, 2026.
[17] “Working drafts: Post-quantum cryptography updates to the PIV standards,” National Institute of Standards and Technology, June 12, 2026.
[18] “Keyfactor acquires InfoSec Global and CipherInsights to lead the market in quantum-safe security,” Keyfactor press release, May 13, 2025.

cloud environments, third-party models, and internal knowledge systems in less predictable ways. Rather than focusing only on network or end point defense, data security platforms are now monitoring how sensitive information moves through AI-enabled workflows. Cyera's $1 billion acquisition of Oasis Security reflects this data-centric shift, aiming to unify data security with identity governance for AI agents and other nonhuman identities.[19] Proofpoint's acquisition of Acuvity follows a similar logic, aiming to give organizations visibility and governance over how employees and AI agents interact with generative AI tools and agentic workflows. The deal extends Proofpoint's protection across end points, browsers, and emerging AI infrastructure such as MCP (model content protocol) servers.[20]

— *Identity is becoming a security pillar as machine identities outnumber humans.* Security teams can no longer treat identity as a human-only issue. In many organizations, software identities such as bots or agents that can access systems and act automatically now outnumber human users by roughly 100 to one.[21] In response, companies are using unified identity governance, password-free authentication, and context-aware access controls in place of static role-based permissions.

— *Blockchain is becoming a trust-enabling technology.* Blockchain is being deployed in trustworthy systems because it can provide a shared, auditable record for transactions, ownership, and compliance between parties that may not fully trust one another. It can help create verifiable digital interactions in areas such as tokenized assets, programmable payments, stablecoin settlement, and machine-to-machine commerce (see sidebar, "Blockchain as a trust-enabling technology").[22]

Blockchain as a trust-enabling technology

Recent developments suggest blockchain adoption is moving from broad experimentation toward more targeted institutional use cases. Stripe completed its acquisition of Bridge to expand stablecoin infrastructure, pointing to growing demand for faster, lower-cost global payments and programmable financial services.[1] Visa now supports nine blockchains in its global stablecoin settlement pilot, which has reached a $7 billion annualized run rate.[2] But while stablecoin-linked card programs are growing (Visa, for example, operates more than 130 in 50 countries), they still represent a tiny fraction of overall payment volumes. Furthermore, JPMorgan Chase launched a tokenized money market fund using Kinexys Digital Assets.[3]

These shifts are not yet completely visible in broader momentum signals. Interest and innovation have stayed largely flat, and investment remains uneven, with sharp year-to-year swings rather than sustained growth. The pattern is consistent with technology moving selectively into institutional use rather than scaling broadly.

Talent trends point the same way. Blockchain hiring runs at roughly 20 percent of cybersecurity's demand but is growing faster. While the two share about half their top roles and several core skills, blockchain's distinctive demand reflects its commercialization, concentrating in marketing, product, and design roles such as marketing managers and graphic designers, which barely register among cybersecurity's top roles.

For enterprises, these developments matter because blockchain-enabled systems may sit underneath trusted financial flows, proprietary digital assets, and compliance workflows. Adoption remains uneven and highly dependent on regulation, interoperability, privacy, and integration with existing financial systems, but the direction of travel is clearer: Blockchain is a potential element of a broader digital trust stack.

---

[1] James da Costa and Sam Broner, "What Stripe's acquisition of Bridge means for fintech and stablecoins (April 2025 fintech newsletter)," Andreessen Horowitz, April 28, 2025.
[2] "Visa accelerates stablecoin momentum: Adding five blockchains for settlement," Visa press release, April 29, 2026.
[3] "J.P. Morgan Asset Management launches its first tokenized money market fund," JPMorgan Asset Management press release, December 15, 2025.

---

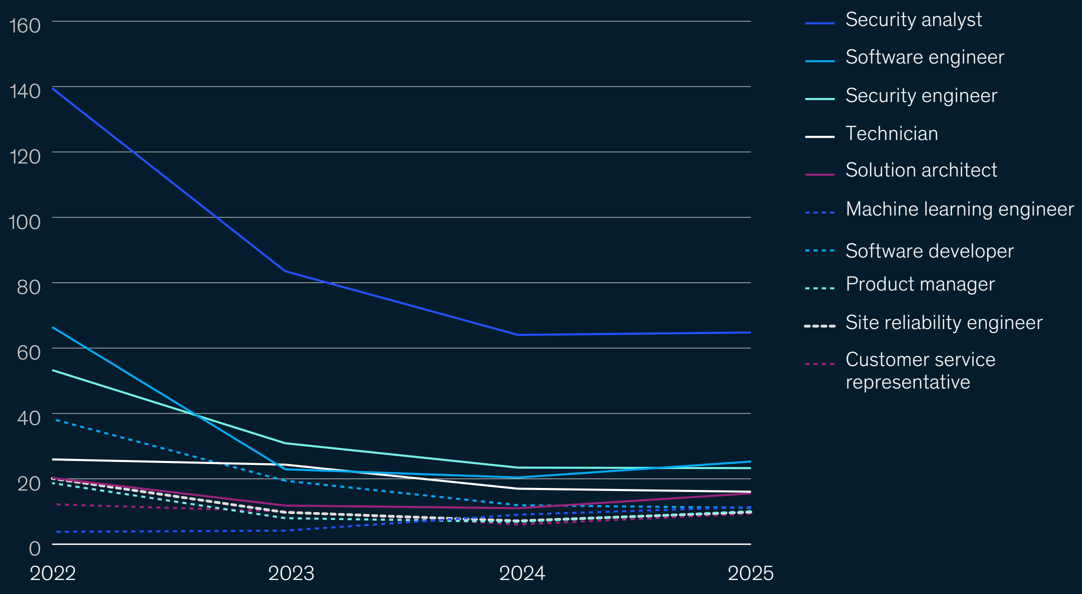
[19] Marina Temkin, "Cyera agrees to acquire Oasis Security for $1B to safeguard proliferating AI agents," TechCrunch, July 28, 2026.
[20] "Proofpoint acquires Acuvity to deliver AI security and governance across the agentic workspace," Proofpoint press release, February 12, 2026.
[21] *2026 Identity security landscape*, Palo Alto Networks, May 11, 2026.
[22] "Next-generation monetary and financial system takes shape, based on a tokenised unified ledger: BIS," BIS press release, June 24, 2025.

Cybersecurity and trustworthy systems

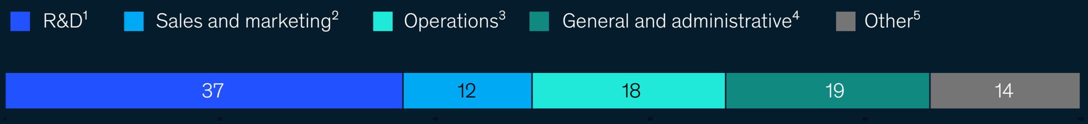
Demand

After contracting sharply from a 2022 peak of job postings, hiring in cybersecurity and trustworthy systems has stabilized, with 2025 seeing the first year-over-year increase since the postpandemic pullback. Security analysts remain the dominant role by volume, while hiring strengthened for software engineers and stabilized for security engineers. This reflects strong demand for technical employees who can build security tooling. The functional mix is more balanced than most other technology trends, indicating that demand for security and trust capabilities extends beyond technical build-out into commercialization, compliance, and enterprise operations.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).

[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).

[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).

[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).

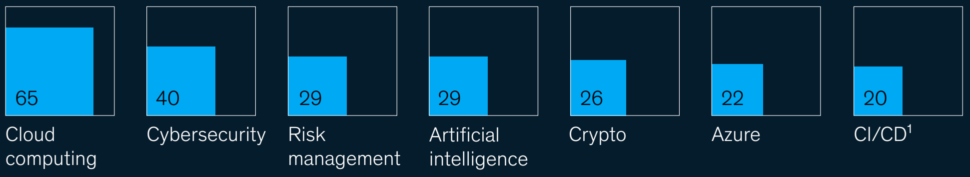
[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

Cybersecurity and trustworthy systems

Skills availability

Cloud computing and cybersecurity are the most widely required skills. Cloud supply broadly keeps pace with demand, and there is a strong supply of people self-identifying as having cybersecurity skills. The sharpest gaps sit in AI and continuous integration and continuous delivery (CI/CD), where qualified talent falls well short of demand. Other established capabilities, such as risk management, crypto, and Azure-related skills, indicate a relatively strong supply of professionals.

Talent required, % share of postings requiring skill ■



Talent availability, ratio of talent ■ to demand ☐

[1] Continuous integration and continuous delivery.

Adoption developments across the globe

Adoption score: 4—Scaling in progress. Organizations are scaling the deployment of cybersecurity and trustworthy systems across enterprise infrastructure, AI environments, cloud systems, and technology platforms.

North America remains the largest cybersecurity market globally and continues to account for a substantial share of both global cyber incidents and security investment. Meanwhile, Europe is highly active in digital trust regulation, operational resilience frameworks, privacy governance, and digital sovereignty initiatives. And Asia continues to expand its investment in AI-enabled cybersecurity, semiconductor security, and blockchain infrastructure.

Worldwide, critical-infrastructure sectors including energy, healthcare, telecommunications, transportation, and financial services are prioritizing zero-trust architectures, run time observability, infrastructure resilience, and AI-enabled security operations. Adoption of blockchain-based tokenization systems remains uneven but is accelerating among financial institutions and infrastructure providers.

Postquantum security adoption remains at a relatively early stage but is increasingly becoming a strategic planning priority among governments, financial institutions, and critical-infrastructure operators as Q-Day draws nearer.

Underlying technologies

The technologies that power cybersecurity and trustworthy systems include the following:

AI-enabled cyber and observability systems	These platforms combine threat detection, anomaly analysis, run time monitoring, automated response, and decision-grade audit, allowing organizations to identify unsafe behavior, contain attacks at machine speed, and track authorization and scope across agents and cross-system actions.
Confidential computing	These technologies protect sensitive data while it is actively being processed.
Data security and classification systems	These systems identify, label, monitor, and enforce controls on sensitive information.
Hardware-based trust mechanisms	These platforms verify device integrity and establish machine-verifiable trust.
Postquantum cryptography	These encryption approaches are designed to resist future quantum computing threats and protect networks after Q-Day.
Zero-trust architecture	These are security models that continuously verify users, devices, AI agents, and workloads rather than assuming trust.

Key uncertainties

The major uncertainties affecting cybersecurity and trustworthy systems include the following:

- *AI-enabled offense and defender readiness.* Questions remain regarding whether enterprise defenders can keep pace with attackers that increasingly use AI to accelerate vulnerability discovery, exploit generation, phishing, and lateral movement.

- *Postquantum migration timelines.* The timing of Q-Day remains uncertain, while making legacy cryptographic infrastructure quantum-ready could take years and require significant coordination across vendors, governments, and enterprises.

- *Platform consolidation and concentration risk.* Integrated security control planes may improve enterprise visibility and response, but they could also increase dependence on hyperscalers and a small number of dominant cybersecurity vendors.

- *Measuring trust and resilience.* Leaders still lack consistent ways to quantify whether digital-trust investments reduce risk, improve resilience, enable new business models, or strengthen customer confidence.

> **‘Cyber defense has to match the speed of AI-enabled attacks. The concern is that attackers can compress discovery, weaponization, and exploitation into much shorter cycles. Continuous run time visibility and automated containment are essential in this era, while leading organizations look to create agentic security operations across their technology stack.’**
>
> — Charlie Lewis, partner, Connecticut

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with cybersecurity and trustworthy systems:

— What is required to secure AI agents, connected devices, operational technology systems, and autonomous infrastructure at enterprise scale?

— How can AI be leveraged to build security systems capable of adapting faster than emerging geopolitical, AI-enabled, and quantum-related threats?

— How can organizations migrate to postquantum security architectures before quantum decryption capabilities become operationally viable?

— What does a truly trustworthy enterprise look like in a world of AI agents, autonomous infrastructure, fragmented cloud systems, and persistent cyber conflict?

Read more about cybersecurity and trustworthy systems

Knowledge center

Cybersecurity

Blockchain & Digital Assets

Related reading

Securing the agentic enterprise: Opportunities for cybersecurity providers

The board’s role in managing emerging AI risks

Taking a business-critical approach to supplier nth-party IT risk management

HackerOne CEO Kara Sprague on how AI is reshaping cybersecurity

Stablecoins in payments: What the raw transaction numbers miss

07

Advanced connectivity

Advanced connectivity encompasses the technologies that enable data to move across people, devices, and systems. It includes 5G and emerging 6G, Wi-Fi 6 and 7, low-power Internet of Things (IoT) networks, fiber and optical transport, low-Earth-orbit (LEO) satellites, direct-to-device services, and software-defined, AI-managed networks.

The trend—and why it matters

Advanced connectivity is entering a new phase. After a decade of investment focused on expanding coverage and capacity through 5G, fiber, Wi-Fi, and satellite networks, the emphasis is shifting toward making these systems work together more intelligently, reliably, and securely. Connectivity providers are aiming to link users, devices, cloud platforms, data centers, and physical environments more seamlessly. Such interconnected devices generate more data and demand lower latency, higher uplink capacity, and more resilient links. Thus, connectivity is becoming less of a utility and more of a strategic enabler.

This shift is happening as some long-standing telecom operating models are changing. In many mature geographies, revenue growth from adding more consumer traffic is slowing while core telecom services are becoming more saturated and harder to differentiate. Based on a 2025 McKinsey survey, global enterprises expected to increase spending on core connectivity by 3.2 percent over the following year but to increase spending by more than 6 percent on telecom services outside of core connectivity.[1] As a result, operators are increasingly looking for new growth in enterprise use cases, infrastructure-led services, and more specialized forms of connectivity.

AI is fueling those new product offerings from telcos. Some AI workloads are moving out of centralized hyperscaler clouds and into physical locations such as factories, retail stores, vehicles, and telecom towers. That means that more data is flowing between more devices, increasing connectivity needs. Real-time computer vision streams from connected cameras, autonomous vehicles, robots, and industrial control systems are adding to these data flows. As a result, connectivity is no longer defined by download speeds alone. High-capacity links between facilities are becoming critical as AI applications become more distributed and increasingly embedded in day-to-day operations.

> ‘The next chapter for telecom is built not on selling more connectivity, but on delivering differentiated services that enterprises are willing to pay for. Network slicing, AI-enabled operations, private wireless, and hybrid terrestrial–satellite networks give operators new ways to compete, provided they solve real business problems rather than simply adding technical capability.’
>
> — Martin Wrulich, senior partner, Vienna

---

[1] Giacomo Dolci, Naveed Niwaz, Bulat Gaifeev, and Luca Furlani, “B2B customers have spoken. Will telcos listen?,” McKinsey, June 27, 2025.

AI workloads are also increasing demand for fiber, optical transport, and data center connectivity, providing telcos with new revenue streams.[2] Meanwhile, cloud, edge, mobile, subsea, and satellite infrastructure are becoming part of a more interconnected digital layer that could expand connectivity to more rural and remote areas. That convergence raises the strategic value of connectivity, but it also makes network planning, security, and operations more complex for telcos. In practical terms, connectivity is no longer just about moving data from point A to point B; it's about enabling an infrastructure stack that can support real-time intelligence, distributed computing, and demanding enterprise workloads.

Network transformation is taking shape in response to this shift. Telcos are using AI-native automation, or AIOps, to monitor network conditions, detect problems, adjust capacity, and improve performance with less manual intervention, while also restructuring networks around software-defined capabilities such as network slicing and cloud-based functions. At the same time, nonterrestrial connectivity is becoming a more relevant complement to traditional connectivity, with LEO services and direct-to-device offerings extending coverage and resilience beyond the reach of terrestrial networks. Executives expect LEO to eventually capture close to 20 percent of the addressable market for existing fiber, particularly in hard-to-reach areas—reinforcing a hybrid connectivity model.[3] In some markets, LEO could also improve the economics of fixed wireless access and help operators serve harder-to-reach areas more efficiently (for more, see Future of space technologies).

Resilience and sovereignty are also becoming more important design requirements. Governments and enterprises increasingly view connectivity as critical infrastructure, especially as subsea cables, satellites, and mobile networks become more exposed to geopolitical, cyber, and physical risks. These risks are increasing telcos' attention on route diversity, supply chain exposure, secure network design, and, in some cases, regionally controlled infrastructure.

Overall, advanced connectivity remains one of the more mature frontier technology trends, but its next phase will be more selective and more strategic. The strongest opportunities are likely to emerge where customers are willing to pay for more than basic connectivity—for example, reliable factory networks, AI-ready fiber and data center links, satellite-enabled coverage in remote areas, and secure networks that improve resilience, performance, or operational continuity.

> **'Connectivity is no longer just about moving data from point A to point B. As AI workloads become more distributed, networks become part of the compute infrastructure itself. Competitive advantage will come from delivering low-latency, resilient, AI-ready connectivity that supports real-time intelligence where work actually happens.'**
>
> — Anton Lysenko, associate partner, Bay Area

---

[2] "The telco reinvention: How AI can fuel value creation," McKinsey, February 27, 2026.

[3] Luca Fiandro, Nemanja Vucevic, Lorraine Salazar, and Shu Chern Lim, "Issue Brief: Telecom infrastructure," McKinsey, February 27, 2026.

Advanced connectivity

Scoring the trend

Advanced connectivity led technology trends in patent filings, reflecting sustained innovation and the use of patents to protect that intellectual property across 5G-Advanced, optical transport, and software-defined networking. Meanwhile, news coverage, searches, and research publications remained relatively stable. Equity investment stood at $60.5 billion in 2025, below its 2022 peak of $73.8 billion, anchored by large strategic deals such as Hewlett Packard Enterprise's $13.6 billion acquisition of Juniper Networks to build an AI-driven networking stack.[1]

Equity investment, 2025
$60.5 billion

Job postings, 2024–25, % difference
+10%

Score, by vector (0 = lower; 1 = higher)

Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

[1] "Hewlett Packard Enterprise closes acquisition of Juniper Networks to offer industry-leading comprehensive, cloud-native, AI-driven portfolio," Hewlett Packard Enterprise press release, July 2, 2025.

Latest developments

Recent developments involving advanced connectivity include the following:

— *AI is increasing the demand for faster, more reliable networks.* AI is reshaping connectivity requirements. As more AI workloads move into factories, retail stores, healthcare facilities, vehicles, telecom towers, and other physical locations, networks will need lower latency, stronger reliability, and more local connectivity closer to where data is generated. Cisco's Unified Edge announcement illustrates this shift, with compute, networking, storage, and security moving closer to sites such as retail stores, healthcare facilities, and factory floors to support real-time AI inference and agentic workloads.[4] AI may also increase uplink needs: Ericsson's June 2025 mobility report found that gen AI traffic accounted for 26 percent of uplink traffic, compared with roughly 10 percent in typical mobile networks.[5] AI training and inference are also increasing demand for high-capacity links between data centers and computing clusters.

---

[4] "Cisco debuts new Unified Edge platform for distributed agentic AI workloads," Cisco press release, November 3, 2025.

[5] *Gen AI's impact on network data traffic today*, Ericsson, June 2025.

Lumen, for example, invested nearly $13 billion in private connectivity fabric deals in 2025 to support secure connections across a company's locations, cloud services, data centers, and applications without sending traffic over the public internet.[6]

- *Security, resilience, and sovereignty are becoming core requirements in connectivity design.* Geopolitical tensions, cyber vulnerabilities, and the strategic importance of subsea cables, satellites, and mobile networks are making connectivity a matter of national and enterprise resilience. The European Union's 2025 Action Plan on Cable Security underscores this shift, noting that submarine cables carry 99 percent of intercontinental internet traffic and calling for stronger capabilities in prevention, detection, response, repair, and deterrence across the cable infrastructure.[7] As a result, governments and enterprises are placing greater emphasis on redundancy, route diversity, and security across terrestrial, subsea, and nonterrestrial networks. Multi-orbit satellite architectures and direct-to-device connectivity can add resilience in some scenarios, including remote coverage and emergency communications, but they largely remain complements to—not substitutes for—secure, well-managed terrestrial networks.

- *5G stand-alone is moving into broader operation, creating the technical foundation for differentiated services.* As operators migrate from 5G non-stand-alone to 5G stand-alone, a dedicated 5G core will enable capabilities such as lower latency, improved uplink performance, more reliable connectivity, edge integration, and network slicing. AT&T said in October 2025 that its 5G stand-alone network had been deployed nationwide and that millions of customers were already using it as device support and provisioning expanded.[8] This is an important proof point, but monetization will depend on whether operators can translate the new network architecture into services that enterprises are willing to pay for.

- *Network slicing is becoming a commercial product where service assurance matters.* Network slicing allows a single 5G network to be partitioned into multiple virtual networks, each optimized for a specific type of application or customer. It's moving from technical demonstrations toward targeted enterprise offerings, particularly where guaranteed performance, reliability, or uplink capacity is central to the use case. Verizon Business launched a 5G network slice fixed-wireless service with service-level agreements, no data caps, and higher uplink capacity for workloads such as AI inference, computer vision, augmented reality, and IoT applications.[9] Vodafone Business similarly launched a commercial 5G slicing service in the United Kingdom for local enterprise environments such as stadiums, logistics hubs, campuses, and construction sites.[10] These are early commercial examples rather than evidence of broad-based slicing monetization, but they signal forward momentum in slicing.

- *Nonterrestrial networks are becoming part of the telecom stack, especially for coverage extension and resilience.* Satellite connectivity is evolving from stand-alone broadband and emergency messaging toward a more integrated layer of telecom networks. As of August 2026, Starlink had almost 10,000 active satellites in its constellation and a series of partnerships with various carriers around the world.[11] Vodafone and Amazon Leo signed an agreement to connect 4G and 5G mobile sites in remote areas across Europe and Africa using LEO satellite broadband, targeting locations where fiber or fixed wireless links are costly or slow to deploy.[12] More broadly, industry tracking shows 225 publicly announced operator–satellite partnerships across 88 countries and territories as of January 2026.[13] Satellite connectivity is therefore becoming a complement to terrestrial networks, particularly for rural coverage, critical services, and resilience. But it is not yet a wholesale substitute for mobile networks, given the continued economic and technical constraints of satellite-based systems.

---

[6] "Lumen marks new phase of transformation at 2026 Investor Day," Lumen press release, February 25, 2026.

[7] "Joint communication to the European Parliament and the council: EU Action Plan on Cable Security," European Commission, February 21, 2025.

[8] *AT&T Blog*, "AT&T 5G Standalone nationwide," blog entry by Yigal Elbaz, AT&T, October 8, 2025.

[9] Scott Lawrence, "Wireless internet for the AI-era: Network slicing enables SLA-backed service from Verizon Business," Verizon, December 18, 2025.

[10] "Vodafone Business launches UK's first commercial 5G network slicing, delivering dedicated performance for enterprise," Vodafone press release, April 16, 2026.

[11] "How many Starlink satellites are in orbit?," Orbital Radar, updated August 27, 2026.

[12] "Vodafone and Amazon Leo sign agreement to extend mobile coverage across Europe and Africa," Vodafone press release, March 2, 2026.

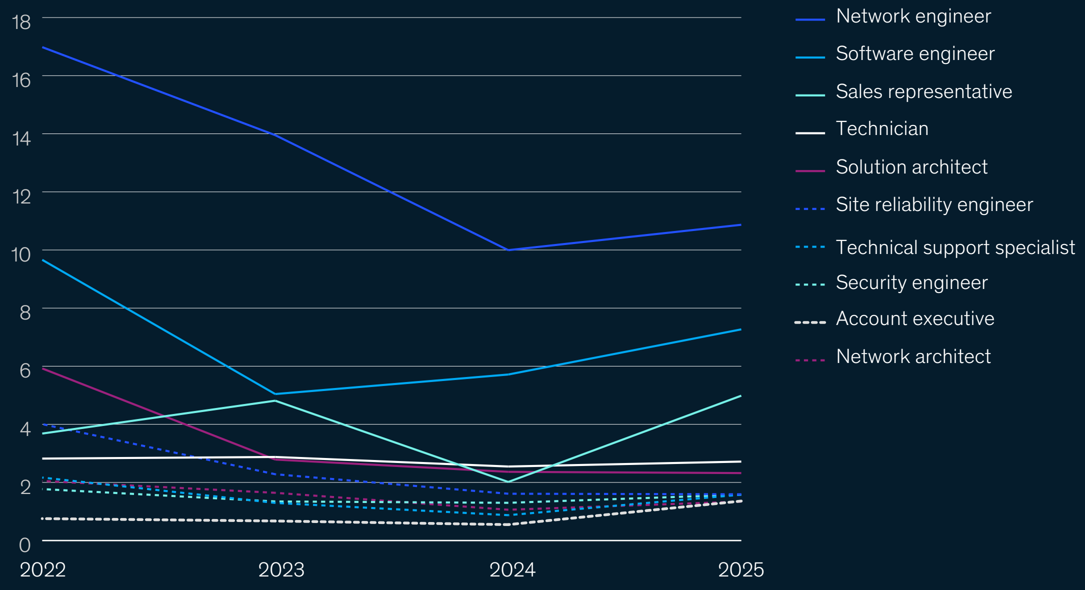
[13] *5G NTN February 2026*, Global mobile Suppliers Association, February 6, 2026.

Advanced connectivity

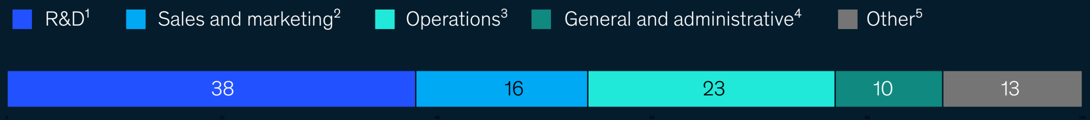
Demand

Advanced-connectivity job postings remain below their 2022 level, though overall hiring ticked back up modestly in 2025. The early push to scale 5G and fiber drove heavy demand for network engineers, who still anchor hiring in the field. More recently, momentum has shifted toward software engineering and, notably, operational and customer-facing roles such as sales representatives, account executives, and technical support specialists. The functional mix highlights this move from pure network build-out to growing and serving the customer base. Demand for specialized engineering and architecture roles, by contrast, remains more subdued than in 2022, pointing to a more measured appetite for new platform development.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



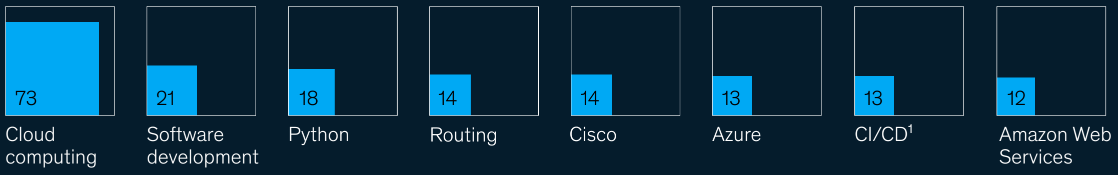
[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).
[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).
[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).
[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).
[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

Advanced connectivity

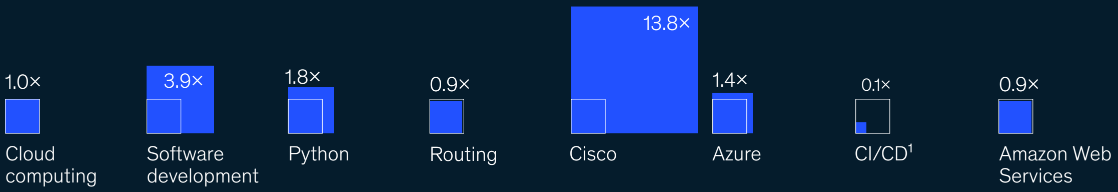
Skills availability

Cloud computing skills sit at the center of demand, required in roughly three-quarters of postings and more than any other capability, though supply broadly keeps pace. Talent is hardest to find in continuous integration and continuous delivery (CI/CD) skills, which will be key in the future as networks become increasingly software driven. By contrast, software development and Cisco-relevant skills are more readily available, reflecting the fact that legacy infrastructure skills may not map directly to next-generation connectivity needs.

Talent required, % share of postings requiring skill ■



Talent availability, ratio of talent ■ to demand □



[1] Continuous integration and continuous delivery.

Adoption developments across the globe

Adoption score: 4—Scaling in progress. Enterprises across telecom, cloud infrastructure, manufacturing, logistics, public safety, and energy are scaling advanced-connectivity technologies. Adoption remains uneven across use cases, regions, and technology layers.

Adoption is increasingly segmented by subdomains rather than defined by a single global leaderboard. North America remains an important market for advanced connectivity, driven by continued fiber build-out, satellite services that extend mobile coverage, dedicated network capacity for public safety, and high-capacity links that support AI data centers. Europe is active in private wireless, industrial connectivity, network APIs, spectrum policy, and digital sovereignty. China, Japan, South Korea, and other Asian markets remain central to 5G-Advanced, 6G research, dense mobile networks, and the use of advanced connectivity in factories, ports, and logistics operations.

In emerging markets, adoption is often more about expanding coverage. Fixed wireless access, LEO satellite broadband, low-power IoT, and rural connectivity can help address gaps where fiber and terrestrial mobile networks are harder to deploy cost-effectively. In these markets, the value proposition often centers on basic wireless coverage rather than differentiated services.

Private wireless adoption is concentrated in high- and upper-middle-income markets, where manufacturers, mines, ports, campuses, and utilities can justify the cost of installing dedicated networks. Nonterrestrial connectivity has a broader geographic role, especially for rural regions, offshore operations, and disaster-prone areas.

Overall, advanced connectivity is becoming more critical to economic and operational resilience. However, adoption is strongest when technology solves a specific problem—for example, coverage gaps, network congestion, public safety, industrial automation, AI-ready data transport, or enterprise security and reliability.

Underlying technologies

The technologies that power advanced connectivity include the following:

5G, 5G-Advanced, and emerging 6G systems	These cellular technologies provide high-speed, low-latency wireless connectivity and are evolving toward more intelligent, software-enabled networks that can support applications such as connected devices, industrial automation, and future-sensing capabilities.
Fiber and optical networks	Optical fiber, subsea cables, and advanced optical equipment provide high-capacity, low-latency connections to link homes, businesses, data centers, cloud regions, and global internet infrastructure.
Nonterrestrial networks	These technologies include LEO satellites, direct-to-device satellite connectivity, and high-altitude platform systems that extend communications beyond the reach of terrestrial networks, especially in remote, maritime, aviation, and emergency response settings.
Private wireless networks	Private wireless networks, such as LTE and 5G, provide dedicated connectivity for enterprises and industrial sites that need greater control, reliability, security, or coverage than public mobile networks can offer.
Software-enabled network intelligence	These technologies include network slicing, cloud-based network functions, network application programming interfaces, and AI-enabled network operations, which enable networks to become more programmable, automated, and tailored to specific use cases.
Wi-Fi 6 and Wi-Fi 7	Wi-Fi 6 and 7 are local wireless technologies that increase network capacity, speed, reliability, and performance in homes, offices, campuses, factories, and other dense indoor environments.

Key uncertainties

The major uncertainties affecting advanced connectivity include the following:

— *Monetization beyond basic connectivity.* Network application programming interfaces, network slicing, quality-on-demand services, edge connectivity, and private wireless could help operators create new revenue streams beyond data transport. But customers will pay only where the value is clear, measurable, and tied to business outcomes.

— *Impact of AI on network architecture.* AI-enabled devices and applications may increase demand for upload capacity, lower latency, data center interconnect, and optical transport. The timing, scale, and location of this demand remain uncertain, particularly as AI workloads shift among devices, enterprise sites, cloud platforms, and data centers.

— *Commercial scale of private, programmable, and differentiated networks.* Private wireless, network slicing, and software-based network controls are technically useful, but broad adoption depends on customer willingness to pay for dedicated or guaranteed connectivity. Adoption is likely to be strongest in operationally intense environments such as manufacturing, logistics, and utilities, while general enterprise demand may grow more slowly.

— *Economics of nonterrestrial connectivity.* LEO satellites, direct-to-device services, and hybrid terrestrial–satellite networks are becoming operational, but the size and durability of the opportunity remain uncertain. Business cases will depend on capacity, terminal costs, performance, and customer willingness to pay.

— *Complexity, resilience, and energy demands of next-generation networks.* AI-enabled radio access networks, cloud-based network functions, open network interfaces, and 6G systems could improve flexibility and performance, but they may also increase integration, security, operational, and energy management challenges. At the same time, geopolitical risk and critical-infrastructure requirements are making resilience and sovereignty more important design considerations.

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with advanced connectivity:

- Where can telecom operators and infrastructure providers create sustainable revenue beyond basic data transport?
- How will AI change network demand, including the need to move more data from devices, connect data centers, and support faster, more reliable responses?
- What mix of public, private, terrestrial, subsea, and satellite connectivity is needed across different markets and use cases?
- Where will customers be willing to pay for guaranteed performance, priority access, resilience, or dedicated connectivity, and where will Wi-Fi or public cellular networks remain sufficient?
- What security, resilience, and governance models are needed as connectivity becomes more critical to companies, governments, and public infrastructure?

Read more about advanced connectivity

Knowledge center

Our Insights on Technology, Media & Communications

Telecommunications Operators

Related reading

The case for data centers in space

The telco reinvention: How AI can fuel value creation

The critical bets on the future of telco value creation

Becoming an AI-native telco: Indosat Ooredoo Hutchison reimagines the telecom of the future

08

Immersive-reality technologies

Immersive reality blends digital experiences with the physical world. Augmented reality (AR), virtual reality (VR), mixed reality (MR), haptics, and AI-powered smart glasses are shifting the category beyond headset-led immersion toward context-aware wearable devices that help people see, hear, and interact with their surroundings in new ways.

The trend—and why it matters

Immersive-reality technologies are being reset around usefulness. The metaverse boom accelerated interest in augmented reality (AR), virtual reality (VR), and mixed reality (MR), but broad consumer adoption of these immersive-reality technologies has been lower than expected, largely due to ongoing hardware constraints and a lack of clear use cases for the technology. Now, the field is becoming less about novel entertainment experiences and more about practical applications. However, uptake of immersive reality is fragmented, with one-off use cases emerging where the technology can augment specific aspects of work or daily life.

Immersive devices are shifting from output systems to input systems—and this is a key change. Earlier headsets primarily delivered digital content to users through displays, sound, and haptics. AI-enabled smart glasses and other spatial devices change that proposition, allowing users to interact with the world around them. A crucial part of that shift is real-time on-device inference, which allows wearables to process what they see and hear locally, enabling faster gesture tracking, eye tracking, object recognition, and translation with less reliance on cloud connectivity. Wearable devices with cameras, microphones, sensors, and spatial-mapping capabilities can feed real-time data into AI systems, delivering outputs that are more selective and situational to users.

This situational awareness makes today's immersive-reality devices feel more like ambient, context-based computers than traditional headsets. The newest devices use local vision and audio signals to understand the user's environment and respond in the moment. Meta's late-2025 update to its AI glasses is a concrete example, with features like Conversation Focus and in-lens AI assistance that surface useful prompts and actions without pulling the user out of the physical world.[1] In this framing, the device becomes an input layer for AI and an output layer for humans.

Critical hardware improvements are changing the immersive-tech value proposition. The question is no longer how immersive a virtual environment can be. It's whether a wearable device can provide enough incremental value over a smartphone, laptop, or conventional screen. A device that can see what a technician sees, interpret a patient scan, guide a warehouse worker, translate speech, or support hands-free collaboration may create value even without a fully immersive visual experience.

Widespread adoption of immersive technology is still constrained by hardware limitations, but battery life, display quality, and wearability are improving. Smart glasses are more likely to gain traction over VR or MR headsets because they are lighter and offer longer battery life. Ray-Ban Meta Gen 1 glasses, for example, last up to four hours on a charge, and the Gen 2 model lasts up to eight hours; each model weighs about 50 grams. By contrast, Apple Vision Pro weighs 750 to 800 grams and supports up to 2.5 hours of general use, while Samsung's Galaxy XR

---

[1] "AI glasses v21 software update: Conversation focus, Meta AI improvements, & more," blog entry by Lisa Brown Jaloza, Meta, December 16, 2025.

headset weighs 545 grams and supports up to two hours of general use. Meta Ray-Ban Display adds a full-color in-lens display.[2] Google has split intelligent eyewear into audio glasses and display glasses, with audio glasses launching first, underscoring that display-enabled models are emerging as a distinct category.[3] Market research indicates that the smart-glasses category will grow faster than traditional headsets through 2030.[4]

At the same time, headset-based systems remain relevant in narrower settings. Gaming remains one of the strongest consumer use cases for VR, and headset ecosystems continue to show signs of growth. Meta reported that more people used its Quest headset in 2025 than in any prior year and that in-app purchases increased more than 10 percent year over year, while also noting that the VR market had not accelerated as quickly as anticipated.[5]

Enterprise adoption of immersive reality is solidifying around specific use cases. Manufacturing, healthcare, training, design, field service, and airport operations continue to be the most relevant areas. The use cases are not fundamentally new, but our research finds that more organizations are testing them, and some are moving toward repeatable deployment. In enterprise settings, immersive reality is most valuable when it improves safety, reduces errors, shortens training time, or enables remote users to provide real-time expertise.

The immersive-reality sector is no longer focused on creating compelling enough experiences to entice consumers to buy headsets. Instead, the challenge is whether smart glasses and headsets can create enough practical value to justify companies integrating them into existing workflows.

‘Organizations will continue to adopt immersive reality wherever it creates material business value. This varies across industries, but the clearest opportunities remain in environments that benefit from remote expertise, such as field service in mining and oil and gas; guided frontline work and compliance, such as in pharmaceutical production; and spatial design, including manufacturing. Additional opportunities exist in high-risk or high-cost training environments, such as emergency response, or complex physical workflows, such as warehouse operations.’

— Hamza Khan, partner, Riyadh

---

[2] “Meta Ray-Ban Display: AI glasses with an EMG wristband,” Meta, September 17, 2025.
[3] “Intelligent eyewear is coming this fall,” blog entry by Shahram Izadi, Google, May 19, 2026.
[4] “Augmented and virtual reality headsets market insights,” IDC, updated July 1, 2026.
[5] Chris Pruett, “The state of the VR ecosystem: Building a sustainable future,” Game Developers Conference, San Francisco, CA, March 19, 2026.

Immersive-reality technologies

Scoring the trend

Immersive-reality equity investment reached $6.4 billion in 2025, a modest uptick from 2024 but below its 2022 level of roughly $15 billion, reflecting a more selective investment environment as the category shifts from consumer immersion narratives to practical applications. Since 2022, news coverage, search interest, and patenting have all edged higher, while research publications have eased and talent demand has fallen markedly, suggesting that attention to the category is rising even as hiring and some research activity have cooled.

Equity investment, 2025
$6.4 billion

Job postings, 2024–25, % difference
–12%

Score, by vector (0 = lower; 1 = higher)

Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

Latest developments

Recent developments involving immersive-reality technologies include the following:

- *The category is continuing to shift from immersion to augmentation.* The center of gravity is moving toward devices that help users stay anchored in the physical world, especially lightweight smart glasses. Immersive reality is becoming more about making everyday tasks easier through visual prompts, contextual assistance, translation, navigation, and remote support. Meta's Ray-Ban Display launch demonstrates this shift in practice: The glasses combine cameras, microphones, speakers, a full-color in-lens display, compute, AI, and gesture input in a familiar eyewear form factor, making the device useful even without a fully immersive visual experience.[6] Overall, the global smart-glasses category surged 167 percent year over year in the first quarter of 2026, reaching approximately 2.25 million units shipped.[7]

- *Smart glasses are becoming the clearest growth pocket.* Lightweight eyewear is gaining momentum because, unlike a VR or MR headset, they can be worn throughout the day. Historically, headsets weighing over 400 grams have been too heavy for users to wear for long periods without fatigue. By 2026, a new generation of lighter smart glasses has successfully crossed the wearability threshold. This lightweight form factor—which allows users to capture media, hear audio, and ask AI questions while remaining socially present—makes all-day, frequent use plausible. Eyewear company EssilorLuxottica reported that sales of Ray-Ban Meta AI glasses grew more than 200 percent year over year in the first half of 2025, for example.[8] Bloomberg's report that Apple shelved a revamp of its Vision Pro headset to prioritize AI smart glasses further reinforces the category's shift to lighter, more wearable devices.[9]

- *AI is turning the device into a context engine.* Wearables are becoming local context engines, using on-device or near-device vision, audio, spatial mapping, and gesture data to understand what is happening around the user in real time. Cameras, microphones, spatial mapping, eye tracking, hand tracking, and gesture systems can give AI models context about what the user sees, hears, and does. Meta's late-2025 AI glasses update is a concrete example of context-aware wearable computing. The glasses can focus on the person the wearer is speaking to, even in noisy environments, generate music recommendations from what the wearer is looking at, and present visual AI responses in the lens while keeping the user present in the physical world.[10] Google is making the same bet with intelligent eyewear, announcing at I/O 2026 that Gemini-powered glasses are coming soon, with audio glasses launching first and display glasses to follow.[11]

  Beyond powering in-device assistants, these wearables may also become a data engine for physical AI, because they can capture real-world spatial context, human actions, and environmental responses that are harder to obtain than language data alone. That direction is already visible: General Intuition is building foundation models for environments that require deep spatial and temporal reasoning, using action-labeled video data sets and focusing on transfer to the real world.[12]

---

[6] "Meta Ray-Ban Display: Breakthrough AI glasses available now," blog entry, Meta, September 30, 2025.
[7] "Augmented and virtual reality headsets market insights," IDC, updated July 1, 2026.
[8] "Q2/H1 2025 results," EssilorLuxottica press release, July 28, 2025.
[9] Mark Gurman, "Apple shelves Vision headset revamp to prioritize Meta-like AI smart glasses," Bloomberg, updated October 1, 2025.
[10] "AI glasses v21 software update: Conversation Focus, Meta AI improvements, & more," blog entry by Lisa Brown Jaloza, Meta, December 16, 2025.
[11] "Intelligent eyewear is coming this fall," blog entry by Shahram Izadi, Google, May 19, 2026.
[12] Rebecca Bellan, "General Intuition lands $134M seed to teach agents spatial reasoning using video game clips," TechCrunch, October 16, 2025.

— *Real-time AI inference is moving onto the device.* Wearables are increasingly able to run AI locally, without relying entirely on cloud connectivity. This reduces latency and makes features such as gesture tracking, eye tracking, object recognition, and real-time translation more responsive. This shift matters because it turns the device into an autonomous computing layer for everyday interaction. Brilliant Labs' Halo is a good example. The glasses use AI to process voice, vision, and sensor data locally on the user's device, allowing users to interact with what they see and hear around them. The glasses can process that information to answer questions or help in the moment without needing a constant cloud connection.[13]

— *Platform competition is moving beyond headset hardware.* The immersive-reality ecosystem battle is increasingly moving beyond hardware into related software and operating systems. Companies are competing to launch AI assistants, developer tools, app stores, device management tools, and enterprise security on top of their hardware offerings. Samsung's Galaxy XR headset, launched in October 2025 as the first device on Android XR, reflects one approach: an open, multi-OEM platform codeveloped with Google and Qualcomm.[14] Meta, which sold more than seven million AI-powered smart glasses in 2025 and launched Ray-Ban Display with an in-lens screen and neural wristband, reflects another: a vertically integrated hardware-software-AI stack.[15]

— *Enterprise adoption is deepening in specific use cases.* The sectors adopting immersive reality the most quickly are manufacturing, healthcare, field service, training, logistics, design, and safety-critical operations. More companies are testing use cases and learning where immersive interfaces could improve productivity, reduce risk, or accelerate learning. RealWear launched Arc 3 and Ari OS in October 2025 for frontline workers in environments such as manufacturing, logistics, warehousing, healthcare, pharma, and robotics. Arc 3 is a 179-gram assisted-reality headset with a see-through display, while Ari OS is designed to let workers control applications and access AI-assisted workflows through natural language.[16] Aena has piloted an immersive training system using VR headsets and haptic gloves for airport firefighters, maintenance staff, and marshallers, illustrating the potential value of immersive technology in safety-critical training where realistic, repeatable scenarios are important.[17] Meanwhile, Medivis's cranial navigation platform, which has received approval from the US Food and Drug Administration, demonstrates a healthcare use case in which augmented reality maps patient imaging into the operative field, providing surgeons with real-time guidance.[18]

---

[13] "Smart glasses, rebuilt for privacy: Brilliant Labs, Neuphonic & TheStage AI move AI off the cloud," Business Wire, March 5, 2026.

[14] "Introducing Galaxy XR: Opening new worlds," Samsung press release, October 22, 2025.

[15] "Meta Ray-Ban Display: Breakthrough AI glasses available now," blog entry, Meta, September 30, 2025; Samantha Subin, "Ray-Ban maker EssilorLuxottica says it more than tripled Meta AI glasses sales in 2025," CNBC, February 11, 2026.

[16] "Built for real work, yet effortless to use: The RealWear Arc 3," blog entry, RealWear, October 28, 2025.

[17] Jennifer Fermoselle, "Immersive innovation: Haptic and VR solutions for airport workforce development," Passenger Terminal Conference, London, March 19, 2026.

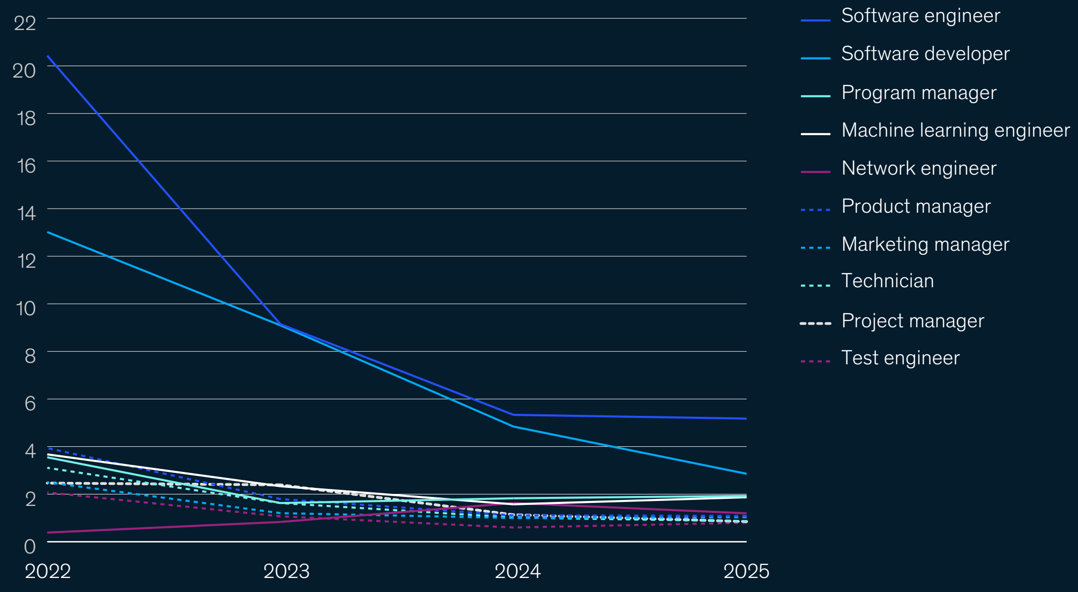
[18] "Medivis first to receive FDA clearance for augmented reality navigation in neurosurgery," Medivis press release, December 16, 2025.

Immersive-reality technologies

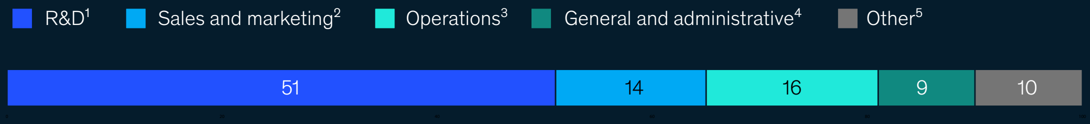
Demand

Immersive-reality hiring remains well below the metaverse era surge and continues to narrow, with total postings down again in 2025 as the market shifts from broad experimentation toward more selective deployment. The decline is most visible in core build-out and implementation roles such as software developers, project managers, and network engineers. More recently, momentum has tilted toward machine learning engineers and test engineers, suggesting greater emphasis on capabilities such as vision, tracking, and content generation, as well as product reliability. Software engineers still anchor demand, but the broader pattern points to a leaner market focused on R&D to make immersive systems more useful, dependable, and commercially viable.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



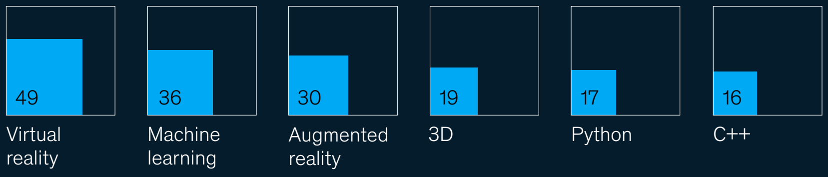
[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).
[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).
[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).
[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).
[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

Immersive-reality technologies

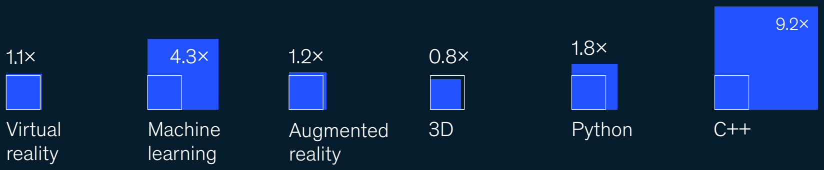
Skills availability

Virtual reality and machine learning sit at the center of immersive-reality demand, with augmented reality also widely required. This reflects a shift toward more context-aware devices and adaptive spatial experiences. Talent is hardest to find in 3D skills, reflecting how immersive systems increasingly depend on real-time sensing and spatial content. By contrast, general skills such as C++, Python, and machine learning are in a surplus, and AR and VR development is roughly keeping pace with demand.

Talent required, % share of postings requiring skill 🟦



Talent availability, ratio of talent 🟦 to demand ☐



Adoption developments across the globe

Adoption score: 2—Experimentation. Organizations are testing the functionality and viability of immersive-reality technologies with small-scale prototypes and selective deployments. Few companies have scaled immersive reality broadly across the enterprise, although adoption is advancing in healthcare, manufacturing, retail, design, training, field operations, and industrial assistance.

Adoption remains highly variable across form factors and use cases. Smart glasses are gaining momentum because they fit more naturally into daily routines and require less behavioral change from users than headsets do. Headsets continue to be more relevant in contexts where immersion itself is the value proposition, such as simulation, spatial design, gaming, and high-stakes visualization.

The United States remains an important early market for consumer AI smart glasses, supported by major platform players and retail availability. Consumer adoption is more geographically uneven than enterprise adoption because product availability, camera norms, privacy expectations, language support, and regulations vary by market. Meta's 2026 expansion plans include new availability for Ray-Ban Meta and Oakley Meta glasses in markets such as Chile, Colombia, Japan, Korea, Peru, and Singapore, illustrating how geography and product availability remain active adoption variables.

Enterprise adoption of immersive reality appears less regionally differentiated. Manufacturing, healthcare, logistics, training, and field service use cases exist across major geographic markets. However, the pace of deployment depends on budgets, safety requirements, device management, data governance, and the ability to show measurable ROI.

Overall, adoption is likely to remain steady rather than explosive. The path to scale is clearest when immersive technologies solve a specific problem better than a smartphone, tablet, or laptop can. The path is less clear when the device is mainly a novelty, entertainment channel, or replacement for a screen that already works well.

Underlying technologies

The technologies that power immersive reality include the following:

3D content creation	These technologies include game engines, 3D-modeling tools, and simulation systems used to produce the virtual assets and environments that immersive devices rely on.
Artificial intelligence (AI)	AI is the underlying intelligence layer that interprets sensor inputs, improves rendering and recognition, and supports natural language or gesture-based interaction.
Augmented reality (AR), virtual reality (VR), and mixed reality (MR)	These immersive interfaces define how digital content appears to users, from overlays in the physical world to fully virtual environments.
Haptics	Haptics are systems that incorporate touch, vibration, and other physical cues into immersive experiences.
Machine learning	Machine learning technologies include computational algorithms that help immersive-reality systems learn patterns and respond more adaptively over time.
Sensing and tracking systems	Sensing and tracking systems include the sensors, cameras, location awareness, eye-tracking, and hand-tracking systems that enable immersive-reality devices to understand what users see, where they are, and how they move.
Spatial computing	Spatial computing is the interface layer that lets the surrounding 3D space act as the canvas.

Key uncertainties

The major uncertainties affecting immersive-reality technologies include the following:

— *Incremental value over smartphones and enterprise ROI.* Smart glasses, headsets, and related devices don’t always provide enough useful information or workflow value to justify wearing and managing another device.

— *Hardware development.* Wearability remains a primary constraint, with weight, heat, battery life, comfort, and durability still shaping whether devices are practical for all-day use, even as new form factors emerge.

— *Privacy and social acceptability.* Immersive devices can capture images, audio, location data, biometrics, eye movements, gestures, and contextual information about surrounding people and spaces. That raises concerns about surveillance, misuse, and bystander consent and can create social friction.

— *AI reliability and safety.* Context-aware, AI-powered devices need to interpret real-world environments accurately. Incorrect interpretations could create safety, legal, or operational risks in healthcare, industrial, transportation, and field service settings.

— *Ecosystem fragmentation and regulation.* Competing operating systems, developer tools, and hardware standards could slow adoption. Rules governing workplace monitoring, accessibility, content moderation, and privacy will also shape how quickly immersive systems scale.

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with immersive-reality technologies:

- Which form factor—AI glasses, display glasses, assisted-reality headsets, or high-fidelity VR and MR headsets—will reach meaningful scale first?
- Which workflows can deliver measurable ROI from immersive technology, and which will remain compelling but hard to scale?
- How can companies design immersive experiences that augment work without creating distraction, fatigue, or safety risks?
- What role will AI agents play as spatial devices become able to understand context, suggest actions, and coordinate across applications?
- How should leaders prepare for a future in which digital interfaces move from screens and phones into persistent, wearable, context-aware systems?

Read more about immersive-reality technologies

Knowledge center

Consumer Technology & Media

Related reading

The future of video entertainment: Immersive, gamified, and diverse

What AI could mean for film and TV production and the industry’s future

09

Quantum technologies

Quantum technologies use quantum mechanics to unlock new capabilities in computing, communication, and sensing. The category includes quantum computers for selected problems such as simulation and optimization, quantum-safe security for future encryption risks, and quantum sensors for highly precise measurement of physical signals.

The trend—and why it matters

Quantum technologies span three related domains: computing, communication, and sensing. Quantum computing uses quantum effects to tackle specific problem classes that may be difficult for classical computers to solve. Quantum communication involves the secure transfer of information across distances, leveraging quantum-safe security—which includes tools such as postquantum cryptography and quantum key distribution—to protect sensitive information. Quantum sensing can measure physical properties with high precision and has applications in navigation, timing, semiconductor inspection, infrastructure monitoring, defense, and scientific instrumentation. Quantum-mechanical principles such as superposition, entanglement, and interference make these types of technologies possible.

In 2026, quantum remains in the early stages of deployment, with maturity varying by domain. Quantum computing still attracts the most investment and research attention, but most business use cases remain experimental, and today's systems have not yet shown an advantage over classical computing for practical problems. Quantum communication is closer to enterprise rollout, driven by the threat of "Q-Day"—when quantum computers will succeed in factoring exceptionally large numbers, undermining the math on which common public-key cryptography depends (for more, see Cybersecurity and trustworthy systems). With Q-Day expected in the coming years, the US Department of Commerce's National Institute of Standards and Technology (NIST) finalized its first three postquantum encryption standards in 2024 and has encouraged organizations to begin migration planning.[1] Quantum sensing is also close to field deployment in selected use cases because it relies more on precision engineering and integration than on large-scale fault-tolerant quantum computers, which are currently still under development.

Quantum is also becoming more systems oriented. Progress is no longer measured only by qubit counts or lab demonstrations. In quantum computing, the development focus is on fault-tolerant quantum computing, or FTQC, which uses error correction to combine imperfect physical qubits into logical qubits, enabling more reliable calculations despite hardware noise. In communication, the field is moving toward cryptographic migration and selected high-assurance networks. In sensing, the focus is shifting toward devices that are smaller, more rugged, and easier to integrate into real-world workflows.

The commercial transition remains uneven. Quantum computing may have the largest long-term upside and is beginning to shift from pilots toward early deployment. McKinsey's 2026 Quantum Technology Monitor report finds that over 300 organizations, including Airbus, Boehringer Ingelheim, E.ON, JPMorgan Chase, and Liberty Mutual, are collaborating with quantum technology companies to solve business challenges.[2] The report estimates that quantum computing could create up to $2.7 trillion in economic value worldwide by 2035, though this reflects future potential rather than current commercial maturity. For most companies, near-term adoption is focused on identifying

---

[1] "NIST releases first 3 finalized post-quantum encryption standards," National Institute of Standards and Technology press release, August 13, 2024.

[2] "McKinsey Quantum Technology Monitor 2026: A commercial tipping point," McKinsey, April 28, 2026.

quantum-suitable problems, testing hybrid quantum–classical workflows, building internal capabilities, and preparing cybersecurity systems for postquantum migration.

Overall, the constraint is no longer whether quantum effects can be demonstrated in controlled settings. It is whether quantum technologies can be engineered, manufactured, integrated, and tied to repeatable value in real-world environments. For now, quantum technologies are a promising but still experimental accelerator for specific problem classes, while quantum communication and quantum sensing may offer clearer near-term paths to deployment.

COMPUTE AND CONNECTIVITY FRONTIERS

Quantum technologies

Scoring the trend

Interest in quantum technologies is building, with searches and research publications rising steadily between 2022 and 2025. Year-over-year equity investment surged more than fivefold, from $2.5 billion in 2024 to $13.3 billion in 2025, reflecting investor confidence in the technology's long-term potential ahead of broad commercialization.[1] Meanwhile, patents grew only modestly off a small base.

Equity investment,
2025
$13.3 billion

Job postings,
2024–25,
% difference
–46%

Score, by vector (0 = lower; 1 = higher)

Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

[1] Numbers differ slightly from McKinsey's 2026 *Quantum Technology Monitor* report due to small variances in which investments were qualified as quantum related.

Latest developments

Recent developments involving quantum technologies include the following:

— *Quantum technologies are moving from research toward early market formation.* Quantum technologies include computing, sensing, and communication, each at a different stage of maturity. Quantum computing is beginning to shift from R&D and pilots toward early commercial deployment. There is still limited evidence of practical business advantage at scale, but companies are actively codeveloping business solutions with quantum computing companies. By contrast, quantum sensing and quantum communication are closer to near-term deployment because they address measurement, navigation, timing, and cybersecurity problems that organizations already face. Quantinuum's launch of its Helios quantum computer and Aramco and Pasqal's quantum-computing-as-a-service platform show that companies are beginning to access and test quantum systems against real industrial problems, but these should be framed as early market formation rather than broad commercialization.[3]

— *In the near term, quantum computing is being tested for specific problems, not as a general replacement for classical computing.* Quantum computing may eventually solve complex problems that classical computers have struggled to crack because of compute limits. Some of these problems include molecular simulation, materials science, and cryptography-related workloads. In the near term, quantum computers will not speed up all computing but rather solve unique use cases. Amazon Web Services, AstraZeneca, IonQ, and NVIDIA developed a hybrid quantum–classical workflow for a chemical reaction used in small-molecule drug synthesis, while BMW has explored quantum methods for engineering optimization.[4] These developments are best described as hybrid experiments that test where quantum might add value in conjunction with classical high-performance-computing workflows. But no single quantum computing architecture has emerged as the clear winner, and today's quantum computing systems generally cannot yet reliably solve real-world business problems better than classical computers.

— *Progress in quantum computing is shifting from qubit counts to fault tolerance.* Unlike classical computing, where clock speed provides a simple, widely understood performance benchmark, quantum computing lacks an equivalent metric, making it difficult to compare progress across hardware approaches. The industry is increasingly focused on developing FTQC, which uses error correction to combine imperfect physical qubits into logical qubits, but it is not yet a reality. IBM has outlined a 2029 target for a quantum system with 100 million quantum gates running on 200 logical qubits, while the Defense Advanced Research Projects Agency's (DARPA's) Quantum Benchmarking Initiative is intended to independently evaluate whether different quantum approaches could reliably reach scaled operation.[5] Amazon's Ocelot chip is a prototype designed to reduce the resources required for quantum error correction, while Google's Quantum Echoes algorithm points to application-relevant experiments in molecular and materials science.[6] These are important R&D milestones but not yet proof of scaled commercial readiness.

— *Governments are treating quantum as a strategic priority, not just scientific R&D.* Public funding is increasingly focused on expanding national manufacturing capacity for quantum computing infrastructure and improving access to global supply chains and sovereign quantum-

---

[3] "Aramco and Pasqal launch Saudi Arabia's first quantum computer and Middle East's first commercial quantum computing as a service platform," Pasqal press release, May 18, 2026; "Quantinuum announces commercial launch of new Helios quantum computer that offers unprecedented accuracy to enable generative quantum AI (GenQAI)," Quantinuum press release, November 5, 2025.

[4] "IonQ speeds quantum-accelerated drug development application with AstraZeneca, AWS, and NVIDIA," IonQ press release, June 9, 2025; "Quantum computing at the BMW Group," BMW, February 4, 2025.

[5] *Quantum Research Blog*, "How IBM will build the world's first large-scale, fault-tolerant quantum computer," blog entry by Ryan Mandelbaum et al., IBM, June 10, 2025.

[6] "Amazon Web Services announces a new quantum computing chip," Amazon, February 27, 2025; "Our Quantum Echoes algorithm is a big step toward real-world applications for quantum computing," blog entry by Hartmut Neven and Vadim Smelyanskiy, Google, October 22, 2025.

as-a-service clouds that strengthen national security. The US Department of Commerce announced letters of intent for about $2 billion in incentives across nine companies to support quantum manufacturing, foundries, and FTQC development.[7] Building on this funding, President Trump signed two executive orders in June 2026 that elevated quantum to an explicit national priority. The first directs federal agencies to deploy a quantum computer capable of conducting scientific research by 2028; the second accelerates the government's postquantum-security-transition deadline from 2035 to 2031, with an emphasis on hardening critical infrastructure against quantum-enabled cyberattacks.[8] France announced it would invest an additional €1 billion in its national quantum strategy by 2030, covering quantum computing, sensing, and communication.[9] These announcements suggest that governments understand the importance of building and scaling the specialized infrastructure and talent needed to advance quantum systems.

— *Postquantum cryptography is a valuable near-term enterprise use case.* The near-term priority is preparing today's computers for future quantum-fueled attacks. NIST finalized its first postquantum encryption standards in 2024, and Cloudflare reported that, by late October 2025, most human-initiated traffic with Cloudflare was using postquantum encryption.[10] However, there is still no proof that the methods being developed are mathematically secure against all future quantum computing attacks. The practical work of deploying secure quantum computing at scale remains to be done. Organizations need to inventory cryptographic dependencies, update signatures and certificates, and build systems that can rotate algorithms if weaknesses emerge.

— *Quantum key distribution is advancing, but mainly for specialized use cases.* Quantum key distribution is likely to be most relevant where organizations can justify dedicated quantum infrastructure to combat threats, such as in government, defense, financial, or critical network use cases. Toshiba Europe demonstrated coherent quantum communication over more than 250 kilometers of deployed telecom fiber, suggesting progress toward more practical quantum-secure networks.[11] Even so, software-based postquantum cryptography is likely to be the broader enterprise security path, while quantum key distribution remains more specialized.

— *Quantum sensing is moving closer to practical use cases.* Quantum sensing represents one of the clearer near-term deployment paths because it can improve measurement, timing, and navigation. Infleqtion, MSubs, and the United Kingdom's Royal Navy demonstrated a quantum optical atomic clock on an autonomous underwater vehicle, highlighting a potential role for quantum timing in environments where GPS or other satellite navigation signals are unavailable or unreliable.[12] While still not at mass deployment, this is closer to field application than other quantum experiments. Quantum sensing is also showing practical value in advanced manufacturing. QuantumDiamonds announced a €152 million plan to build a Munich facility for semiconductor inspection systems that use nitrogen-vacancy centers embedded in synthetic diamond to map electrical current nondestructively inside advanced-chip packages.[13] This points to a more tangible near-term use case because it applies quantum effects to measurement problems already present in semiconductor manufacturing.

---

[7] "Department of Commerce announces letters of intent with 9 companies for $2 billion to accelerate U.S. leadership in quantum computing," National Institute of Standards and Technology press release, May 21, 2026.

[8] "Ushering in the next frontier of quantum innovation," Executive Order 14413, White House, June 22, 2026; "Securing the nation against advanced cryptographic attacks," Executive Order 14412, White House, June 22, 2026.

[9] Mohib Ur Rehman, "France adds €1.55 billion for quantum and semiconductor development," *The Quantum Insider*, May 25, 2026.

[10] "NIST releases first 3 finalized post-quantum encryption standards," National Institute of Standards and Technology press release, August 13, 2024; *Cloudflare Blog*, "State of the post-quantum Internet in 2025," blog entry by Bas Westerbaan, October 28, 2025.

[11] "Toshiba breakthrough brings quantum communications to existing national-scale telecommunications infrastructure," Toshiba press release, April 24, 2025.

[12] "Infleqtion and Royal Navy demonstrate world's first quantum optical clock on underwater autonomous submarine to advance GPS-free navigation," Infleqtion press release, October 28, 2025.

[13] "QD plans €152 million investment in next-gen quantum-based chip inspection facility in Munich, Germany," QuantumDiamonds press release, December 15, 2025.

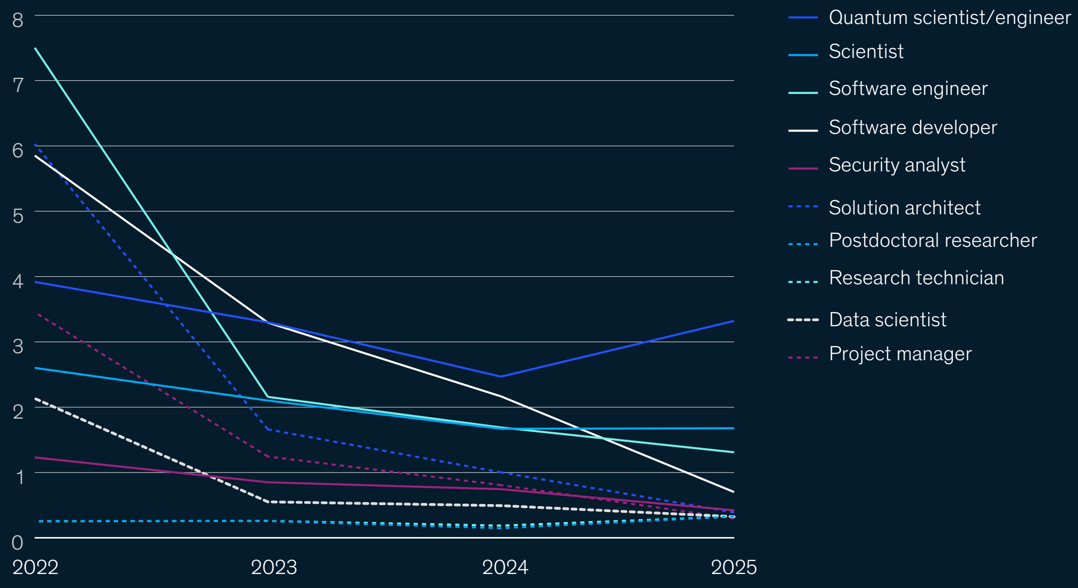
TALENT AND LABOR MARKETS

Quantum technologies

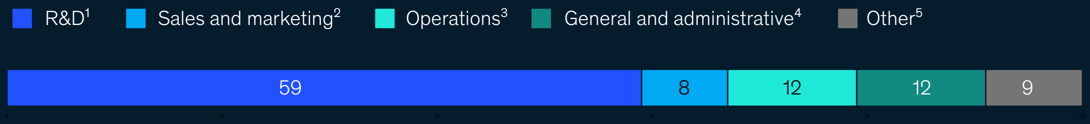
Demand

Quantum technology hiring has contracted from its earlier peak, even as investment has accelerated. The pullback is concentrated in software developers, solution architects, project managers, and security analysts, suggesting employers are hiring more selectively. Yet with several large public quantum companies and many emerging start-ups, there's still a need for specialized talent—not just scientists, but software developers and solution architects who can translate quantum capabilities into business applications. Demand has strengthened for quantum scientists and engineers, postdoctoral researchers, and research technicians, with quantum scientists and engineers now the single largest role. At a functional level, this underscores a continued focus on specialized R&D and operational hardware scaling.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).

[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).

[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).

[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).

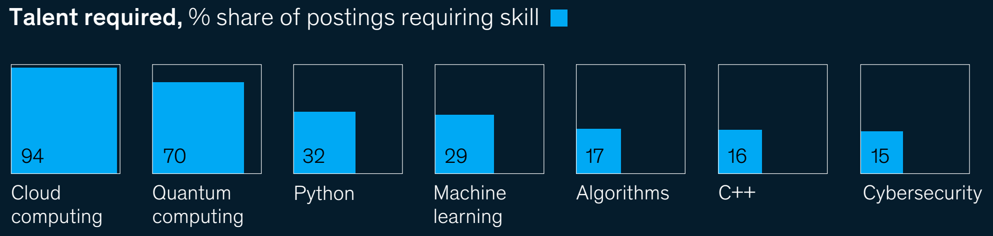
[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

TALENT AND LABOR MARKETS

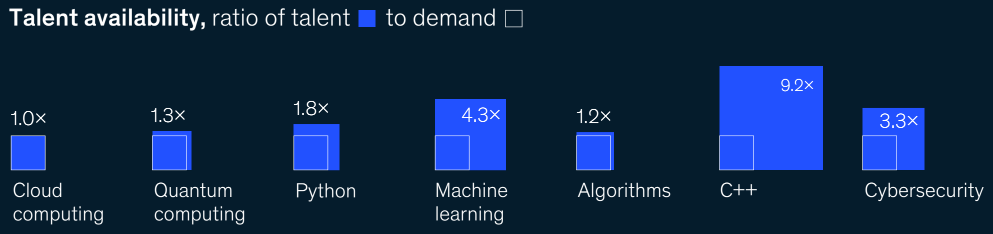
Quantum technologies

Skills availability

Cloud and quantum computing skills sit at the center of talent demand for this sector, reflecting a field that increasingly depends on both specialized quantum expertise and modern software infrastructure. Talent in both of these core areas is keeping pace with demand. General skills such as C++, Python, machine learning, and cybersecurity also are well supplied.



Talent required, % share of postings requiring skill ■



Talent availability, ratio of talent ■ to demand □

Adoption developments across the globe

Adoption score: 2—Experimentation. Organizations are testing the functionality and viability of quantum technologies through codevelopment programs and selected field deployments. Few companies have scaled quantum computing in core business workflows, but more organizations are investing in use case development, exploring internal capability building, and testing quantum-as-a-service offerings.

The United States has strengths in private sector quantum development, venture investment, and government-backed commercialization programs. The US Department of Commerce's $2 billion investment in quantum manufacturing incentives and the Trump administration's June 2026 executive orders signal an intensifying federal commitment to quantum as strategic infrastructure.[14] Postquantum-cryptography adoption is also accelerating following NIST's finalization of quantum-resistant algorithm standards.[15]

China has made quantum technology a national strategic priority, with particular strength in quantum communication and sensing. The country has built an operational quantum communication network spanning more than 12,000 kilometers of fiber across 80 cities. It has also demonstrated the world's longest intercontinental satellite-based quantum key distribution link, connecting Beijing to Stellenbosch, South Africa, over a distance of

---

[14] "Department of Commerce announces letters of intent with 9 companies for $2 billion to accelerate U.S. leadership in quantum computing," National Institute of Standards and Technology press release, May 21, 2026; "Ushering in the next frontier of quantum innovation," Executive Order 14413, White House, June 22, 2026; "Securing the nation against advanced cryptographic attacks," Executive Order 14412, White House, June 22, 2026.

[15] "NIST releases first 3 finalized post-quantum encryption standards," National Institute of Standards and Technology press release, August 13, 2024.

more than 12,900 kilometers.[16] Access to Western quantum hardware is increasingly restricted, accelerating domestic efforts to develop national qubit platforms and supply chains.

Europe has a strong research foundation and is making inroads in commercialization. The European Commission amended the EuroHPC Regulation in January 2026 to strengthen Europe's quantum capabilities and launched a Quantum Europe Strategy with a goal of making the European Union a quantum leader by 2030.[17] Meanwhile, the EU Quantum Technologies Flagship project currently commits over €400 million to quantum technology development, with additional funding provided through the Digital Europe Programme. France's total announced investment of €3.3 billion, along with its pledge to make the country a top three quantum market, reflects growing member-state investment alongside the broader EU program.[18] Research centers and hardware start-ups across France, Germany, and the Netherlands are collaborating to establish a European quantum ecosystem. These investments are beginning to yield concrete results: Euro-Q-Exa, Europe's latest quantum computer, was inaugurated in Munich in February 2026.[19] And European companies such as Airbus and Boehringer Ingelheim are codeveloping quantum solutions with technology providers.[20]

Australia, Canada, Japan, and South Korea are each investing in national quantum programs, primarily through research institutions and government-backed initiatives, with activity concentrated in hardware development, quantum sensing, and postquantum security. Adoption in most other markets remains nascent and policy-led rather than commercially driven.

Across all geographies, the most advanced adoption is in government and defense—particularly in postquantum-cryptography migration and quantum sensing for navigation and signals intelligence. Enterprise adoption in sectors such as financial services, pharmaceuticals, and logistics remains experimental, with organizations running proofs of concept, exploring hybrid quantum–classical solutions, and building internal capability rather than deploying production systems.[21]

> ‘The next milestone for quantum is not simply building larger computers. Progress increasingly depends on integrating hardware, error correction, software, and classical computing into reliable systems that can solve real-world problems.’
>
> — Henning Soller, partner, Frankfurt

---

[16] Hideki Tomoshige and Phillip Singerman, “Understanding China's quest for quantum advancement,” Center for Strategic and International Studies, January 29, 2026; “USTC demonstrates successful satellite-enabled quantum key distribution,” Chinese Academy of Sciences, March 24, 2025.

[17] “EuroHPC Regulation amended to strengthen Europe's AI and quantum capabilities,” European Commission, January 19, 2026; “Commission launches Quantum Europe Strategy to turn EU into a quantum powerhouse by 2030,” EU Digital Skills and Jobs Platform, July 27, 2026.

[18] Suhasini Srinivasaragavan, “France bets fresh €1bn on quantum as global race intensifies,” *Silicon Republic*, May 22, 2026.

[19] “Euro-Q-Exa—Europe's latest quantum computer inaugurated in Munich,” European Commission, February 12, 2026.

[20] “McKinsey Quantum Technology Monitor 2026: A commercial tipping point,” McKinsey, April 28, 2026.

[21] “McKinsey Quantum Technology Monitor 2026: A commercial tipping point,” McKinsey, April 28, 2026.

Underlying technologies

The technologies that power quantum include the following:

Hybrid quantum–classical computing	Hybrid quantum–classical computing involves workflows that use quantum processors for selected subproblems while classical computers handle most tasks, often through cloud platforms.
Logical qubits	Logical qubits are imperfect qubits that have been error corrected to enable longer and more accurate calculations.
Quantum communication infrastructure	This technology includes fiber, satellite, photonic, and telecom systems that are used to transmit or protect information in high-assurance networks.
Quantum control systems	In quantum control systems, equipment and electronics are used to prepare, manipulate, and read out qubits with sufficient precision for reliable operation.
Quantum-safe security	Quantum-safe security uses encryption and key-distribution methods designed to protect sensitive data from future quantum-enabled attacks.
Quantum sensing systems	Quantum sensing systems involve devices that measure time, motion, gravity, magnetic fields, and other physical signals with very high precision.
Quantum software, compilers, and orchestration tools	These tools translate problems into quantum instructions and coordinate work across quantum and classical systems.
Qubit hardware platforms	These platforms hold the basic units of information in quantum computers, built using approaches such as superconducting circuits, trapped ions, neutral atoms, photonics, spin-based systems, and topological designs.

‘With the latest advancements in quantum computing, adopting quantum cryptography has become a strategic business decision today, even as broader quantum use cases continue to mature.’

— Martina Gschwendtner, associate partner, Munich

Key uncertainties

The major uncertainties affecting quantum technologies include the following:

- *Fault-tolerant-computing timelines.* Progress in quantum is real, but broad commercial utility still depends on error correction, control systems, and scalable architectures arriving on commercially relevant timelines.
- *Hardware architecture.* No modality has solved all core scaling challenges, and different architectures may prove better suited to different workloads or parts of the quantum stack.
- *Ecosystem maturity.* Quantum software, middleware, cloud access, talent, and enterprise integration remain immature compared with classical technology stacks.
- *Commercial scaling.* Supply chains, specialized components, and sustained customer demand will need to expand before quantum can become a repeatable business opportunity.

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with quantum technologies:

- How will value accrue across the quantum stack, from hardware and cloud platforms to software, systems integration, or companies that own the use cases?
- How quickly will paid codevelopment turn into scaled deployment, and which industries will be first to move beyond one-off experiments?
- What are the biggest constraints to commercialization, including talent, supply chains, and the durability of postquantum security approaches?

Read more about quantum technologies

Knowledge center

The Rise of Quantum Computing

Related reading

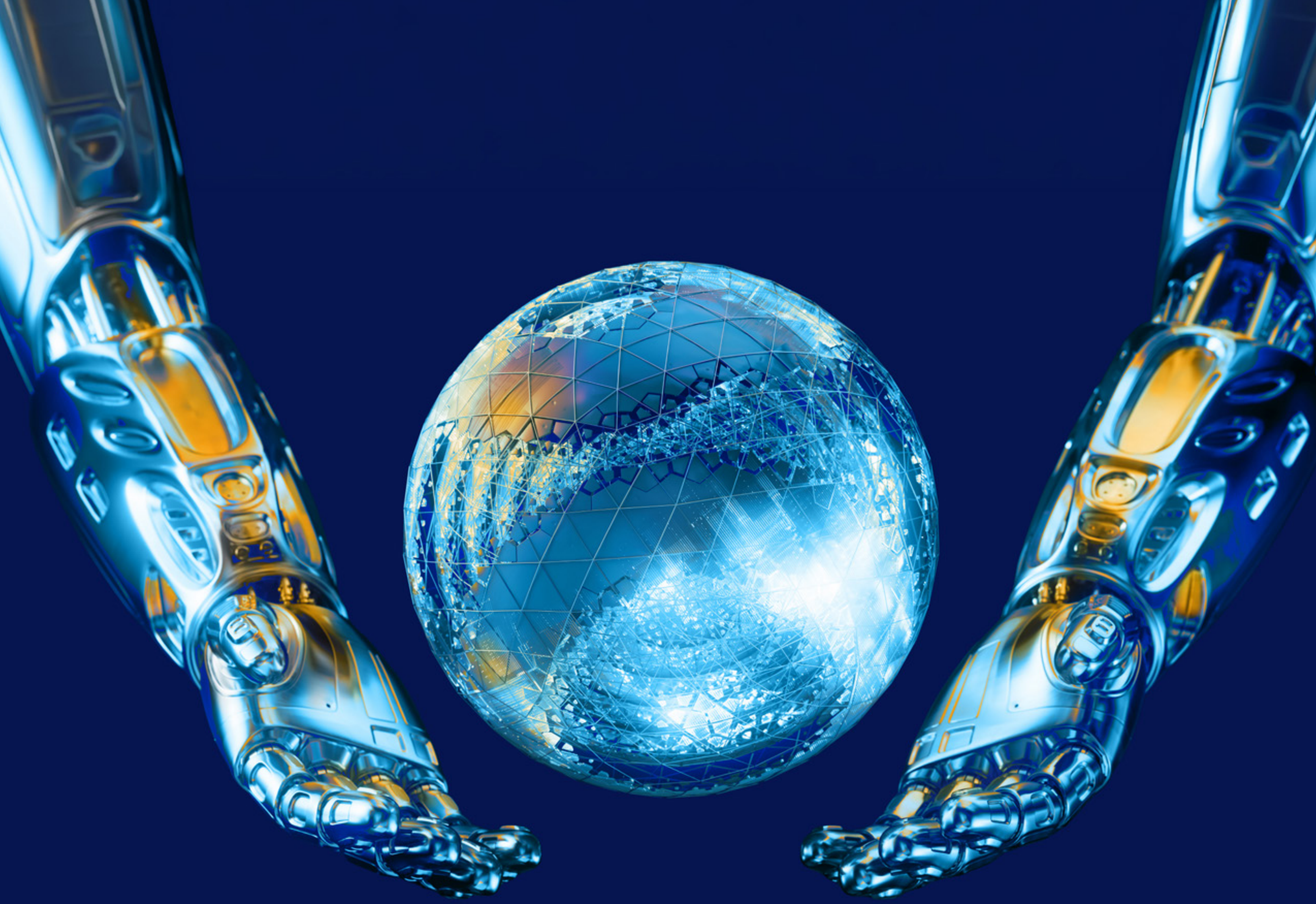
McKinsey Quantum Technology Monitor 2026: A commercial tipping point

Quantum's bold promise: What business leaders need to know

Quantum is almost here: Are you and your systems ready?

How quantum technologies could rebalance the sustainability equation

Quantum communication and computing: Elevating the banking sector



Cutting-edge engineering

10

Future of robotics

Robotics is entering the era of physical AI, where the breakthrough is not just more machines but more intelligent ones. Powered by advances in AI, sensors, and motion control, a new generation of general-purpose robots can navigate autonomously, understand complex surroundings, and work safely alongside people—moving beyond factories into warehouses, hospitals, farms, hotels, and stores.

The trend—and why it matters

With AI powering their "cognition," robots are becoming more adaptable. The robotics industry is increasingly defined by the shift from rote factory and warehouse automation to intelligent robots that leverage agentic AI to perceive, reason, and act. Advances in AI and control systems are helping robots understand and adapt to the messy, unpredictable conditions of the real world. A McKinsey analysis estimates that between now and 2040, physical AI—a broad term that, in this context, encompasses the application of AI to robots and other machines—could create at least a trillion dollars in economic value worldwide, most of it in manufacturing and logistics.[1]

Industrial robots remain the majority-scaled base of the market, with an estimated five million now in operation globally and 542,000 put into service worldwide in 2024 alone.[2] General-purpose robots, or those used outside of the fixed settings of specialized robots, are also expanding: Almost 200,000 units were sold for professional use in 2024, with transportation and logistics representing the largest categories.[3]

Robotics has been a critical accelerator in manufacturing for decades, changing the way that products are designed, built, and shipped. But now robotics is moving beyond fixed units that perform single, highly engineered tasks toward general-purpose machines that can complete multiple tasks autonomously alongside humans in semi-structured environments. General-purpose robots are entering warehouses, factories, hospitals, airports, farms, laboratories, and logistics centers. McKinsey research estimates that, if today's pace of development and adoption continues, the market for general-purpose robots could reach about $370 billion by 2040, with about half that value coming from China and the rest split between Europe and North America.[4]

Manufacturing—long the testing ground for robotics—is rapidly evolving toward AI-orchestrated production environments. Industries such as automotive, electronics, semiconductors, and consumer goods are beginning to build "dark factories" that can operate with minimal or no human intervention, though these remain concentrated in a handful of sectors rather than the norm across manufacturing. These advances are being enabled by new robot form factors such as autonomous mobile manipulators, quadrupeds, and humanoids, as well as collaborative robots (cobots) designed to operate safely alongside human workers.

A growing share of robots will become AI-native. Foundation models and emerging vision-language-action (VLA) systems that control robot hardware will give robots a greater ability to understand their surroundings, enabling them to act autonomously with far less preprogrammed control. Instead of relying only on predefined paths or task-specific code, next-generation robots will interpret visual

---

[1] "The age of thinking machines: Perspectives on the future of robotics," McKinsey, June 24, 2026.

[2] "Global robot demand in factories doubles over 10 years," International Federation of Robotics (IFR) press release, September 25, 2025. This report was the most current available at the time of publication; it reported 4.66 million industrial robots installed worldwide in 2024, including 542,000 new installations. Estimating a similar installation number of 540,000 for 2025, the total installed base at the end of 2025 would be approximately 5.2 million.

[3] "Service robots see global growth boom," IFR press release, October 7, 2025.

[4] "Will embodied AI create robotic coworkers?," McKinsey, June 30, 2025.

inputs, follow natural-language instructions, reason and make decisions, and generate motor actions on the fly.

The shift to autonomy changes where value is created. It is becoming faster and cheaper to design powerful robots that can complete complex workflows, but hardware challenges will still limit scaling. Untethered robots will continue to depend on better power systems and batteries, while dexterous hands, sensing, and other high-performance subsystems remain bottlenecks. As robots become software-defined, the battleground is moving beyond hardware to the proprietary data, foundation models, and software that power them. The emerging robotics stack combines on-robot AI intelligence and traditional rule-based control systems while relying on cloud-based training, synthetic data, and digital twins to improve performance. Another part of the stack, teleoperation, allows direct human control of individual robots, while fleet management software helps coordinate many robots across large-scale operations.

Humanoid robots, with their uncanny antics, have captured people's attention. Today, however, the gap between what is technically demonstrated in pilots and what is commercially viable at scale remains wide. The prototypes making headlines are impressive but still far from performing reliably and consistently in the real world. The hard problems are cost, uptime, dexterity, and safe operation at scale, not just the underlying AI capability, which is advancing quickly. Few humanoid robots can yet clear McKinsey's four bridges to commercial viability: radical cost reduction, greater dexterity and mobility, sustained uptime, and safety for fenceless operation.[5]

Despite these challenges, global humanoid robot adoption is gaining ground. Shipments rose sharply from a small base to reach 18,000 units in 2025, while market revenue increased more than 500 percent year on year to $440 million.[6] Chinese firms lead shipments, with Agibot, Ubtech, and Unitree among the top producers—a position supported by China's strengths in the robotics supply chain.[7] China manufactures many key components, including most high-precision gear systems, servo motors, and battery systems. Also, the Chinese government has named humanoid robots a strategic technology priority in its 15th Five-Year Plan, with commercialization expected toward the end of the plan.[8]

The near-term and long-term stories are best kept distinct. For the next decade, growth is concentrated in what is already scaling: industrial robots in structured, task-defined environments. In the near term, most value will come from integrating robots into existing manufacturing and logistics facilities even while more new dark factories will be built. The clearest near-term opportunities for robots remain materials handling, inspection, and laboratory automation. These repetitive tasks share similarities with manufacturing: labor constraints, measurable outputs, and enough environmental structure for robots to operate reliably.

Large-scale rollouts of general-purpose robots operating in unstructured environments, such as homes, hospitality settings, and public spaces, are further away. By contrast, professional environments such as hospitals sit in a middle ground: They are more structured than homes or public spaces but still less controlled than warehouses, and they already support robots for tasks such as moving supplies and drugs. Long term, humanoid form factors will become more commonplace, even though their installed base is today a fraction of industrial robotics.

Overall, the future of robotics is best understood as the continuation of a mature market in industrial robots, with selective scaling of new general-purpose form factors. The entire sector is innovating around AI enablement, and technical capabilities are advancing quickly. But the next phase will be determined by whether robots can cross the gap from promise to performance in real-world environments. Widescale adoption of autonomous robots will depend on the economics of deployment and integration with existing systems. Robotics-as-a-service models could help reduce capital barriers, but companies will need to clearly see where robots can create value rather than just technical wizardry.

---

[5] "Humanoid robots: Crossing the chasm from concept to commercial reality," McKinsey, October 15, 2025.
[6] Lily Li, *Worldwide humanoid robotics market analysis 2026: Breakthroughs, commercialization, and emerging landscape*, IDC, January 2026.
[7] "Chinese firms dominated global humanoid robot shipments in 2025," Bloomberg, January 8, 2026.
[8] "China makes AI-powered robots core of national strategy," IFR press release, May 5, 2026.

Future of robotics

Scoring the trend

Driven by advances in AI, interest in robotics increased across searches, research publications, patents, and equity investment from 2022 to 2025. Investment climbed to $19.5 billion in 2025, its highest level yet, and job postings grew modestly year over year. Improvements in perception, manipulation, and reasoning are extending what robots can do across manufacturing, logistics, and healthcare. Technical advances and research are spurring early commercial momentum.

Equity investment, 2025

$19.5 billion

Job postings, 2024–25, % difference

+12%

Score, by vector (0 = lower; 1 = higher)

Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

Latest developments

Recent developments involving the future of robotics include the following:

- *AI is turning robots into adaptable systems, not simply better machines.* Robots are becoming able to turn camera input and natural-language instructions into actions, reducing reliance on task-specific programming. The emergence of multimodal world models—systems that can understand and predict the physical world around them—is accelerating this shift. Google DeepMind's Gemini 2.0 introduced native multimodal reasoning and agent capabilities that laid the foundation for physical AI, while Gemini Robotics and Gemini Robotics-ER extend these capabilities into embodied reasoning and VLA control.[9] Together, they show how robots can begin to reason through multistep physical tasks rather than simply follow predefined scripts. OpenVLA-OFT, a fine-tuning recipe for open-source VLA models, and the Open X-Embodiment data set that now covers more than one million real robot

---

[9] "Introducing Gemini 2.0: Our new AI model for the agentic era," blog entry by Demis Hassabis, Koray Kavukcuoglu, and Sundar Pichai, Google, December 11, 2024; "Gemini Robotics ER 2," Google blog entry, July 30, 2026.

trajectories across 22 embodiments are making it easier to adapt robot policies to new tasks and platforms.[10] Meanwhile, Physical Intelligence's generalist robotic foundation model makes the case that general-purpose robotic-AI models could be generalized across form factors with little or no extra training.[11] These advancements are creating a research-to-commercial pipeline where academic discoveries could translate into deployable robots.

- *The physical AI stack is built on AI compute.* Robots need high-performance AI for perception and reasoning, but they also need deterministic, low-latency control for motors, sensors, safety, and real-time responsiveness. New compute partnerships and specialized robotics processors are beginning to combine these layers into more integrated stacks. For instance, Texas Instruments is integrating radar devices, sensors, and motor controllers into NVIDIA Isaac Sim and using NVIDIA's Holoscan Sensor Bridge to accelerate sensor and actuator integration.[12] AMD is moving in a similar direction with its Ryzen AI Embedded P100 Series processors designed for AI-based collaborative robots, autonomous industrial machines, and mission-critical physical AI with deterministic latency and real-time control.[13] These moves show that physical AI requires more than GPU-scale intelligence; it requires timing, sensing, power, and safety systems that can operate reliably at scale.

- *Wheeled platforms and mobile manipulators are emerging as alternatives to humanoids.* The general-purpose robotics market is taking a more pragmatic view of potential form factors. Rather than force every use case into a humanoid bipedal design, both established robotics companies and emerging cobot start-ups are testing mobile manipulators and wheeled platforms that can address many of the same industrial tasks with less complexity, lower cost, and fewer balance and safety challenges. ABB is expanding mobile manipulation capabilities, for example.[14] Siemens's test of the HMND 01 wheeled Alpha humanoid at its Erlangen factory and BMW Group's pilot of the wheeled humanoid AEON at Plant Leipzig are other early field tests of the wheeled approach.[15]

- *Humanoid robots are gaining momentum, but scale is still early.* Shipment volumes rose sharply in 2025 from a small base, but many deployments are still tied to pilots and narrow industrial use cases rather than repeatable enterprise ROI. Agibot and Ubtech have both achieved scaled humanoid production, but the market is still far from widespread adoption.[16] A few companies are making an early bet on consumer humanoids for use in the home: 1X launched Neo in October 2025, positioning it as a consumer-ready humanoid robot for chores and personalized assistance.[17] However, as a precautionary measure, Neo still relies on remote human supervision for more complex tasks. The company has stated that teleoperation is a critical next step toward safe, more fully autonomous deployment. The momentum matters because more deployed humanoid robots can generate the real-world data needed to improve physical AI models. The logistics and manufacturing sectors are likely to be the first adopters of humanoids over the next decade, due to their extensive experience with robots, highly structured environments, and economically viable use cases.

- *Uptime is improving, but dexterity and manipulation remain the hardest bridges to cross.* Battery management, fleet operations,

---

[10] Moo Jin Kim, Chelsea Finn, and Percy Liang, "Fine-tuning vision-language-action models: Optimizing speed and success," arXiv, April 28, 2025, arXiv:2502.19645; Abby O'Neill et al., "Open X-embodiment: Robotic learning datasets and RT-X models," arXiv, May 14, 2025, arXiv:2310.08864.

[11] Bo Ai et al., "π0.7: A steerable generalist robotic foundation model with emergent capabilities," arXiv, April 24, 2026, arXiv:2604.15483v1.

[12] "TI accelerates the next generation of physical AI with NVIDIA," Texas Instruments press release, March 5, 2026.

[13] "AMD introduces Ryzen AI embedded processor portfolio, powering AI-driven immersive experiences in automotive, industrial and physical AI," AMD press release, January 5, 2026.

[14] "ABB expands AI-powered autonomy to Flexley Mover mobile robot," ABB press release, June 4, 2025.

[15] "Siemens and Humanoid bring physical AI to the factory floor: deploying humanoids in industrial operations with NVIDIA," Siemens press release, April 16, 2026; "First humanoid robot introduced in Plant Leipzig," BMW Group, March 9, 2026.

[16] "Agibot reaches 10,000 units as real-world demand for robots accelerates," Agibot, March 30, 2026; "Ubtech humanoid robot Walker S2 begins mass production and delivery, with orders exceeding 800 million yuan," Ubtech press release, November 17, 2025.

[17] "Neo home robot," 1X Technologies, accessed July 2, 2026.

and energy requirements are making continuous operation of robots more feasible, yet fine manipulation is still the bottleneck. The NIST Task Board benchmark, used by roboticists to measure manipulation capability, notes that in-hand manipulation and grasping are among the most difficult capabilities to achieve and measure.[18] This capability gap limits how quickly robots can move from impressive demonstrations to dependable production use.

— *Robotics as a service (RaaS) is accelerating commercial adoption.* One of the structural barriers to robotics scaling is the up-front capital cost: humanoid robots are currently priced at $150,000 to $500,000 per unit.[19] RaaS models, where operators pay per robot, per hour, or per task completed, shift the financial model toward operational expenditure instead of capital outlay. This could reduce adoption friction and make humanoids accessible to organizations that cannot justify large up-front commitments before ROI is clear. RaaS is demonstrating growing commercial maturity: Locus Robotics surpassed five billion warehouse picks across its global customer base, providing a clear proof point of scalability.[20] The RaaS model creates a structural feedback loop with aligned incentives: The deployer receives performance and reliability data, the robot operator receives utilization and failure-mode data, and both parties have direct incentives to improve uptime and task reliability.

— *Dark factories are accelerating as manufacturers integrate AI, robotics, and digital twins.* Manufacturers across automotive, electronics, semiconductors, and white-goods industries are building dark factories where robots work unsupervised. These factories integrate robotics with AI and digital-twin technologies, enabling robotic fleets to operate autonomously and maintain themselves. In China, firms including Haier, Midea, and Xiaomi are expanding highly automated manufacturing facilities—even drawing tourists to witness their robot workforces.[21] Meanwhile, South Korea is accelerating its investment in AI-powered smart factories and autonomous manufacturing systems across industrial sectors.[22] Companies in the country, including Hyundai, LG, and Samsung, are increasingly integrating AI-driven production lines and intelligent factory platforms to modernize manufacturing operations. Even so, most manufacturers will capture value not by building these greenfield showcases but by integrating robots into existing brownfield plants.

> **‘For decades, robots could repeat only what they were programmed to do. Physical AI breaks that limit: With foundation models, machines can now perceive, reason, and act in environments no one scripted in advance, and work safely alongside people. The opportunity is vast, but it will be won less by the impressive demo than by making these systems reliable enough for everyday operations.’**
>
> — Christian Jansen, partner, Hamburg

---

[18] “Robot grasping and manipulation for assembly,” National Institute of Standards and Technology, accessed July 2, 2026.

[19] “Agents, robots, and us: Skill partnerships in the age of AI,” McKinsey Global Institute, November 25, 2025.

[20] “Locus Robotics surpasses 5 billion pick milestone, accelerating global adoption of mobile warehouse automation,” Locus Robotics press release, April 24, 2025.

[21] “Introducing Xiaomi Smart Factory,” Xiaomi, accessed July 2, 2026; Zeyi Yang, “China’s electric-vehicle factories have become tourist hot spots,” *Wired*, June 23, 2025.

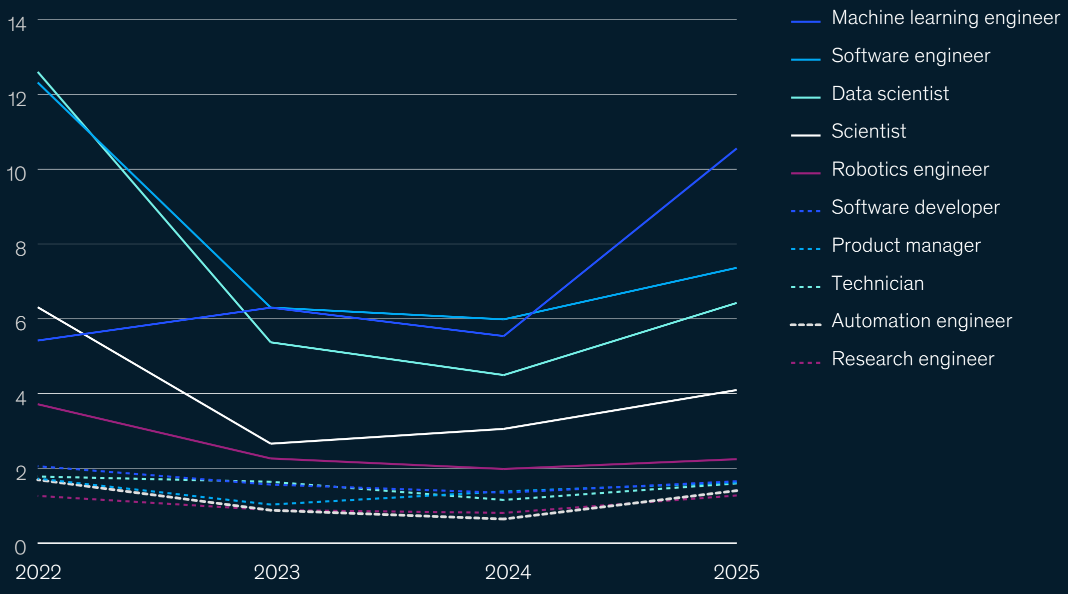
[22] “Korean government backs AI-driven autonomous factories,” *Asia Manufacturing Review*, September 18, 2025.

Future of robotics

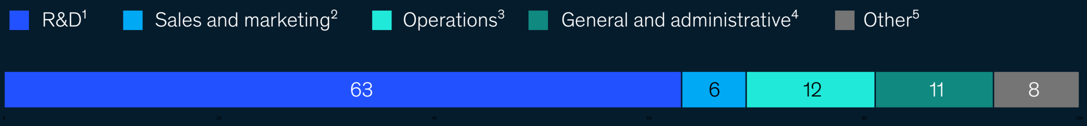
Demand

Robotics hiring has declined overall from its earlier highs, but with a noted uptick in 2025. Demand tilted toward roles in machine learning, automation, and research. Machine learning engineers are now the single largest role. Automation engineers and data scientists also saw strong growth, reflecting the rising importance of skills related to robot perception, model training, and control systems. The functional mix skews toward R&D in this frontier field.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



---

[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).

[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).

[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).

[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).

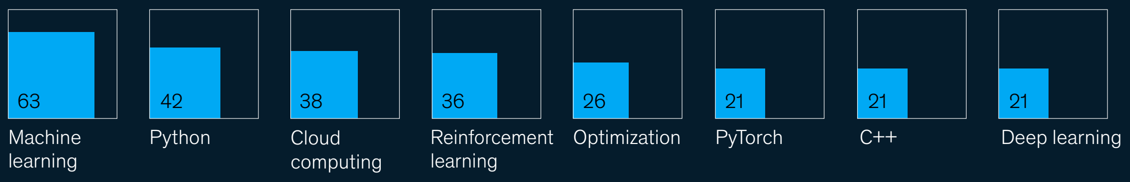
[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

Future of robotics

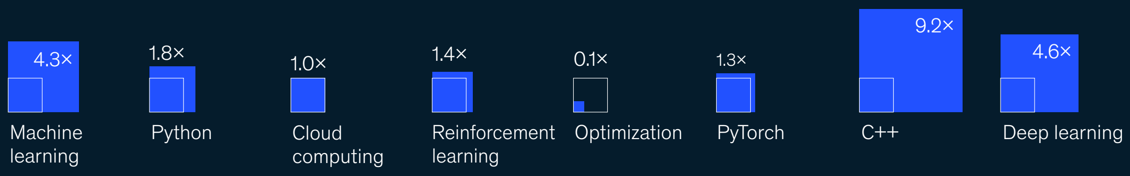
Skills availability

Robotics talent is scarcest in optimization, a key skill where demand clearly outstrips supply. This capability is central to making robotics systems work in real settings. Technical skills such as C++, deep learning, and machine learning are abundant. The harder constraint is finding people who can connect AI with hardware, controls, and deployment.

Talent required, % share of postings requiring skill ■



Talent availability, ratio of talent ■ to demand ☐



Adoption developments across the globe

Adoption score: 2—Experimentation. Organizations are testing the viability of newer robotics technologies with small-scale prototypes and targeted deployments. Industrial robots are already scaled in many manufacturing settings, and warehouse robots are increasingly common. However, AI-enabled mobile manipulation, humanoids, and broader service robots remain early.

Adoption remains uneven across geographies and sectors. China is the largest market for industrial robots and is becoming the most important early market for humanoid production and deployment. China had about two million industrial robots in operational use in 2024, and 54 percent of annual industrial robot installations worldwide were deployed there.[23] China has also placed robotics at the center of its 15th Five-Year Plan, signaling a strategic shift toward AI-powered robots as part of its modern industrial system.

The United States remains a major center for robotics start-ups, physical AI models, software, and venture-backed humanoid development, but scaling hardware remains more challenging. Europe is active in areas such as industrial automation, advanced manufacturing, and healthcare. Its strengths include high-precision mechatronics, industrial engineering, and safety frameworks. Japan has strengths in mechatronics, components, and industrial automation, supported by long-standing robotics expertise and deployments.

South Korea also is a relevant market, supported by electronics, automotive, and industrial conglomerates with robotics ambitions. Adoption in emerging markets remains more selective, but companies are making strides in robots for logistics, agriculture, mining, and infrastructure inspection, where labor shortages or safety concerns create clear use cases.

Across sectors, manufacturing and logistics continue to lead in robotics deployments.

---

[23] "Global robot demand in factories doubles over 10 years," IFR, September 25, 2025.

Healthcare, agriculture, hospitality, retail, and food service are advancing more gradually. Construction is promising but remains one of the hardest environments for robots because sites are dynamic and safety critical. Near-term construction robotics is therefore more likely to focus on specific tasks such as layout, inspection, and repetitive installation.

Underlying technologies

The technologies that power the future of robotics include the following:

Actuators, motors, and dexterous end-effectors	Hardware elements such as these turn software-based decisions into movement and manipulation.
Deterministic controls	Real-time systems manage low-latency motion, safety, and control loops.
Fleet software	These systems coordinate many robots across sites and workflows.
Physics-aware simulation	Simulation environments model physical interactions with high fidelity, enabling robots to learn, test, and validate behaviors.
Power and batteries	These energy systems determine uptime and untethered operation in robots.
Safety tools	These technologies enable safe deployment and include collision avoidance, run time monitoring, fail-safe controls, and human oversight mechanisms.
Sim-to-real tools	Training and validation environments let robot policies improve in simulation before being deployed in the real world.
Tactile sensing	These vision, force, depth, radar, and touch systems let robots understand their environments.
Teleoperation	These tools collect data and enable humans to supervise and manage robot fleets from afar.
World models	These architectures are designed to simulate physical environments, predict outcomes, and support planning for robotics and autonomous systems.
Vision-language models (VLMs) and vision-language-action (VLA) systems	These models help robots interpret scenes, reason about what is happening, and convert visual and language inputs into actions.

‘The next wave of robotics adoption will be determined less by impressive demonstrations than by measurable business outcomes. Organizations will scale general-purpose robots where they improve safety, productivity, and operational continuity in repeatable workflows. The companies that lead will be those that redesign operations with robotics.’

— Ani Kelkar, partner, Boston

Key uncertainties

The major uncertainties affecting the future of robotics include the following:

- *Real-world reliability.* Performance in warehouses is improving, but broader deployment depends on robots becoming reliable enough in semistructured and unstructured environments.
- *Sim-to-real transfer.* Robots still need faster and cheaper ways to move from simulation training to dependable real-world performance.
- *Sensor requirements.* The field still must prove whether vision-only VLA systems are sufficient, or whether tactile and other sensory inputs become essential for dependable manipulation.
- *Dexterity and manipulation.* Fine motor control remains one of the hardest bottlenecks, and faster progress will be needed to unlock broader use cases.
- *Humanoid economics.* Humanoid robots still need to achieve attractive economics relative to specialized automation, especially outside clear workflows and controlled environments.
- *Governance and trust.* More autonomous and human-adjacent robots will require safety, liability, cybersecurity, and insurance frameworks that can keep pace.

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with robotics:

- Where is ROI already clear enough to justify scaling, rather than continued piloting?
- How much future value will accrue to hardware versus models, simulation, data, compute, fleet orchestration, and integration services?
- What will it take to deploy robotics successfully in brownfield environments rather than greenfield facilities?
- What safety, liability, cybersecurity, and software-assurance controls are required as robots become more autonomous and connected?

Read more about the future of robotics

Knowledge center

Robotics and Industrial Automation

Related reading

The Next Normal: The future of robotics

Turning humanoid supply chain constraints into billion-dollar wins

The age of thinking machines: Perspectives on the future of robotics

Agents, robots, and us: Skill partnerships in the age of AI

11

Future of mobility

The technologies shaping the future of mobility are changing how people and goods move across land, air, and water. Electric, autonomous, and software-defined vehicles, along with charging infrastructure, shared mobility, drones, air taxis, and maritime systems, are making transportation more connected and automated.

The trend—and why it matters

Autonomous electric air taxis and self-navigating cargo ships are no longer science fiction; they are being tested today and could enter commercial service in a few years. Meanwhile, autonomous ridesharing has already become a reality in multiple cities. The future of transportation will leverage software and AI to make moving people and goods safer, faster, cleaner, and cheaper. The traditional vehicle could evolve from a mode of transportation into a connected, AI-enabled second space, with a cabin redesigned around work, communication, and homelike functions.

Next-generation mobility solutions are already being deployed, but adoption remains uneven across regions, transport modes, and sectors. Progress depends on continued development of new technologies, including advances in batteries and charging systems, electronics and sensors, AI-enabled perception systems, software-defined vehicle architectures, and remote-assistance tools. It also hinges on solving wider challenges. Scaling will require governments and the private sector to collaborate to update infrastructure, define new regulatory and safety frameworks, and meet consumer demand for new modes of transport. In the past decade, much of the progress in mobility focused on advancing discrete innovations such as electric vehicles (EVs), autonomous-driving software, and shared micromobility. Now, the central challenge is making these technologies work reliably inside heavily regulated transport systems such as freeways, rail networks, and aviation pathways.

Electrification remains the largest and most mature part of the trend, but its next phase is becoming more operationally complex. Global electric-car sales increased by more than 20 percent in 2025 over 2024, reaching 21 million units, and one in four cars sold worldwide was electric. China is the largest market for electric cars, capturing more than half of annual car sales for the first time in 2025. Europe saw electric-car sales grow by more than 30 percent in 2025, supported by stricter EU emissions standards, targeted national incentives, and improved model affordability, although policy support varied by country. In the United States, electric-car sales declined by 2 percent year over year in 2025, largely due to the elimination of federal tax credits after September.[1]

Beyond passenger vehicles, electrification is rapidly taking hold in commercial logistics, with sales of electric heavy-freight trucks tripling in 2025 compared with 2024, reaching more than 200,000 units globally.[2]

Autonomous mobility is also becoming more widespread, but in narrower domains than early forecasts suggested. L4 robotaxis are becoming a repeatable commercial service in many cities, and autonomous trucks are moving from supervised pilots to driverless freight operations on select routes.[3] At the same time, fully autonomous privately owned passenger vehicles are likely to scale more gradually than robotaxis or other constrained commercial use cases. Some automakers are pursuing higher levels of autonomy for consumer vehicles, but broad deployment will depend on

---

[1] *Global Energy Review 2026*, IEA, April 20, 2026.

[2] *Global Energy Review 2026*, IEA, April 20, 2026.

[3] L1, L2, L3, and L4 refer to levels of driving automation defined by SAE International, on a scale from zero to five. In L1, the vehicle can assist with a single driving task, such as steering, braking, or accelerating, while the driver remains fully responsible. In L2, the car can simultaneously assist with steering, braking, or accelerating, but the human driver must stay engaged and be ready to take control of the vehicle at any time. In L3, the vehicle can perform all driving tasks under certain conditions, but the driver must be prepared to take control when requested. In L4, the vehicle can drive itself without human intervention within specific mapped areas or operating conditions.

safety validation, regulatory approval, liability models, and consumer trust. For the mass market, advanced driver assistance features are likely to remain the primary near-term pathway. A 2025 McKinsey survey of autonomous-vehicle experts found that L4 robotaxis are expected to be the first commercial L4 application at scale, while 49 percent of experts expect the mass market for privately owned vehicles to center on L2 functions through 2035.[4]

Novel mobility solutions are now moving toward selective deployment. Commercial drone delivery, for example, is expanding in the retail and healthcare sectors, where speed, small payloads, and access matter. Electric air taxis are approaching early passenger service, but infrastructure, air traffic integration, and battery performance remain constraints. Micromobility solutions such as e-scooters and e-bikes are stabilizing after earlier oversupply and are increasingly being integrated into city transport systems. Maritime and industrial mobility are also advancing through the electrification and automation of boats and ferries, as well as through port system integration.

Mobility is also becoming more software defined. Vehicle systems increasingly integrate sensors, AI systems, over-the-air updates, fleet management, payments, and customer applications. At the same time, electrification is linking mobility more closely to energy systems—creating interdependencies that could limit the scaling of mobility solutions. Uptake of EVs, fleets, and charging networks will depend on grid capacity, interconnection, and charging economics. Scaling will also depend on battery supply chains, including access to critical minerals, refining capacity, and recycling.

Overall, the future of mobility is best understood as selective scaling. Technologies are advancing, but the pace of adoption depends not only on how quickly they reach commercial viability but also on infrastructure readiness, regulatory approval, safety validation, and user trust.

> **‘Mobility is becoming a software-defined system where value comes as much from data, software, and fleet operations as from the vehicle itself. The winners will be those that integrate electrification, autonomy, and digital infrastructure into a seamless customer and operating experience.’**
>
> — Andreas Breiter, partner, Bay Area

---

[4] McKinsey conducted a survey of leaders in the autonomous-driving industry in January 2025. Respondents were 91 decision-makers (43 from the European Union, 35 from North America, and 13 from Asia) and included leaders from start-ups and incumbents in automotive, transportation, and software, as well as supporting institutions such as universities and mapping and navigation companies. For more, see “Where to next? Insights from autonomous-vehicle experts,” McKinsey, January 6, 2026.

CUTTING-EDGE ENGINEERING

Future of mobility

Scoring the trend

Research publications for the future of mobility have fallen sharply since 2022, the trend's largest shift, and equity investment has eased as well, reflecting a sector moving from broad experimentation toward more selective deployment in use cases such as robotaxis and autonomous freight. Patent activity, by contrast, has increased, and search and news interest have edged higher, pointing to continued, but more focused, innovation as the industry consolidates around proven applications.

Equity investment, 2025
$63.5 billion

Job postings, 2024–25, % difference
–11%

Score, by vector (0 = lower; 1 = higher)

Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

Latest developments

Recent developments involving the future of mobility include the following:

— *Robotaxis are moving from demonstration to repeatable service in some cities.* Autonomous ridesharing is becoming commonplace in select cities, but its commercial progress remains highly localized. Waymo reported more than 15 million trips in 2025 and exceeded 20 million lifetime trips.[5] Baidu's Apollo Go reported 3.4 million fully driverless operational rides in the fourth quarter of 2025 and more than 20 million cumulative public rides by February 2026, with most of these rides in China. Baidu also announced select early deployments in the Middle East and pilots in Europe in early 2026.[6] These examples show that robotaxis can become repeatable urban services, but deployments remain operationally complex and dependent on regulatory approval, fleet execution, and public trust. Regional patterns

---

[5] "Delivering more for our riders in a year of incredible growth," Waymo, December 10, 2025.
[6] "Baidu announces fourth quarter and fiscal year 2025 results," Baidu press release, February 26, 2026.

are also emerging: The United States has seen significant city-level robotaxi deployments; China is building a dense local ecosystem of operators and routes; and Europe has strong pilot activity as more players announce city trials. An important regulatory milestone came in June 2026, when the United Nations Economic Commission for Europe (UNECE) adopted the first global regulatory framework for automated driving systems. The framework provides China, the European Union, the United States, and other major markets with a shared technical foundation for developing approval processes.[7]

— *Electric-vehicle adoption is growing, but strategies are becoming more regional and affordability focused.* Electric vehicles continue to gain share globally, with electric-car sales reaching 21 million in 2025, or about one-quarter of all new-car sales.[8] Growth remains uneven, however, with adoption shaped by local prices, charging access, consumer preferences, and, particularly in Europe, regulatory mandates. The European Union's fleetwide emissions standards effectively require automakers to increase EV sales or face penalties. The bloc's planned requirements for new cars to meet a 90 percent CO_2 reduction target after 2035 adds further pressure for carmakers to shift production toward electric models.[9] China's low battery prices continue, with McKinsey Battery Insights research finding that some Chinese manufacturers have reached lithium iron phosphate (LFP) pack costs of approximately €64 per kilowatt-hour (kWh) and about €82 per kWh for nickel manganese cobalt (NMC) packs. Leveraging simplified pack architecture, lower-cost LFP chemistry, and optimized component costs, Chinese OEMs achieve battery production costs that are up to 40 percent lower than global peers.[10] In China, price parity between EVs and internal-combustion cars has largely been achieved, driven by not only cheaper batteries but also lower costs across the entire vehicle.[11] Automakers in areas with high battery costs are adjusting portfolios to include plug-in hybrids, extended-range EVs, and lower-cost battery chemistries such as LFP.

— *The electrification bottleneck is shifting from vehicle supply to energy access.* There are plenty of EVs available, but adoption at scale is increasingly constrained by access to reliable charging infrastructure and electricity. While global public charging networks surpassed seven million points by the end of 2025, large-scale charging is becoming a multiyear project development challenge.[12] Lead times for the large transformers and switchgear required for high-power depots have stretched to more than two years in some major markets.[13] In markets where public charging infrastructure is still a bottleneck, some vehicle makers are focusing less on cars and leapfrogging to freight vehicles, where routes are more predictable and depot charging can be planned. As a result, electric heavy-freight truck sales tripled in 2025 to more than 200,000 units globally, with growth concentrated in China.[14] However, scaling of electric heavy-freight vehicles will depend on grid capacity, equipment availability, and the pace at which high-power standards, such as

---

[7] "UNECE adopts first-ever global rules allowing fully autonomous vehicles," United Nations Economic Commission for Europe press release, June 24, 2026.

[8] *Global EV Outlook 2026: Growing sales amid an energy crisis*, IEA, May 20, 2026.

[9] Marta Pacheco, "EU carmakers to comply with 90 percent emissions reduction by 2035 as full combustion engine ban scrapped," Euronews, December 16, 2025.

[10] Clemens Cepnik, Martin Linder, Andreas Haunreiter, and Luca Buscaglione, "The future of affordable EVs: Breakthroughs in battery pack costs," McKinsey, January 2, 2026.

[11] "Pathways to global EV cost-competitiveness," in *What next for the global car industry? An Energy Technology Perspectives special report*, IEA, November 18, 2025.

[12] "Charging outlook," in *Global EV Outlook 2026: Growing sales amid an energy crisis*, IEA, May 20, 2026.

[13] Benjamin Boucher, Devin Thomas, and Michael Mendrek-Laske, "Mind the gap: Tackling supply-chain challenges in the electric T&D sector," Wood Mackenzie, October 15, 2025.

[14] *Global EV Outlook 2026: Growing sales amid an energy crisis*, IEA, May 20, 2026.

megawatt charging systems, move from early deployment to broader availability.

- *Autonomous trucking is entering early deployment.* Driverless trucking is finding an early commercial foothold along constrained routes and in repeatable logistics settings, rather than nationwide autonomy. Aurora began commercial driverless trucking operations on public roads between Dallas and Houston.[15] Meanwhile, Kodiak and Atlas Energy Solutions expanded driverless operations in the Permian Basin.[16] These examples suggest that freight autonomy is more likely to scale first in fixed corridors, semistructured industrial environments, and hub-to-hub networks, where safety and operating economics are easier to establish.

- *Software is increasingly important for creating automotive value.* Automotive vehicles are being built around centralized computing, connected architectures, and over-the-air updates, allowing automakers to improve functionality and offer software-enabled features. McKinsey analysis estimates that the global automotive software and electronics market could grow at a 5.9 percent compound annual growth rate and reach $548 billion by 2035, compared with roughly 1 percent annual growth for the overall vehicle market.[17]

- *Drone delivery is moving from pilot programs toward retail expansion.* Autonomous delivery via drones is beginning to move from proof-of-flight demonstrations toward more structured operations. This shift is supported by regulatory progress. In 2025, the Federal Aviation Administration proposed a rule to enable routine beyond-visual-line-of-sight (BVLOS) drone operations, allowing drones to fly beyond the operator's direct line of sight without requiring case-by-case waivers. If this becomes legally binding, it will remove a key barrier to scaling commercial applications such as delivery, infrastructure inspection, and emergency response in the United States.[18] Walmart's expansion of drone deliveries to 66 stores across four states, following more than one million completed drone deliveries, illustrates how the business case is increasingly concentrating on small-payload delivery.[19] With a delivery weight limit of 2.8 pounds, Walmart's drone delivery operating model, like those of other retailers, is constrained by drones' weight capacity and flight ranges. Amazon's Prime Air targets eligible packages weighing five pounds or less, and Keeta Drone, in China, focuses on deliveries within three kilometers among 70 set routes.[20] In practice, drone delivery is still concentrated in suburban or office-dense areas where a central pickup station is easier to manage than true door-to-door residential delivery.

- *Ground-based autonomous delivery is advancing selectively.* Ground delivery by mobile robots is also gaining traction in specific use cases, particularly short-distance food, grocery, and campus delivery. Serve Robotics built more than 2,000 delivery robots by the end of 2025.[21] Meanwhile, Starship Technologies announced in late 2025 that it had completed more than nine million deliveries via a fleet of over 2,700 robots across more than 270 locations and

---

[15] "Aurora begins commercial driverless trucking in Texas, ushering in a new era of freight," Aurora Innovation press release, May 1, 2025.

[16] "Kodiak delivers customer-owned autonomous RoboTrucks to Atlas Energy Solutions, completes 100 loads of proppant with first-ever driverless commercial semi-truck service," Atlas Energy Solutions press release, January 24, 2025.

[17] "The automotive software and electronics market through 2035," McKinsey, January 6, 2026; figures have been revised upward since publication.

[18] "Beyond visual line of sight (BVLOS)," Federal Aviation Administration, August 6, 2025.

[19] "Walmart celebrates 1 million drone deliveries, marking a major milestone in customer convenience," Walmart, May 29, 2026.

[20] Tyler Greenawalt, "Amazon's drones deliver items in 60 minutes or less—here's how we simplified the process," Amazon News, May 20, 2025.

[21] "Serve Robotics builds 2,000 autonomous delivery robots, creating largest sidewalk delivery fleet in the U.S.," Serve Robotics press release, December 12, 2025.

had partnered with Uber Eats to expand autonomous sidewalk delivery across Europe and the United States. The company expects to have 12,000 robots in service by the end of 2027.[22] Ground-based autonomous delivery is also gaining traction in China. Neolix said it has surpassed 10,000 cumulative L4 autonomous-vehicle deliveries in more than 300 cities and is seeing commercial use expand across urban delivery scenarios such as express freight, fresh food, cold chain, and supermarkets.[23]

— *Electric air taxis are moving toward launch readiness, but scaled service remains unproven.* Urban air taxis, or electric vertical takeoff and landing (eVTOL) vehicles, are progressing from prototypes toward certification, operations planning, and infrastructure development. Commercial flights could happen in a few years. Joby completed more than 850 flights across its electric air taxi fleet in 2025 and positioned itself for Federal Aviation Administration certification flight testing.[24] Meanwhile, other players are testing eVTOL aircraft in local operating conditions and planning early launch markets. These developments are important signs of operational maturity, but passenger eVTOL services still face challenges in addition to certification, such as air traffic integration, cost, noise, and public acceptance before they can scale.

— *Shared micromobility is becoming more integrated with urban transportation systems.* After a period of market consolidation, the industry is entering a new phase focused on steady, profitable growth as operators work to improve cost management and unit economics. Shared bikes and scooters are now increasingly being treated as part of the urban mobility mix rather than as a stand-alone novelty. Micromobility company Lime, for example, surpassed one billion lifetime rides in 2025.[25] Certain cities are seeing wide uptake of shared-mobility solutions. For example, Seattle's bike and scooter share program reached a record ten million rides in 2025, with trips up more than 60 percent from the prior year as the city added protected bike lanes and expanded parking infrastructure.[26] Shared-micromobility use has shifted toward utilitarian trips, with commuting now a key purpose for rides, even as leisure, errands, and appointments remain common across rider age groups. These examples suggest that micromobility can scale when operators improve unit economics and cities invest in safety, affordability, and network integration—often through close collaboration between the two, particularly on regulation and pricing structures.

— *Maritime mobility is an early test bed for electrification and autonomy.* Autonomous electric boats are being deployed on predictable routes where infrastructure can be integrated with public transit and port systems. Stockholm expanded service for Candela's electric hydrofoil ferry following a successful pilot, demonstrating how electric propulsion and hydrofoil technology can reduce energy consumption and enhance the passenger experience in urban water transit.[27] In parallel, Singapore's Maritime and Port Authority and NYK Group collaborated to advance trials of port-to-port maritime autonomous surface ships.[28] These developments point to an early operational phase for autonomous maritime systems, but the segment remains focused on route-specific deployments.

---

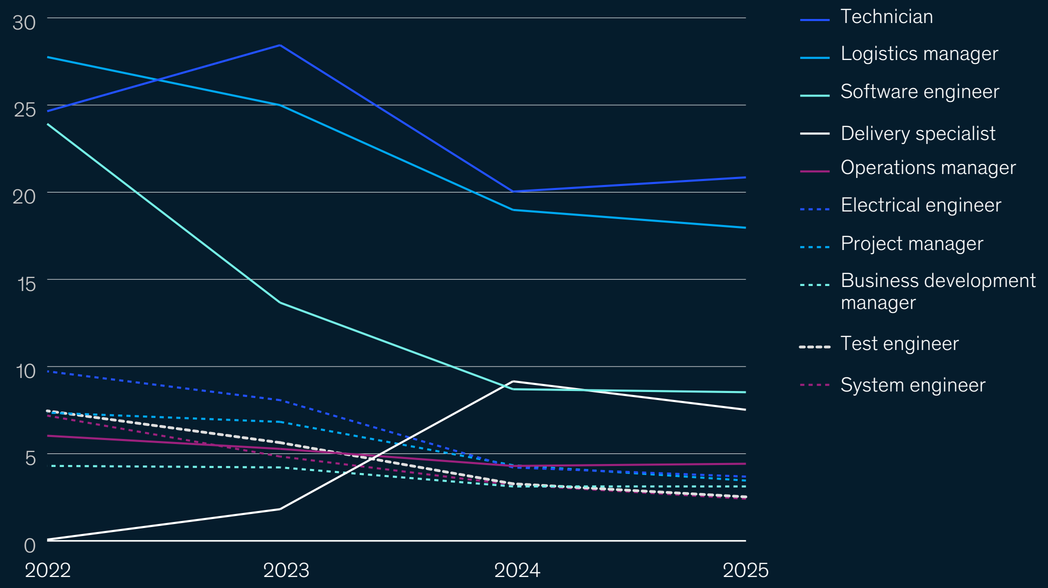
[22] "Starship Technologies and Uber Eats launch autonomous delivery partnership," Starship Technologies press release, November 20, 2025.
[23] "Neolix raises over $600M in Series D funding, setting new record in China's autonomous driving sector," Neolix press release, June 28, 2025.
[24] "Joby caps year of flight, demonstrating global commercial readiness and operational momentum towards FAA certification," Joby Aviation press release, December 15, 2025.
[25] *The Lime Times*, "Lime reaches 1 billion rides worldwide," blog entry, Lime, October 17, 2025.
[26] *SDOT Blog*, "10 million rides: Seattle bike & scooter share sets new record in 2025," blog entry by Amy Abdelsayed, Seattle Department of Transportation, December 23, 2025.
[27] Carla Giaume, "Stockholm expands electric hydrofoil ferry service after successful pilot," EU Urban Mobility Observatory, May 6, 2025.
[28] "MPA and NYK Group advance collaborative efforts on maritime autonomous surface ship trials," NYK press release, June 5, 2025.

Future of mobility

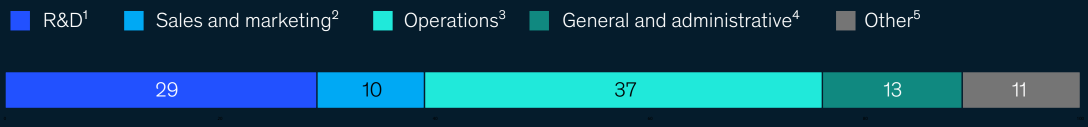
Demand

Mobility job postings remain well below their 2022 peak and continued to decline from 2024 to 2025. The steepest pullbacks have come in engineering roles such as electrical, software, systems, and test engineers as the industry moves from build-out toward operations. Roles for technicians, now the single largest role, and operations managers have held up better comparatively, pointing to a market increasingly focused on running and maintaining mobility systems at scale. At a functional level, over a third of mobility job postings are in operations, the highest share of any trend.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).

[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).

[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).

[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).

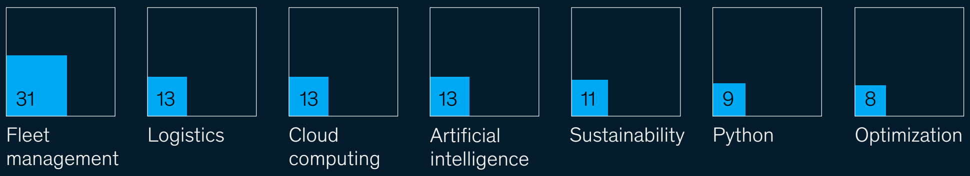
[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

Future of mobility

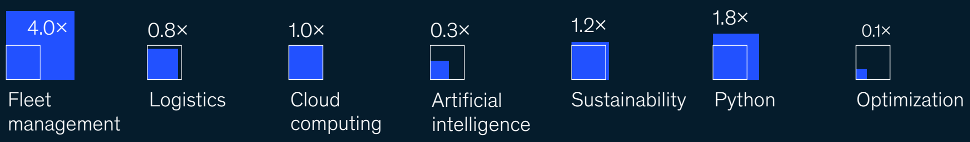
Skills availability

Fleet management, the most in-demand skill, is well supplied, and the pool of professionals with Python expertise exceeds the number of relevant openings. However, in artificial intelligence and optimization, supply falls short of open roles. These gaps could constrain growth in autonomous-vehicle expansion.

Talent required, % share of postings requiring skill ■



Talent availability, ratio of talent ■ to demand ☐



Adoption developments across the globe

Adoption score: 3—Piloting. Organizations are deploying mobility technologies in several specific business use cases but are mostly launching pilots to test feasibility. EVs and micromobility are moving toward maturity, while autonomous vehicles, drones, eVTOL aircraft, and maritime autonomy remain more selective and use case specific.

Adoption varies significantly across regions. China remains a top market for EVs, battery production, public charging, and robotaxis. In 2025, electric cars accounted for more than half of annual car sales in China. The country holds the most public charging infrastructure, representing more than 65 percent of public charging points globally at the end of 2025. China's public charging infrastructure grew from nearly 3.4 million charge points at the end of 2024 to more than 4.7 million at the end of 2025, accounting for over 75 percent of global growth.[29] China excels not only in technology development but also in manufacturing scale, cost competition, and integrated industrial policy.

Europe's mobility progress continues to be shaped by regulation, sustainability targets, and urban policy. Electric-car sales in the European Union increased 30 percent in 2025, while Europe's public charging network expanded to roughly 1.3 million charging points in 2025.[30] Micromobility is increasingly integrated into urban transportation, and European cities are active in shared mobility, public-transport integration, and maritime electrification.

The United States is making strides in robotaxis, autonomous trucking, drone delivery, software-defined vehicles, and mobility start-ups. However, EV adoption is more regionally uneven and sensitive to policy, consumer economics, charging availability, and political context. US electric-car sales declined 2 percent year over year in 2025, while the public

---

[29] *Global EV Outlook 2026: Growing sales amid an energy crisis*, IEA, May 20, 2026.
[30] *Global EV Outlook 2026: Growing sales amid an energy crisis*, IEA, May 20, 2026.

charging stock was roughly 235,000 points at the end of 2025.[31]

The Middle East is becoming more visible in advanced air mobility, smart-city mobility, and digital-transport infrastructure. The region's role is less about broad-based adoption today and more about serving as an early launch market for regulated advanced mobility services, supported by urban development agendas, infrastructure investment, and government-backed innovation programs.

India and Southeast Asia are important markets for lower-cost electrification, especially for two- and three-wheelers, buses, and high-utilization commercial fleets. In 2025, India's annual EV sales reached 2.3 million units, while electric-car sales increased more than 75 percent year over year. Electric-car sales in Southeast Asia more than doubled from 2024 to 2025, with Indonesia growing 125 percent.[32] In these markets, affordability, density, and high vehicle utilization could support faster adoption of lightweight, shared electric mobility models rather than private electric cars.

Africa and parts of Latin America remain earlier in mobility adoption, but select markets are making advances in electrifying buses, motorcycles, and fleets. Electric-car sales in Latin America and the Caribbean grew by about 70 percent in 2025 compared with 2024, reaching nearly 350,000 units.[33] Constraints include a lack of charging infrastructure, affordability concerns, power reliability, and vehicle availability.

Across regions, adoption is strongest when mobility technology solves a clear operating problem, such as reducing fuel costs, serving underserved areas, or reducing emissions. Adoption is slower when the business case depends primarily on novelty or on infrastructure that does not yet exist.

> **'The next challenge is not proving that autonomous and electric mobility work. It's building the charging networks, regulatory frameworks, and public trust needed to scale them. Commercial success will depend not just on ecosystem coordination but also on advances in vehicle technology.'**
>
> — Ruth Heuss, senior partner, Berlin

---

[31] *Global EV Outlook 2026: Growing sales amid an energy crisis*, IEA, May 20, 2026.
[32] *Global EV Outlook 2026: Growing sales amid an energy crisis*, IEA, May 20, 2026.
[33] *Global EV Outlook 2026: Growing sales amid an energy crisis*, IEA, May 20, 2026.

Underlying technologies

The technologies that power the future of mobility include the following:

Advanced air mobility	These technologies include electric aircraft designed for short-range passenger and cargo flights, and the infrastructure needed to support them.
AI and simulation	AI and simulation tools train, test, and validate autonomous systems before and during real-world deployment.
Autonomous-driving software	This software layer enables vehicles to perceive their environment, make decisions, and drive with limited or no human input.
Batteries and electric powertrains	Energy storage and propulsion systems replace internal combustion engines across passenger, freight, and shared vehicles.
Charging infrastructure	This infrastructure includes the physical and digital systems that deliver electricity to vehicles, from public chargers and depot stations to battery-swapping networks.
Drone and autonomous delivery	These technologies enable unmanned aerial and ground-based vehicles to transport goods.
Fleet management and mobility platforms	These platforms include operations software that helps companies run shared, autonomous, and electric-mobility services at scale.
Grid integration	Grid integration systems manage the connection between electric vehicles, charging networks, and the power grid.
Maritime mobility	Electric propulsion, autonomous navigation, and port systems support lower-emission movement across waterways.
Materials and circularity	Lightweight and recyclable materials, battery recycling, and manufacturing processes reduce vehicle weight and life cycle emissions.
Sensors and positioning	Hardware such as lidar, radar, and cameras gives vehicles the ability to identify objects and environments, locate themselves, and navigate safely.
Software-defined vehicle architectures	Computing and connectivity platforms allow vehicles to be updated and improved after they leave the factory.

Key uncertainties

The major uncertainties affecting the future of mobility include the following:

- *EV infrastructure and supply chains.* Electric-vehicle adoption will depend on how incentives, battery costs, and charging availability evolve. Scaling will also depend on whether charging infrastructure, grid capacity, and battery supply chains can keep pace with electrification needs, especially for commercial fleets and heavy-duty transport.

- *Affordability and regional competitiveness.* Lower-cost EV platforms could broaden adoption, but it remains uncertain how quickly automakers can deliver affordable vehicles across markets with very different trade policies, consumer preferences, and charging infrastructures.

- *Public acceptance of autonomous vehicles.* Autonomous mobility will need to demonstrate safety and reliability to gain widespread public trust. Robotaxis must prove they can scale outside dense urban markets, while autonomous trucking must move beyond constrained routes and private industrial settings.

- *Scalability of emerging delivery models.* When it comes to delivery, drones and ground-based robots have already proven their utility in early deployments, but their broader commercial viability remains uncertain. Scaling will depend on regulation, operating economics, battery performance, and community acceptance.
- *Mobility-system software risk.* As vehicles become more connected and software-defined, they become more exposed to software failures and cyberattacks. Since a single over-the-air update can reach an entire fleet, one vulnerability could have consequences at a scale the automotive industry has not had to manage before.

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with mobility technologies:

- Which mobility technologies are ready to scale now, and which should remain in pilots until safety, economics, infrastructure, and customer acceptance improve?
- Where should autonomy be prioritized first across robotaxis, trucking, industrial sites, ports, public transit, and advanced driver-assistance systems?
- How should companies and governments coordinate the infrastructure needed for electrified, autonomous, and multimodal mobility?
- Which players will capture value as mobility becomes more software-defined: automakers, battery makers, semiconductor companies, software providers, fleet operators, charging networks, or infrastructure owners?
- How should companies manage trust, safety, cybersecurity, and operational reliability as mobility systems become more autonomous?

Read more about the future of mobility

Knowledge center

Our Insights on Future Mobility

Related reading

How consumers are reshaping the future of mobility

Reinventing autonomous driving in the age of generative AI

The automotive software and electronics market through 2035

The future of affordable EVs: Breakthroughs in battery pack costs

The future of air mobility: Electric aircraft and flying taxis

12

Future of space technologies

Space technologies span four connected layers: launch systems that send payloads into orbit, spacecraft and infrastructure that operate in space, Earth-based ground networks, and services such as connectivity, observation, satellite servicing, and orbital computing.

The trend—and why it matters

People may not yet be colonizing Mars, but the future of space technologies has made bold leaps in the past year. The sector is increasingly moving away from stand-alone spacecraft to build an integrated infrastructure layer that powers the terrestrial economy. A joint research report by McKinsey and the World Economic Forum found the global space economy could grow from $630 billion in 2023 to $1.8 trillion by 2035, or about 9 percent annually.[1] That definition includes both core space industries—such as launch systems, satellites, communications, Earth observation, and defense—and adjacent industries enabled by space-based services, including ridesharing and delivery tracking.

This broader space-enabled layer is expected to grow quickly. Companies in the transportation, communications, food and beverage, and retail sectors are increasingly leveraging data and services originating in space to improve operational performance here on Earth. What's more, emerging super-heavy launch systems could enable more mass-intensive use cases, potentially including orbital computing infrastructure, as the space economy continues to expand.

We see several shifts taking place in the future of space technologies. For one, launch economics are becoming more viable. Reusable rockets, more efficient manufacturing, and higher launch frequency have reduced the cost of launching heavy payloads into low Earth orbit (LEO) around 95 percent over the past four decades, to a low of about $1,500 per kilogram, from roughly $65,000 per kilogram.[2] Further progress in super-heavy launch could make it more practical to deploy heavier payloads, including larger satellites and orbital infrastructure, although the commercial timeline is uncertain. The decrease in launch costs has made space-based assets increasingly affordable across communications, sensing, navigation, Earth observation, defense, and the adjacent industries that leverage satellite data for logistics, agriculture, and climate monitoring.

Satellite connectivity solutions that bring data from space to Earth also are rapidly evolving. LEO satellite broadband has scaled, extending space-based internet access to homes, businesses, ships, aircraft, and remote sites through dedicated terminals. Direct-to-device (D2D) connectivity is a more nascent category. Unlike satellite broadband services, which require a dedicated terminal, D2D aims to connect satellites directly to standard mobile phones or Internet of Things (IoT) devices. Today, the market is split across several use cases: established satellite communications and emergency-response services, fast-growing maritime and aviation connectivity delivered through dedicated terminals, and nascent smartphone D2D services that aim to integrate more directly with terrestrial mobile networks.

Earth observation is another growth area. The value is moving from supplying images to supplying operational intelligence. Satellite data is increasingly being combined with onboard AI, ground-based data, weather models, customer workflows, and decision support systems to provide users with custom insights. This is particularly important in sectors such as defense, climate monitoring, agriculture, insurance, energy, and disaster response, where the value lies less in the raw imagery and more in the intelligence derived from those images.

Another area attracting interest is orbital compute, although commercial services may still be years

---

[1] "Space: The $1.8 trillion opportunity for global economic growth," McKinsey, April 8, 2024.

[2] Thomas G. Roberts, "Space launch to low Earth orbit: How much does it cost?," Center for Strategic and International Studies, September 1, 2022.

away. This should be distinguished from space-based data processing, which analyzes data in orbit before sending selected outputs back to Earth. Orbital compute would place larger-scale computing infrastructure directly in orbit. The concept is drawing attention because AI workloads are putting pressure on terrestrial power and data center capacity, but the business case remains technically and economically unproven. Near-term use cases of orbital compute are more likely to focus on "space for space" applications—such as satellite operations and onboard inference—than on direct competition with Earth-based cloud infrastructure.

Alongside these technical developments, commercial space is showing signs of greater financial maturity. Private investment remains strong but selective, with capital concentrating around launch, satellite connectivity, geospatial intelligence, defense-linked services, and orbital infrastructure. Public-market activity is also reopening selectively, with activity concentrated around more proven platforms rather than across the sector as a whole. SpaceX's IPO in June 2026, which raised $85.7 billion, and HawkEye 360's May 2026 IPO illustrate continued investor interest in geospatial and defense-linked space services.[3] Although these examples suggest that public markets can absorb scaled space infrastructure companies, they do not yet prove broad public-market appetite for every space category.

Realizing opportunity in the space sector will depend on three enabling conditions. First, governance will need to keep pace with the multiplication of space players, particularly around spectrum access, orbital congestion, debris mitigation, and cybersecurity. Second, the economics will need to pencil out with lower launch costs, higher asset utilization, and customer willingness to pay for space-enabled services. Third, space systems will need to deliver the safety and reliability required for downstream adoption, especially in sectors where connectivity, sensing, and decision support services are becoming operationally critical.[4]

Overall, the future of space technologies is best understood through the lens of changing economics. Lower launch costs and more capable satellites are enabling a wider range of space-based services. Connectivity, geospatial intelligence, and defense-linked applications remain the clearest near-term markets, while super-heavy launch and orbital compute could expand what's possible over time. The central challenge is turning these capabilities into reliable services that customers on Earth are willing to purchase.

'Space is becoming an infrastructure layer for the terrestrial economy. The opportunity is defined not by getting into orbit but by delivering competitive connectivity, intelligence, and services. The winners will be those that translate space capabilities into operational value on Earth.'

— Ryan Brukardt, senior partner, Miami

---

[3] "Space Exploration Technologies Corp. announces closing of initial public offering, including full exercise of underwriters' option to purchase additional shares," SpaceX press release, June 15, 2026; "HawkEye 360 announces pricing of initial public offering," HawkEye 360 press release, May 6, 2026.

[4] Ryan Brukardt, Jesse Klempner, Bob Sternfels, and Brooke Stokes, "Space: The missing element of your strategy," McKinsey, March 27, 2023.

CUTTING-EDGE ENGINEERING

Future of space technologies

Scoring the trend

Interest in space technologies has picked up since 2022, with gains in news coverage, searches, and equity investment, even as research publications slipped. Equity investment grew by roughly 70 percent between 2022 and 2025, and investor interest accelerated sharply in 2026, with SpaceX raising $85.7 billion in its June IPO, including an overallotment option.[1] These moves signal growing public-market confidence in space infrastructure.

Equity investment, 2025
$15.7 billion

Job postings, 2024–25, % difference
–16%

Score, by vector (0 = lower; 1 = higher)

Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

[1] "Space Exploration Technologies Corp. announces closing of initial public offering, including full exercise of underwriters' option to purchase additional shares," SpaceX press release, June 15, 2026.

Latest developments

Recent developments involving the future of space technologies include the following:

— *Launch economics are changing what can be built in orbit.* Reusable and super-heavy launch systems are progressing through major milestones that could make heavier satellites, communications infrastructure, and future orbital systems more practical. In the past year, SpaceX, for example, flew its Super Heavy booster for a second time, a key step toward achieving the full reusability that underpins the system's economics.[5] Meanwhile, Blue Origin announced that it is building a super-heavy version of its New Glenn booster called 9x4.[6] These developments demonstrate broader progress toward reusable heavy-lift launch systems. However, a recent New Glenn test explosion highlights the uncertainty around launch cadence, reliability, timelines, and commercialization.[7] Thus, the feasibility of routine low-cost, super-heavy

---

[5] "Starship's ninth flight test," SpaceX, May 27, 2025.

[6] "New Glenn update: Upgraded engines and subcooled components drive enhanced performance," Blue Origin press release, November 20, 2025.

[7] Dave Limp, "GS1-3 hotfire updates," Blue Origin press release, June 1, 2026.

launches remains unproven, but recent progress points to that potential.

- *Satellite connectivity is expanding across consumer and enterprise use cases.* LEO broadband constellations are moving from niche to mainstream commercial service for homes, businesses, maritime, aviation, and remote sites, while D2D connectivity is a more nascent category connecting satellites directly to standard smartphones. LEO broadband services are already operating at scale through constellations such as Starlink, which has thousands of satellites in orbit, and OneWeb.[8] Amazon Leo, formerly Project Kuiper, began full-scale deployment in 2025 and has since launched more than 300 satellites.[9] Meanwhile, AST SpaceMobile's BlueBird program points to the potential expansion of D2D from emergency messaging toward broader mobile connectivity use cases, although the loss of BlueBird 7 following a New Glenn launch anomaly underscores the deployment risks that remain.[10]

- *Earth observation is becoming more integrated into operational decision-making.* Earth observation has long converted satellite imagery and sensor data into intelligence, but the value is increasingly tied to how those insights can be integrated into customer workflows. Advances in hyperspectral sensing, AI-enabled analytics, cloud-based platforms, and data fusion are making Earth observation outputs more actionable for use cases such as emissions monitoring, climate resilience, agriculture, insurance, energy, and disaster response. In emissions monitoring, GHGSat's methane data was selected for access through NASA's Commercial Satellite Data Acquisition program. Carbon Mapper announced plans to expand its Tanager constellation with next-generation detection technology designed to improve sensitivity to methane super emitters and extend monitoring across agriculture, waste, and energy sectors.[11] These announcements illustrate a shift from collecting data about Earth to helping governments and companies analyze it to decide where to act.

- *Defense applications remain a major backbone of the space economy.* As space systems become more important to communications, sensing, navigation, and national security, defense demand continues to shape investment and deployment. In the United States, for example, the Proliferated Warfighter Space Architecture's first Tranche 1 launch shows how commercial-style data transport satellites are being used for national security. For instance, these satellites can provide US defense users with operational capabilities from space, including integrated sensing and communications.[12]

- *Orbital compute is moving into early experimentation.* Space-based computing remains nascent, but recent projects are making the idea more concrete. Axiom Space and Spacebilt announced plans for an orbital data center node aboard the International Space Station.[13] Starcloud, a member of the NVIDIA Inception program, is using NVIDIA accelerated computing platforms for next-generation space missions.[14] These are early proof points for processing data closer to where it is generated in space but not yet evidence that orbital data centers can effectively and efficiently compete with terrestrial cloud infrastructure. Starcloud CEO Philip Johnston reiterated this reality in a recent interview, saying that near-term use cases of orbital compute would focus mostly on providing inference onboard spacecraft.

---

[8] Tereza Pultarova, "Starlink satellites: Facts, tracking and impact on astronomy," Space.com, updated July 30, 2026.

[9] "Amazon Leo mission updates: 375+ satellites now in orbit after successful Atlas V launch," Amazon News, updated July 2, 2026.

[10] Mike Wall, "Blue Origin's huge New Glenn rocket grounded after launch mishap," Space.com, April 22, 2026.

[11] "Carbon Mapper unveils plan for next-generation methane detection technology to cut climate's strongest pollutant," PR Newswire, April 30, 2026.

[12] "Space Systems Command, Space Development Agency complete successful launch of first Tranche 1 satellites," Space Systems Command press release, US Space Force, September 10, 2025.

[13] "Axiom Space, Spacebilt announce orbital data center node aboard International Space Station," Axiom Space press release, September 16, 2025.

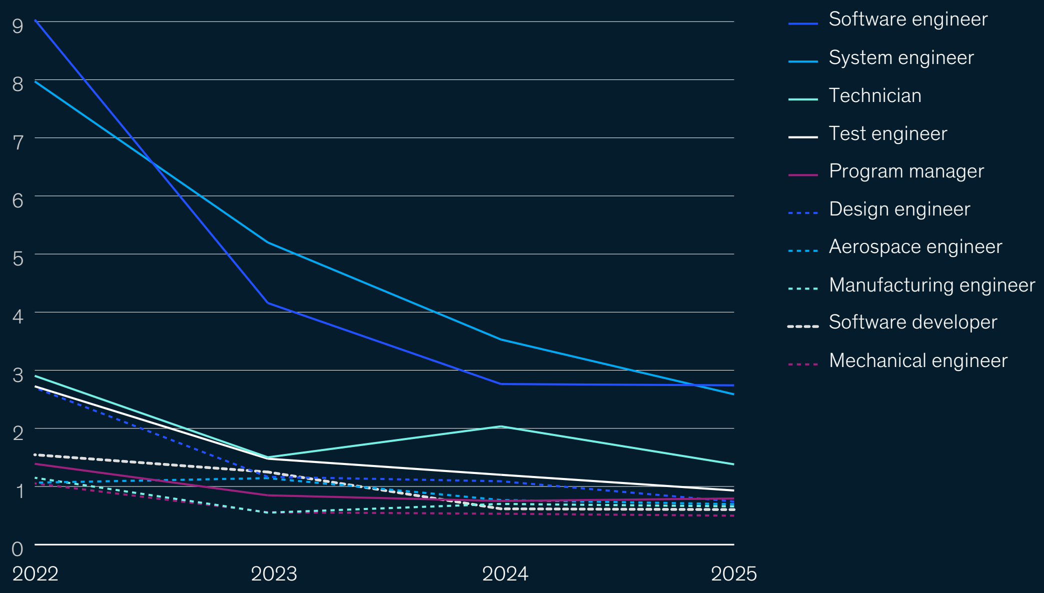
[14] "NVIDIA launches space computing, rocketing AI into orbit," NVIDIA press release, March 16, 2026.

Future of space technologies

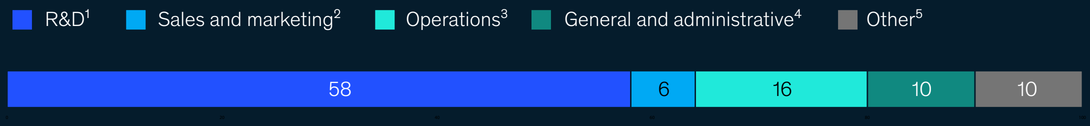
Demand

The space technology talent market has shifted from rapid expansion to consolidation and operational scaling, with job postings well below their 2022 peak and continuing to fall. The pullback has been broad, spanning system engineers, test engineers, technicians, design engineers, and software roles alike. In this frontier sector, R&D represents the majority of job postings.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).

[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).

[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).

[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).

[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

TALENT AND LABOR MARKETS

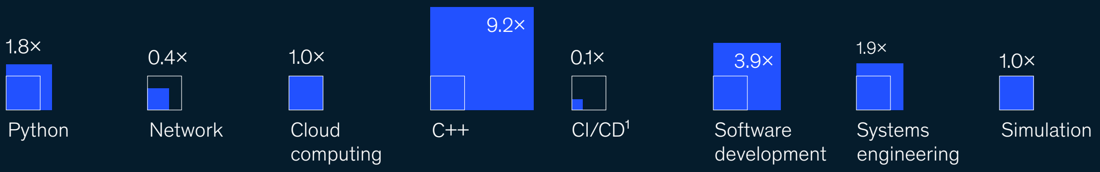
Future of space technologies

Skills availability

The sharpest shortages are in networking and in continuous integration and continuous delivery (CI/CD), pointing to gaps that matter as satellite systems become more software defined and continuously updated. By contrast, traditional development skills, from C++ and software development to systems engineering, are abundant. The pattern reflects a workforce better equipped for traditional engineering than for the modern DevOps (software development and IT operations) and automation capabilities the space sector increasingly needs.

Talent required, % share of postings requiring skill ■

Talent availability, ratio of talent ■ to demand □



[1] Continuous integration and continuous delivery.

Adoption developments across the globe

Adoption score: 2—Experimentation. Few companies have embedded space technologies into recurring business processes at scale, although adoption is advancing in telecommunications, defense, energy, agriculture, insurance, and logistics. Uptake is developing unevenly across regions.

The United States has historically been a leader across all four layers in the space ecosystem, and that continues, including in launch, spacecraft and infrastructure, networks, and space-enabled services.

China is advancing activity across launch, Earth observation, space stations, lunar exploration, and orbital infrastructure. Its state-backed strategy and manufacturing capacity make it a central player in the global space economy.

Europe has developed capabilities in Earth observation, weather forecasting, satellite navigation, and climate monitoring, driven by institutions such as the European Space Agency (ESA) and flagship programs including Copernicus, the European Union's Earth observation program, and Galileo, Europe's global satellite navigation system. Today, the region is increasingly focused on resilience, sovereignty, and commercial satellite services, including through initiatives such as IRIS², the European Union's planned multiorbit satellite communications constellation.[15] Europe's main challenge is scaling private investment, launch access, and downstream commercialization to convert its technical strengths into commercial outcomes.

---

[15] "Welcome IRIS²: Infrastructure for resilience, interconnectivity and security by satellite," European Commission, November 17, 2022.

India and Japan have longstanding space programs with activity across launch, exploration, satellite systems, and downstream applications. Australia, Canada, New Zealand, and South Korea are expanding capabilities across satellite manufacturing, launch, ground infrastructure, and space-enabled services, while the Middle East and Singapore are emphasizing sovereign capabilities and application-led adoption. Across these markets, demand is often concentrated in three overlapping areas: Earth observation applications, satellite communications infrastructure, and defense-related use cases.

Underlying technologies

The technologies that power the future of space include the following:

Earth observation and geospatial intelligence	These technologies include satellite-based sensing, analytics, and data fusion systems to support decision-making across climate, agriculture, energy, logistics, insurance, disaster response, and defense.
Ground systems, terminals, and optical communications	These platforms connect space assets with terrestrial networks and help move data between satellites, users, and cloud or enterprise systems.
In-space servicing	In-space servicing systems can perform inspection, refueling, repair, repositioning, and eventual debris removal on satellites and spacecraft. These systems can extend satellite lifetimes and reduce orbital congestion, although commercial models are still developing.
Onboard processing and orbital compute	These capabilities can analyze data closer to where it is generated, reducing latency and bandwidth needs. Larger-scale orbital compute remains in its early stages and would require advances in power, thermal management, radiation tolerance, networking, and servicing.
Reusable and super-heavy launch systems	Rockets with reusable components and significantly increased payload capacity reduce the cost of reaching orbit and make it easier to deploy larger satellites, constellations, and orbital infrastructure.
Satellite connectivity and nonterrestrial networks	This includes satellite broadband, D2D connectivity, and integration with terrestrial mobile and internet networks to extend communications coverage.
Satellite constellations and orbital architectures	These are networks of satellites across low Earth orbit, medium Earth orbit, geostationary orbit, and cislunar space that support communications, sensing, navigation, defense, and scientific missions.
Space situational awareness and traffic management	These systems track satellites, debris, collision risks, and other orbital activity to help manage increasingly crowded orbital environments.

Key uncertainties

The major uncertainties affecting the future of space technologies include the following:

— *Supply–demand equilibrium for space infrastructure.* The market impact of super-heavy and lower-cost launch will depend on whether supply and demand scale in balance. On the supply side, reliable reuse, higher launch cadence, and lower costs could make it easier to deploy larger constellations, heavier satellites, and orbital infrastructure. On the demand side, downstream adoption of space-enabled services will grow only as use cases are proven. The resulting price point—shaped by both launch supply and customer demand—will determine whether the space industry can scale effectively.

- *Commercial viability of D2D satellite connectivity.* Satellite-enabled D2D connectivity is becoming operational, but the durability of associated profit pools remains uncertain. Constellation business cases may not rely on D2D alone; they may need to combine multiple revenue sources, including consumer broadband, enterprise connectivity, emergency messaging, and maritime and aviation connectivity.

- *Economic feasibility of orbital compute.* Space-based data centers could help address growing terrestrial AI infrastructure constraints, but major technical and economic barriers remain. Radiation exposure, bandwidth costs, latency, launch costs, and hardware replacement cycles continue to pose challenges to commercial viability.

- *Governance, congestion, and security risks.* The growth of satellite constellations is increasing pressure on spectrum coordination, orbital traffic management, and cybersecurity protections. Rising risks from debris, jamming, spoofing, and cyberattacks are increasing the need for more resilient system architectures.

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with space technologies:

- What mass-intensive use cases could be unlocked by lower launch costs and higher payload capacity, and what use cases could instead be enabled by improved size, weight, and power performance in existing satellite form factors?

- How quickly will AI move from ground operations and decision support into trusted autonomy in spacecraft and satellite constellations?

- How will launch access, spectrum coordination, and supply chain fragmentation affect the evolution of space-enabled services and commercial-defense integration?

- Will sustainable profit pools emerge in growing sectors such as D2D connectivity and in-space satellite servicing?

- What would it take for orbital compute to become commercially viable: cheaper launch, better radiation tolerance, improved thermal management, stronger optical networking, routine servicing, or some combination of these?

Read more about the future of space technologies

Knowledge center

McKinsey’s work in the space sector

Related reading

New space, new rules: Commercial space markets are taking off

The case for data centers in space

Is Europe still on the launchpad? Reshaping its space ecosystem to lead

Space: The $1.8 trillion opportunity for global economic growth

13

Future of life sciences and bioengineering

The future of life sciences and bioengineering combines AI with biology, chemistry, and engineering to design, test, and scale new products. Life sciences study how living things work, while bioengineering applies that knowledge to solve medical and technical problems. Organizations are applying these capabilities across pharmaceuticals, gene editing, cell therapies, diagnostics, synthetic biology, precision fermentation, agriculture, and industrial materials.

The trend—and why it matters

Life sciences and bioengineering have moved past the hype phase and are changing people's lives. A case in point: Gene-editing therapies now offer curative treatments for some patients with sickle cell disease. The ability to discover new breakthroughs and engineer biological systems with precision is no longer the only measure of progress. Going forward, the real test is execution.

In life sciences—a broad field including biology, medicine, genetics, and pharmacology—the next stage depends on whether companies can translate fast-moving science into real products that can be manufactured, regulated, and scaled. In bioengineering, the application of engineering principles to understand and manipulate biological systems, companies are making progress in turning biological designs into scalable products with clear customer value, such as industrial materials and biopharmaceuticals.[1] We group these two categories together because they intersect throughout the research, discovery, and product development life cycle.

Companies in several key sectors are leveraging life sciences research and bioengineering to develop new products. For example, pharmaceuticals companies are using AI-enabled drug discovery, gene editing, CRISPR-based approaches, and precision diagnostics to develop therapies and tools that work in clinical practice and help patients live better lives. The most commercially mature activity is in treatments for oncology, immunology, and rare diseases, where pipeline activity is highest and clinical needs are well defined.[2] In food and agriculture, companies are using gene editing, biological pesticides and fertilizers, and precision fermentation to improve crop production and yields and expand the range of possible products. Some of these include cultivated meat grown from animal cells and nitrogen-fixing microbes that could replace synthetic fertilizers. In the industrial sector, companies are using engineered biological processes to produce chemicals, materials, and ingredients that may be more efficient to manufacture or perform better than conventional alternatives.[3]

The most important shift is that the future of life sciences and bioengineering is becoming more application specific, though business models and adoption curves vary widely. Some companies are using AI-driven processes to develop proprietary therapies, while others are building tools, seed technologies, or production capabilities that serve broader markets.

---

[1] "What is bioengineering?," McKinsey, June 23, 2023; "From linear gates to learning loops: Rewiring biopharma R&D with AI," McKinsey, June 29, 2026.

[2] Guang Yang, Jeff Smith, Emily Wingrove, and Eva Lopez Vidal, "How biopharmaceutical leaders optimize their portfolio strategies," McKinsey, January 9, 2025.

[3] Ji Yeon Kim et al., "Beyond petrochemicals: Challenges and opportunities in industrial-scale biomanufacturing," *Nature Communications*, May 30, 2026, Volume 17, Number 4819.

In life sciences, the advantage is likely split across two types of players: large pharmaceutical companies focused on drug discovery and newer biotech companies that can develop a range of products, such as therapies, engineered foods, industrial enzymes, or biomaterials. Larger pharmaceutical companies may be better positioned to bring AI-enabled workflows to scale because they have the data assets, infrastructure, and financial capacity to invest in innovation.[4] At the same time, AI-native biotechs may move faster in selected use cases, especially where they can take concentrated technical risk and prove value before scaling. However, within industrial biotechnology, the approach is the inverse: Instead of continuing to build manufacturing capacity ahead of demand, companies are identifying high-value molecules first and scaling production around them.[5]

AI is accelerating all of these shifts. Researchers are already using AI models to design proteins, identify drug candidates, and analyze genomic data. This work is opening up experiments that would have been too slow or too expensive to run previously (for more, see AI for scientific discovery and engineering). But AI doesn't eliminate the hard parts that come after discovery, such as wet-lab validation, regulatory approval, manufacturing, confirmation of real-world safety and efficacy, and customer validation. In many cases, it's making these bottlenecks more constraining, surfacing more promising candidates than the system can absorb.[6]

Overall, the future of life sciences and bioengineering is best understood as selective scaling across a long adoption journey. Not everything in this space is moving at the same speed. In life sciences, precision diagnostics, AI-enabled discovery, and CRISPR-based therapies are making the transition from research to clinical use, though adoption still depends on proving patient benefit, integration with clinical workflows, and scaling manufacturing. In bioengineering, functional proteins and selected specialty molecules are closest to commercial reality, while commodity chemicals, food replacement, and cultivated meat still depend on regulation and customer demand catching up.

> **‘AI is expanding the number of promising biological candidates quickly, and new tools and workflows are helping support faster validation. However, as discovery accelerates, constraints remain in wet-lab data that can fully support this validation—as well as in the areas of clinical evidence generation, manufacturing in complex modality cases, and regulatory approval. The next winners will connect AI with wet-lab and clinical data collection and end-to-end workflow reinvention.’**
>
> — Erika Stanzl, partner, Zurich

---

[4] Fay Lin, “Pharma races to scale AI as billions flow into drug discovery,” *Genetic Engineering & Biotechnology News*, July 6, 2026.
[5] Mohamed Soufi, “The four molecules to scale for industrial biotech,” SynBioBeta, March 27, 2026.
[6] “From linear gates to learning loops: Rewiring biopharma R&D with AI,” McKinsey, June 29, 2026.

CUTTING-EDGE ENGINEERING

Future of life sciences and bioengineering

Scoring the trend

Bioengineering investment recovered from its 2022 low and has since remained broadly steady, hovering around $100 billion over the past three years, reflecting disciplined capital allocation to more mature companies pursuing commercialization. Research activity surged between 2022 and 2025 to one of the highest levels among trends, driven by advances in gene editing, AI protein design, and advanced therapeutics, although patent filings remained relatively low.

Equity investment, 2025	**Job postings,** 2024–25, % difference
$101.0 billion	**–10%**

Score, by vector (0 = lower; 1 = higher)

Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

Latest developments

Recent developments involving the future of life sciences and bioengineering include the following:

— *Bioengineering is defining where biology has a right to win.* Innovation in bioengineering is moving away from applying biology as a one-size-fits-all platform and toward deploying applications where biology offers a clear edge over conventional chemistry—whether in cost, safety, supply resilience, or access to molecules that simply can't be made any other way. In an industrial-biotechnology example, the industry group Advanced Biotechnology for Sustainability screened more than 300 molecules and highlighted terpenes, peptides, noncatalytic proteins, and hydroxy acids as the most actionable near-term pathways, underscoring a more molecule-led approach to commercialization rather than capacity-first expansion.[7]

— *Generative biology is improving design, but the bottlenecks are moving downstream.*

---

[7] Neil Renninger et al., *The Molecule Manifesto: Four priority molecule families to unlock industrial biotech scale*, Advanced Biotechnology for Sustainability, March 2026.

Multimodal biological foundation models and AI-enabled design tools are making it easier to generate proteins and therapeutic candidates and to explore and exploit biological pathways.[8] This makes the next bottlenecks more physical and operational—for example, data quality, assay throughput, downstream processing, clinical trials, and customer qualification. Adaptyv Bio's protein-engineering foundry is an early proof point: It combines gen AI, open-source software, synthetic biology, robotics, and microfluidics to validate AI-designed proteins for medicines, enzymes, and sustainable materials.[9]

— *Gene editing is moving into therapeutic translation, but scaling is hard.* Gene editing is moving from proof of concept to the development of therapies for specific diseases, but regulatory success does not translate automatically into broad adoption. In vivo base-editing programs are showing early clinical signals. For example, an early study of a gene-editing treatment aimed at lowering LDL cholesterol showed encouraging results in adults with a rare inherited cholesterol condition. After one infusion, cholesterol-related markers fell in a dose-dependent way, and no serious side effects were reported.[10] The result is promising, but broader adoption will depend on durability, safety, and treatment infrastructure. Casgevy, the first approved CRISPR-based therapy, also shows how manufacturing and reimbursement can slow uptake even after approval. In 2025, 147 patients had their first cell collection, and 64 received infusions.[11]

— *Advanced drug modalities are expanding the therapeutic tool kit.* Beyond gene editing, bioengineering is expanding the range of therapeutic formats available to drug developers. Bispecific and trispecific antibodies, antibody drug conjugates, and radiopharmaceuticals are giving companies more precise ways to intervene in disease biology, especially in oncology and immunology. Many of these approaches are not new, but they are becoming more important strategically as AI improves design capabilities and more programs advance toward clinical trials. The shift marks a move away from a narrower monoclonal-antibody paradigm toward a broader set of engineered therapeutic formats, each with its own manufacturing, delivery, safety, and reimbursement requirements. In 2025, advanced drug modalities continued to advance from promise to practice. Antibody drug conjugates and bispecific antibodies reached additional regulatory milestones, including FDA approval of AstraZeneca and Daiichi Sankyo's Datroway for a form of metastatic breast cancer and accelerated approval of Regeneron's Lynozyfic for heavily pretreated multiple myeloma.[12] At the same time, the drug pipeline continued to broaden, with Johnson & Johnson reporting Phase I data for a trispecific antibody in multiple myeloma.[13]

— *Better in vitro models are improving the path from hypothesis to human relevance.* Organoids and organ-on-a-chip systems use human cells to create simplified models of organs or tissues, allowing researchers to test therapeutic hypotheses in vitro rather than relying solely on traditional cell cultures or animal models. These methods can reduce some of the uncertainty in moving from preclinical results to human outcomes, where translation from mice to people remains a major leap. For bioengineering, the methods could apply to multiple disease areas. Better in vitro models could help companies create better validation targets, speed up drug testing and toxicity assessment, and build the evidence needed to advance new therapeutic modalities. This is becoming more important as regulators seek better ways to predict human response before clinical testing. For example, the FDA released a road map to reduce animal testing in preclinical-safety studies, pointing to alternatives such as organ-on-chip systems, computational

---

[8] Huan Yee Koh et al., "AI-driven protein design," *Nature Reviews Bioengineering*, September 8, 2025, Volume 3, Number 12.

[9] "Adaptyv is the cloud lab for protein designers—and now available for everyone," blog entry, Adaptyv, September 18, 2025.

[10] Ping Wan et al., "In vivo base editing gene therapy for heterozygous familial hypercholesterolemia: A phase 1 trial," *Nature Medicine*, March 2026, Volume 32.

[11] "Vertex reports fourth quarter and full year 2025 financial results," Vertex Pharmaceuticals press release, February 12, 2026.

[12] "FDA approves datopotamab deruxtecan-dlnk for unresectable or metastatic, HR-positive, HER2-negative breast cancer," US Food and Drug Administration, accessed August 2026; "FDA grants accelerated approval to linvoseltamab-gcpt for relapsed or refractory multiple myeloma," US Food and Drug Administration, accessed August 2026.

[13] "Early results from Johnson & Johnson's trispecific antibody show promising response in heavily pretreated multiple myeloma patients," Johnson & Johnson press release, June 3, 2025.

modeling, and advanced in vitro assays.[14] In parallel, CN Bio and Pharmaron announced a long-term strategic partnership to validate CN Bio's PhysioMimix organ-on-chip technology and integrate its capabilities into Pharmaron's global R&D platform.[15]

— *Precision diagnostics are moving into workflows.* New forms of precision diagnostics such as liquid biopsy, in vitro companion diagnostics, and circulating-tumor DNA testing are increasingly being embedded into clinical pathways, especially in oncology. NHS England's expansion of circulating-tumor DNA testing is an early proof point that the biggest gains come from rapid molecular diagnosis, more targeted treatment selection, and fewer repeat invasive procedures.[16] However, the evidence for current liquid biopsy use remains mixed across indications. Companion diagnostics, or tests that identify which patients are likely to respond to a specific therapy before treatment begins, are already part of the regulatory and clinical infrastructure for oncology drugs. The FDA says companion diagnostics provide information essential for the safe and effective use of a corresponding therapeutic product, and its oncology guidance supports codevelopment and aligned labeling of the diagnostic and the drug.[17] The question is no longer whether new types of precision diagnostic tests can be developed, but rather whether they can be reimbursed, interpreted, and tied to clinical decisions.

— *Food and agriculture are becoming outcome led.* Bioengineering is bringing technology to farmers' fields and people's kitchens. Gene-edited crops, biological inputs for crops, and precision-fermented ingredients are gaining traction when they deliver measurable value, such as higher crop yield, novel foods, improved nutrition, or extended shelf life of existing products. For example, the European Parliament recently passed a law that allows the expedited approval of many gene-edited crops in Europe to boost innovation, competitiveness, and food security.[18] Cultivated meat continues to advance technically and progress through selected market access pathways, but broad adoption remains a longer-term prospect because cost, scale, and consumer acceptance remain unresolved.[19] The development of novel foods such as cultured meat relies on advances in stem cell biology, tissue engineering, and bioprocessing. While regulators in countries such as Singapore and the United States have created case-by-case approval pathways, major challenges remain in achieving low-cost, industrial-scale production and increasing consumer acceptance.[20]

— *Biomanufacturing capacity is becoming a strategic control point.* The ability to produce biological products reliably and economically is becoming a key growth constraint. This is most visible in industrial biotechnology. The life sciences companies bringing products to market are those that have overcome challenges related to raw materials, fermentation capacity, and the economics of scaling production. Biopharma companies in particular face bottlenecks in scaling cell culture systems, obtaining quality and regulatory evidence, and equipping treatment centers with product knowledge. In practice, these barriers are pushing the life sciences industry toward pilot plants and shared infrastructure before major capital deployments. In April 2025, BioMADE—a biomanufacturing institute backed by the US Department of Defense—announced an agreement to acquire and transition Lygos's pilot-scale biomanufacturing facility into a nationally available, multiuser site. This news illustrates how biomanufacturing could increasingly be organized around shared manufacturing capacity rather than more risky stand-alone build-outs.[21]

---

[14] *Roadmap to reducing animal testing in preclinical safety studies*, US Food and Drug Administration, April 2025.

[15] "CN Bio and Pharmaron establish long-term strategic partnership to develop OOC technologies on a global R&D platform," CN Bio press release, April 24, 2025.

[16] "Implementation of circulating tumour DNA (ctDNA) testing in the NHS," NHS England, April 24, 2026.

[17] "Companion diagnostics," US Food and Drug Administration, accessed August 2026.

[18] "New genomic techniques for plants to boost innovation in sustainable agriculture," European Parliament press release, June 17, 2026.

[19] Sungmin Kim et al., "Cultured meat: Advances in stem cell biology, tissue engineering, and bioprocess optimisation for scalable and sustainable production—a review," *International Journal of Food Science and Technology*, July 2025, Volume 60, Number 2.

[20] "SFA list of approved novel foods," Singapore Food Agency, updated July 22, 2026.

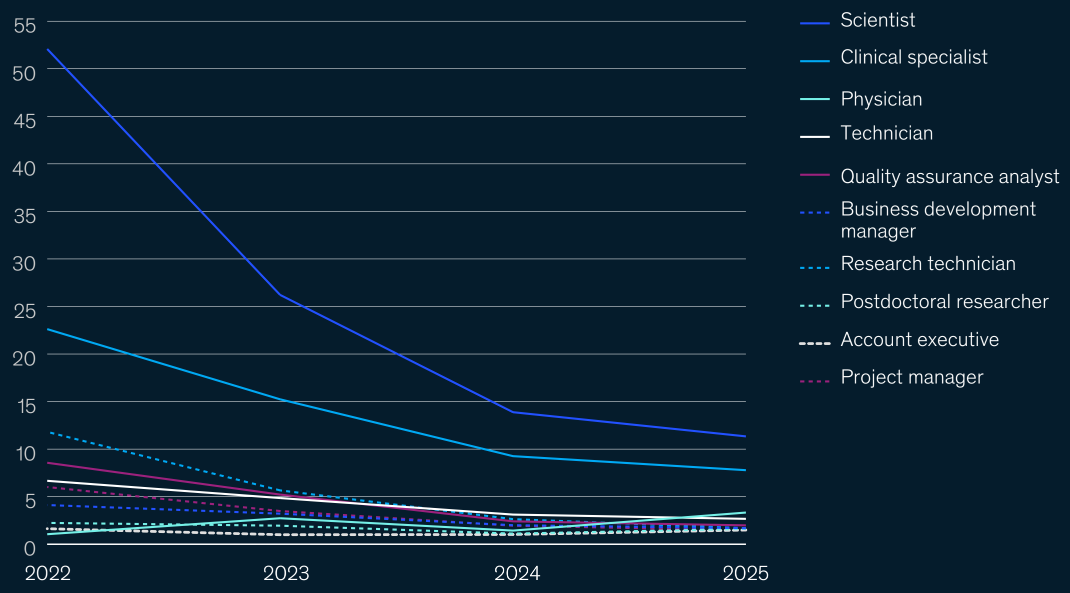
[21] "BioMADE and Lygos to fast-track pilot biomanufacturing facility in California," BioMADE press release, April 30, 2025.

Future of life sciences and bioengineering

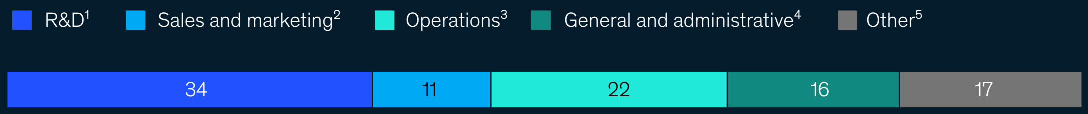
Demand

Life sciences and bioengineering job postings remained well below their 2022 peak and continued to decline from 2024 to 2025, reflecting a sector shifting from broad research expansion toward more selective, application-focused hiring. However, job postings for physicians, postdoctoral researchers, and account executives grew in 2025, pointing to rising demand at the intersection of clinical development and commercial deployment. On a functional level, operations and general and administrative roles account for over a third of job postings, while R&D sits notably below the cross-sector baseline. This reinforces that this trend is broadly transitioning from pure discovery toward scale-up and market access.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).
[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).
[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).
[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).
[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

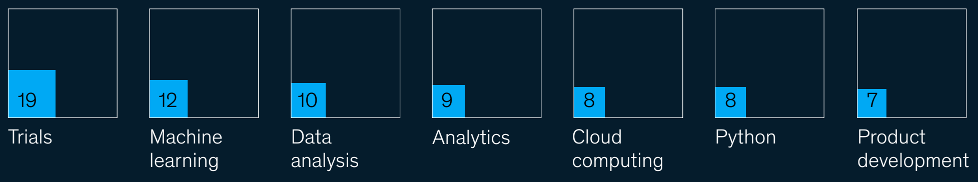
TALENT AND LABOR MARKETS

Future of life sciences and bioengineering

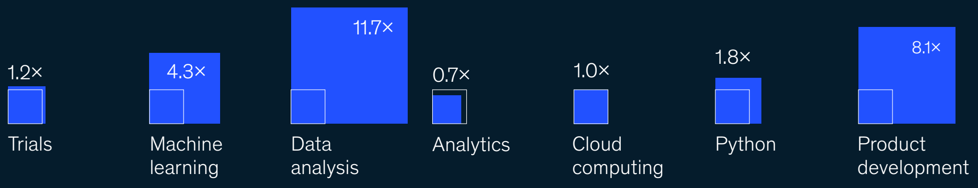
Skills availability

The supply of bioengineering talent is generally strong. Clinical-trials expertise is by far the most in-demand skill, yet supply is roughly in line with demand. Skills such as data analysis, product development, and machine learning all exceed demand. The one area where supply falls modestly short is analytics, and even that gap is small, suggesting talent availability is not currently a binding constraint even as the field scales.

Talent required, % share of postings requiring skill ■



Talent availability, ratio of talent ■ to demand □



Adoption developments across the globe

Adoption score: 4—Scaling in progress. Life sciences and bioengineering companies are scaling the adoption of new technologies in selected areas, especially in therapeutics, diagnostics, R&D, precision fermentation, and biomaterials. However, adoption varies widely across subfields.

Adoption is becoming more regionally differentiated. The United States has a mature ecosystem for clinical bioengineering, AI-enabled discovery, precision fermentation, and early commercialization. It benefits from strong venture capital investment, established pharmaceutical and technology ecosystems, a network of research universities, and flexible regulatory pathways. However, translating lab breakthroughs into economically viable production remains a challenge.

Europe combines strong research institutions, mature manufacturing, and established chemical infrastructure, although higher costs, cautious public attitudes, and regulatory complexity may slow commercialization in some applications, particularly in food and agriculture. China has developed significant manufacturing scale and production capabilities in bioengineering. This scale is likely to influence global cost structures and encourage companies in other regions to differentiate through specialized products, supply chain resilience, or higher-value applications.

India is becoming an important life sciences and bioengineering hub in both agriculture and biomanufacturing. Regulatory changes for certain gene-edited crops, combined with national efforts to expand biomanufacturing hubs, could increase India's role in sustainable manufacturing and food security.

Singapore remains a notable test bed for novel foods. Its regulatory pathways have supported market access for cultivated, fermentation-derived, and other novel food products. This makes it a reference market for companies testing how new food technologies could move from safety review to commercialization.

Progress in emerging markets may be shaped by trade dependencies, food security priorities, and regulatory roadblocks. These regions should not be viewed solely as end markets for imported bioengineered products. Local research institutions, agricultural innovation systems, public-health needs, and regional manufacturing efforts also could shape where bioengineering capabilities develop.

Underlying technologies

The technologies that power the future of life sciences and bioengineering include the following:

Agricultural biotechnology	Gene editing, microbial inputs, and biological crop protection tools are agricultural technologies that improve yield and resilience and reduce reliance on chemical inputs.
AI-enabled biological design	AI-powered tools can support protein design, molecule discovery, pathway design, strain engineering, and experimental automation.
Biomanufacturing infrastructure	These are the facilities, equipment, and quality systems needed to take a biological product from lab scale to commercial production.
Biosecurity, containment, and monitoring	These systems help support the responsible development and deployment of engineered biological systems.
Gene editing	Tools such as CRISPR, base editing, and prime editing, along with delivery systems such as lipid nanoparticles and viral vectors, enable precise changes to biological systems.
Human-relevant preclinical models	Organoids, organ-on-a-chip systems, and microphysiological systems help researchers test biology and therapeutics in vitro in ways that better approximate human responses.
Omics	Single-cell sequencing, proteomics, metabolomics, multiomics, and biological data platforms help generate and interpret high-resolution biological data.
Precision diagnostics	Tests and tools, including liquid biopsies and AI-enabled analysis, improve how diseases are detected and matched to the right treatment.
Precision fermentation and cell culture	These are methods used to produce proteins, food ingredients, peptides, biomaterials, and specialty molecules for food, medicine, and industrial applications.
Protein and therapeutic modality engineering	Antibodies, bispecific and trispecific formats, antibody-drug conjugates, radiopharmaceuticals, and other engineered therapeutic modalities enable targeted intervention in disease biology.
Synthetic biology and metabolic engineering	These approaches are used to design and reprogram cells, microbes, enzymes, and biological pathways for research and discovery.
Tissue engineering	Technologies including 3D bioprinting and biomaterials can grow or reconstruct biological tissue for repair, replacement, or transplantation.

Key uncertainties

The major uncertainties affecting the future of life sciences and bioengineering include the following:

— *Regulation, reimbursement, and market access.* Advanced therapies and drug modalities still face manufacturing, reimbursement, and treatment site bottlenecks. In vivo and personalized gene-editing therapies must demonstrate safety and efficacy at scale, and regulatory systems could converge or fragment across major markets.

— *Diagnostic and model validation.* Precision diagnostics and better in vitro models may improve evidence generation, but they still need to prove routine clinical relevance, regulatory credibility, and workflow fit.

— *Manufacturing scale-up and industrial economics.* Industrial biotechnology needs enough offtake commitments to finance first-of-a-kind and second-of-a-kind facilities. Bio-based chemicals will need either cost parity or differentiated performance in markets shaped by incumbent petrochemical routes and global overcapacity, while feedstock availability, land use, energy costs, and downstream processing could constrain the economics of bio-based production.

— *Consumer adoption and public acceptance.* Cultivated meat and broader novel food markets still need to solve cost, taste, regulatory, and consumer adoption challenges. Gene-edited crops face uneven public acceptance across regions, while biosecurity and public trust could slow adoption of otherwise technically viable products.

— *AI's impact on development economics.* AI is accelerating protein design, molecule discovery, and experimental workflows, but it's not removing all the obstacles of getting products to market. Companies still need to do wet-lab validation and clear regulatory hurdles. The open question is whether AI meaningfully lowers development costs and timelines or mainly increases the number of candidates that still require validation and scale-up.

> **'Without a doubt, the broad advances in biotech, automation, and AI underway will accelerate innovation across the board. But they'll also shift the bottlenecks and change the basis of competition in unexpected ways; it's not yet clear who the future winners will be.'**
>
> — Tom Brennan, partner, Philadelphia

Big questions about the future

Companies and leaders may want to consider a few questions when moving forward with life sciences and bioengineering:

- Which applications create enough differentiated value to justify biological production before full commodity cost parity?
- Which opportunities should be prioritized first: advanced therapeutics, precision diagnostics, better in vitro models, functional proteins, peptides, gene-edited crops, or cultivated foods?
- What is the primary bottleneck for each opportunity: delivery, durability, downstream processing, feedstock, regulatory approval, reimbursement, customer qualification, or offtake?
- How will AI reshape the value chain, from discovery and design to validation, clinical development, process control, and quality assurance?
- Who will capture value across the ecosystem: intellectual property owners, data platforms, contract manufacturers, ingredient companies, brands, health systems, or downstream customers?

Read more about the future of life sciences and bioengineering

Knowledge center

Biotech

Insights on Life Sciences

Related reading

Biopharma R&D: The evolving formula for discovery and development

From linear gates to learning loops: Rewiring biopharma R&D with AI

Pharmaceuticals: Innovating and advancing around the world

Reimagining life science enterprises with agentic AI

14

Future of energy and sustainability technologies

Energy and sustainability technologies are transforming how energy is produced, stored, transported, managed, and used. This trend covers the technologies reshaping energy systems, spanning cleaner electricity, storage, grid modernization, electrification, and low-carbon infrastructure such as nuclear energy, hydrogen, and e-fuels.

The trend—and why it matters

Energy and sustainability technologies don't just support growth and innovation; they make it possible in the first place. Almost every modern industry runs on power. The trend spans mature technologies such as solar, wind, lithium-ion batteries, heat pumps, and energy efficiency systems, as well as earlier-stage technologies still proving themselves through experiments and pilots. These include advanced nuclear fission, next-generation geothermal, long-duration storage, low-emission hydrogen, carbon and direct air capture, and nuclear fusion.

As the world thirsts for more power, energy and sustainability technologies need new innovations to meet that demand. Progress is mixed. Low-emission power generation is scaling, but variable renewable energy will require more storage, grid flexibility, and transmission capacity to evolve.[1] Clean molecules such as hydrogen, ammonia, and methanol are being developed for sectors that are difficult to electrify directly, but these are not yet widely used.[2] Meanwhile, digital systems such as AI-enabled forecasting, digital twins, and satellite-based emissions monitoring are beginning to help companies and grid operators manage more distributed and dynamic energy systems.[3]

The most visible pressure point is meeting surging electricity demand. Data centers are becoming a meaningful source of incremental load, adding demand to already strained grids (for more, see AI infrastructure and model architectures). In the United States, data center electricity consumption could account for 14 percent of total power demand by 2030.[4] Globally, data center electricity demand could approach or exceed 1,000 terawatt-hours by 2030.[5] This demand growth is changing the question for utilities, energy developers, and large-scale power consumers from simply "how cheap is the power?" to "how quickly can reliable power be secured?"

They are finding that grid infrastructure and physical supply chains are central bottlenecks. Multiple power generation and large-load projects are stalled, waiting for grid connections. Renewable-energy sources such as solar, wind, and batteries can often be built faster than the transmission lines, substations, transformers, and interconnection processes needed to connect them.[6] These delays are shifting attention away from individual technologies to the need for integrated energy systems that combine generation and storage with transmission and digital controls. The past year brought several technology milestones that show where the energy system is moving next. Clean firm power gained momentum through new nuclear and geothermal project activity; next-generation solar advanced toward commercialization; and sensor- and satellite-enabled measurement systems became more operational.

At the same time, the policy and investment environment is complex—and only getting more so. In some markets, the emphasis is expanding from meeting carbon targets alone toward a broader set

---

[1] "How grid operators can integrate the coming wave of renewable energy," McKinsey, February 8, 2024.
[2] "What is hydrogen energy?," McKinsey, October 2, 2024.
[3] "Software: The hidden catalyst for decarbonization," McKinsey, December 17, 2025.
[4] "Global Energy Perspective 2025," McKinsey, October 13, 2025.
[5] "Energy supply for AI," in *Energy and AI*, IEA, April 10, 2025.
[6] *Electricity 2026*, IEA, February 6, 2026.

of priorities: reliability, energy security, industrial competitiveness, domestic manufacturing, and speed of deployment.[7] This does not mean decarbonization has become irrelevant. Rather, lower-emission technologies increasingly need to prove that they can support resilient and cost-effective energy systems. Only then can they secure the capital and grid access needed to scale. As a result, decarbonization is likely to advance through pockets of competitiveness. Specific regions and sectors with low-cost resources, supportive policies, and infrastructure availability could see low-emission technologies become economically viable sooner than elsewhere.

Overall, the future of energy and sustainability technologies is becoming as much a question of deployment as of innovation. Companies and governments will need to decide which lower-emission technologies to fund, where to deploy them, and how to integrate them into existing systems. Thus, the leaders most likely to pull ahead will be those that can combine technology with execution discipline: building faster, maintaining affordability and reliability, and reducing emissions where the economics and infrastructure allow.[8]

CUTTING-EDGE ENGINEERING

Future of energy and sustainability technologies

Scoring the trend

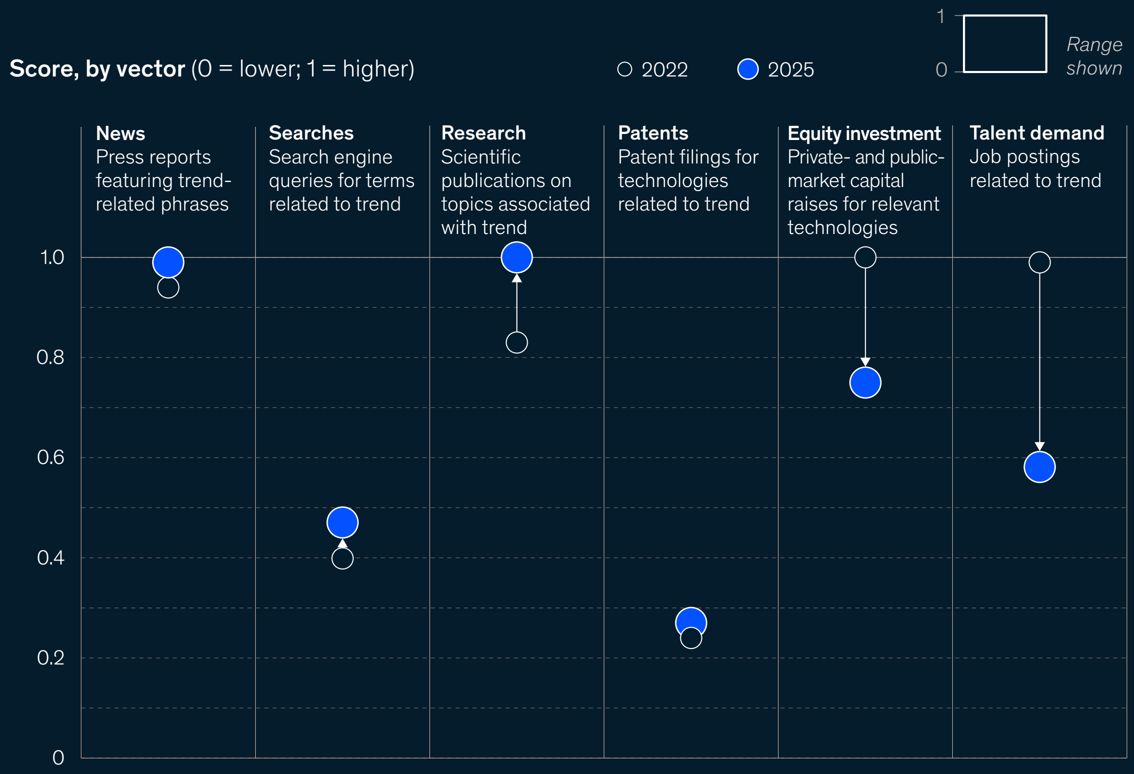
Despite equity investment declining from its 2022 peak, energy and sustainability technologies have the highest research publication score of any trend and one of the highest levels of news coverage, reflecting sustained scientific and public attention to energy supplies and transitions. Investment reached over $194 billion in 2025, the highest of any trend that year, driven by renewables, storage, and grid infrastructure build-outs.

Equity investment, 2025

$194.2 billion

Job postings, 2024–25, % difference

–19%



Score, by vector (0 = lower; 1 = higher)

Note: For each vector, we used a defined set of data sources to find occurrences of keywords associated with each of the 14 trends, screened those occurrences for valid mentions of activity, and indexed the resulting numbers of mentions on a 0–1 scoring scale that is relative to the trends studied.

---

[7] “Global Energy Perspective 2025,” McKinsey, October 13, 2025.
[8] “The hard stuff 2025: Taking stock of progress on the physical challenges of the energy transition,” McKinsey, November 6, 2025.

Latest developments

Recent developments involving the future of energy and sustainability technologies include the following:

- *Large energy users are moving from power procurement to power orchestration.* Rising demand from artificial intelligence and cloud infrastructure is pushing hyperscalers and other large-load customers to secure energy more directly, rather than relying only on standard utility procurement timelines.[9] Some are pursuing long-term clean firm power contracts, dedicated infrastructure, and behind-the-meter (on-site) options to reduce exposure to interconnection delays and power scarcity. Google, Kairos Power, and the Tennessee Valley Authority, for example, announced a power purchase agreement for up to 50 megawatts of advanced nuclear power from the Hermes 2 plant to support Google data centers in Tennessee and Alabama.[10] Amazon Web Services and Talen Energy also expanded a nuclear power agreement up to 1,920 megawatts from the Susquehanna plant in Pennsylvania.[11]

  These agreements are early proof points that large energy users are looking beyond conventional procurement models, but they do not mean advanced nuclear or private power systems are broadly commercialized at scale. Many hyperscalers continue to rely on fossil fuel power to support data center growth, and emissions are not falling across the board. Microsoft's total emissions, for example, rose 25 percent year over year, driven largely by data center expansion. The company also recently

> **'This phase of the energy transition is much more of a balance between inventing entirely new technologies and more about deploying them at scale, balancing affordability, resilience, and emissions. Grid capacity, transmission, storage, and interconnection are becoming the constraints that determine how quickly reliable, low-emission energy can be delivered. Behind-the-meter generation and management of energy are increasingly in focus. The leaders will be those who execute across technology, infrastructure, and policy at the same time.'**
>
> — Mark Patel, senior partner, Bay Area

---

[9] "The next big shifts in AI workloads and hyperscaler strategies," McKinsey, December 17, 2025.
[10] "Google, Kairos Power, TVA collaborate to meet America's growing energy needs," Kairos Power press release, August 18, 2025.
[11] Ethan Howland, "Talen to sell Amazon 1.9 GW from Susquehanna nuclear plant," Utility Dive, June 11, 2025.

signed a 20-year agreement with Chevron for a roughly 2.67-gigawatt natural-gas-fired plant in West Texas dedicated to one of its data centers.[12]

- *Grid constraints are a bottleneck for renewable and large-load projects.* Solar, wind, storage, and large-load projects can often be planned faster than the transmission, substations, and transformers needed to connect them. The IEA estimates that more than 2,500 gigawatts of renewable, storage, and large-load projects are stalled in grid queues worldwide, underscoring that physical infrastructure and regulatory processes have become binding constraints on deployment.[13] This is shifting attention to faster grid expansion, interconnection reform, and grid-enhancing technologies such as dynamic line ratings and power flow control. These tools can unlock capacity in selected markets, but they are complements to—not substitutes for—major transmission and distribution build-outs.

- *Capital is rotating toward the physical backbone of the energy system.* Investment momentum is becoming more selective, with more capital directed to grids, storage, power generation, and commercially contracted infrastructure rather than broad climate technology categories. Global energy investment could reach $3.4 trillion in 2026, with investments in power sector projects alone expected to reach $1.6 trillion.[14] Battery storage is also becoming a more integrated grid asset. For example, in Texas, the Electric Reliability Council of Texas (ERCOT) Real-Time Co-optimization plus Batteries initiative integrates battery storage more directly into real-time grid operations, improving the management of electricity supply and grid-balancing services.[15] The implication is that awareness is growing about the role batteries can play in grid flexibility and reliability.

- *Clean firm power is gaining ground, but commercialization remains uneven.* As companies seek reliable, low-emission power that can run around the clock, advanced nuclear fission and next-generation geothermal are receiving growing attention. The Google–Kairos–TVA agreement is one early signal of how advanced nuclear could serve data center demand if projects can meet technical, regulatory, and cost milestones.[16] Geothermal is also emerging as a potential clean firm power option: Fervo Energy and NV Energy received approval to supply Google data centers with 115 megawatts of enhanced geothermal energy in Nevada.[17] These developments show growing demand for clean firm power, but advanced nuclear, small modular reactors, and enhanced geothermal remain earlier in deployment than wind, solar, and lithium-ion batteries.

- *Next-generation solar and industrial electrification are moving from lab progress toward early commercial proof points.* Mature renewables continue to scale, but innovation is also advancing in technologies that could improve land productivity or decarbonize industrial heat and materials. Oxford PV shipped its first commercial perovskite-on-silicon tandem solar modules to a US customer in 2024 and later signed a patent-licensing agreement with Trinasolar for perovskite-based photovoltaic products in China.[18] In steelmaking, Stegra's hydrogen-based green steel plant in Boden, Sweden, is working to become one of the first commercial-scale facilities to produce primary steel with deep emissions reductions.[19] Both examples are important in field- or industrial-test milestones, but their broader impact will depend on durability, cost, manufacturing scale, and customer willingness to pay for lower-emission products.

---

[12] "Chevron signs 20-year power agreement with Microsoft for West Texas data center," Chevron press release, June 22, 2026; *2026 Environmental sustainability report*, Microsoft, July 2026.

[13] "Grids," in *Electricity 2026*, IEA, February 6, 2026.

[14] *World energy investment 2026*, IEA, May 28, 2026.

[15] "ERCOT goes live with real-time co-optimization plus batteries (RTC+B)," ERCOT news release, December 5, 2025.

[16] "Google, Kairos Power, TVA collaborate to meet America's growing energy needs," Kairos Power, August 18, 2025.

[17] "Our first-of-its-kind partnership for clean energy has been approved in Nevada," blog entry by Briana Kobor, Google, May 13, 2025.

[18] "20% more powerful tandem solar panels enter commercial use for the first time in the US," Oxford PV press release, September 5, 2024; "Oxford PV and Trinasolar announce a landmark perovskite PV patent licensing agreement," Oxford PV press release, April 9, 2025.

[19] Jane Birch, "Slowing momentum and persistent barriers set up a critical year for green steel in 2026," Stockholm Environment Institute press release, February 11, 2026.

- *Hydrogen and alternative fuels are entering a more disciplined phase.* Hydrogen, ammonia, methanol, and synthetic fuels remain relevant for hard-to-abate sectors such as steel, chemicals, aviation, shipping, and select industrial processes. However, elevated costs, infrastructure gaps, and uneven willingness to pay a green premium are forcing more selective project development. Progress is strongest where policy support, low-cost power, infrastructure, and offtake are aligned. In aviation, SWISS integrated Synhelion's solar fuel into flight operations, using fuel derived from Synhelion's DAWN facility.[20] This is a meaningful commercialization signal for solar-derived fuels, but it remains a small-volume proof point rather than evidence of scaled fuel availability.[21]

- *Digital measurement, reporting, and verification are moving toward operational use.* Measurement, reporting, and verification are shifting from periodic manual audits toward more automated, sensor- and satellite-enabled systems. Carbon Mapper's Tanager-1 satellite has published more than 5,300 methane plume observations and more than 1,200 carbon dioxide plume observations, including facility-level detections across oil and gas, waste, coal, agriculture, and power generation.[22] This kind of remote sensing can make emissions more visible to regulators, investors, customers, and corporate supply chains.[23] The technology is increasingly operational, but its influence on regulation, finance, and procurement will depend on data quality, transparency, and the ability of organizations to act on the findings.

- *Energy strategies are becoming more regional and more tied to industrial competitiveness.* The execution of energy and sustainability strategies is increasingly shaped by energy security, supply chain resilience, domestic manufacturing, and trade policy, not only emissions targets. The European Union's Carbon Border Adjustment Mechanism, for example, is designed to place a carbon price on selected carbon-intensive imports and enters its definitive regime in 2026 after a transitional phase.[24] Policies such as this can shift investment decisions, supplier strategies, and industrial competitiveness across regions. The result is not one uniform global energy transition but multiple regional pathways, each shaped by local power costs, resource availability, grid constraints, policy design, and industrial priorities.

> **'Competition for reliable, low-emission power is expected to increase due to the coinciding growth of data center investments and electrification of transport and industry. This will create pressures on both cost and time to power, putting demands on players to secure power supply much more proactively than in the past.'**
>
> — Anna Granskog, partner, Helsinki

---

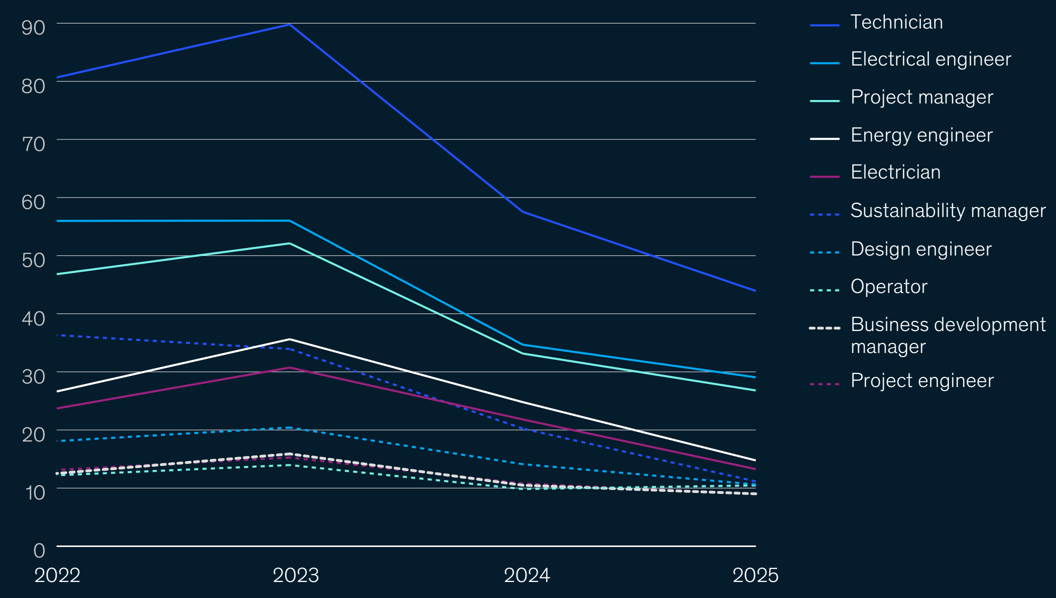
[20] "SWISS becomes first airline to integrate Synhelion's solar fuel into its flight operations," SWISS press release, July 24, 2025.
[21] *Global Hydrogen Review 2025*, IEA, September 12, 2025.
[22] "Tanager-1: One year in space," Carbon Mapper, September 23, 2025.
[23] "First greenhouse gas plumes detected with NASA-designed instrument," NASA Jet Propulsion Laboratory, October 10, 2024.
[24] "Carbon Border Adjustment Mechanism," European Commission Directorate-General for Taxation and Customs Union, accessed August 2026.

Future of energy and sustainability technologies

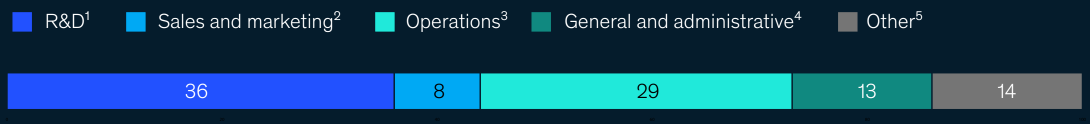
Demand

Energy and sustainability job postings peaked in 2023 and have fallen sharply since, continuing to decline from 2024 to 2025. The pullback has been broad, with most core build-out roles, including technicians, electrical engineers, project managers, and energy engineers, down roughly 40 percent or more since 2022. Sustainability managers have fallen even further, likely reflecting a consolidation of dedicated sustainability functions. At a functional level, job postings skew to non-R&D roles such as operations and general and administrative, suggesting a sector leaning more toward physical deployment and infrastructure.

Job postings, by title, 2022–25, thousands



Job postings, by business function, 2025, % share



[1] Roles that create or improve products, technology, or intellectual property (eg, researchers and scientists, data scientists, software engineers, AI and machine learning engineers, data engineers).
[2] Roles driving demand, customers, and revenue (eg, sales and business development representatives, account executives, customer success, marketing).
[3] Roles that produce, deliver, support, or run the core product or service (eg, technical and customer support, service delivery, field service and IT support technicians, systems and network administrators).
[4] Internal enabling functions (eg, finance, legal and compliance, HR and recruiting, procurement and contracts, security governance).
[5] Mixed or nonstandard titles that did not map directly to any function's criteria (eg, case manager, project lead, director, data specialist).

TALENT AND LABOR MARKETS

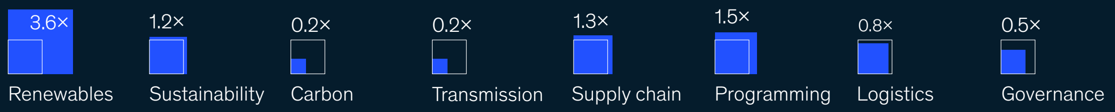
Future of energy and sustainability technologies

Skills availability

Renewables and sustainability are among the most frequently required skills across the sector, and supply has broadly kept up: Renewables run above demand, while sustainability keeps pace. The sharpest shortages are in carbon expertise and grid transmission—gaps tied to cutting emissions and completing infrastructure modernizations.

Talent required, % share of postings requiring skill ■

Talent availability, ratio of talent ■ to demand ☐



Adoption developments across the globe

Adoption score: 4—Scaling. Organizations are moving past pilots and deploying energy and sustainability technologies at scale, notably to add grid capacity and to power data centers. Some technologies, such as solar, wind, lithium-ion batteries, heat pumps, and energy efficiency solutions, have become default areas for energy investment. Others, such as green hydrogen, e-fuels, long-duration storage, next-generation geothermal, small modular nuclear reactors, and direct air capture, remain in early commercialization or project development stages.[25]

Adoption is regionally uneven. China accounts for the largest share of global manufacturing capacity across solar, batteries, electric vehicles, and electrolyzers. The country is also expanding renewable power, nuclear, and hydrogen capacity in ways that could reshape global cost curves and supply chains. China's annual power generation is projected to nearly double, from roughly 10,000 terawatt-hours in 2023 to about 21,000 terawatt-hours by 2050, with low-carbon sources accounting for approximately 92 percent of that total.[26] That said, China still relies on coal to reliably meet industrial demand.

The United States is seeing rapid growth in data center power supply, renewable-energy development, battery storage, nuclear, geothermal partnerships, and industrial policy-driven manufacturing. However, grid interconnection, transmission, permitting, and equipment bottlenecks remain significant constraints.

Europe is focused on energy security, industrial competitiveness, carbon border rules, offshore wind, hydrogen infrastructure, grid modernization, and industrial decarbonization. High energy prices, permitting complexity, and low industrial demand are key challenges, but Europe benefits from strong policy frameworks that continue to spur low-emission investments.

---

[25] "Tracking the energy transition: Where are we now?," McKinsey, January 14, 2026.

[26] "Global Energy Perspective 2025," McKinsey, October 13, 2025.

India is scaling solar, grid infrastructure, electrification, and domestic manufacturing. The country's energy transition is shaped by rapid demand growth, affordability, industrial development, and the need to expand both generation and grids. India's annual power generation could triple, from about 2,000 terawatt-hours in 2023 to roughly 6,000 terawatt-hours by 2050, with low-carbon sources reaching approximately 85 percent of that mix.[27]

The Middle East is investing in solar, hydrogen, ammonia, carbon capture, and energy-intensive industrial development, often supported by existing oil and gas infrastructure. The region could play a larger role in clean fuels and industrial decarbonization, but low-emission hydrogen and ammonia remain challenged by production costs, infrastructure needs, and uncertain export demand.

Emerging and developing economies face the dual challenge of expanding energy access while decarbonizing. Getting access to financing and grid capacity, as well as policy uncertainty, can limit projects even where renewable resources are abundant. These markets will require more tailored financing models and infrastructure support.

Overall, adoption is strongest where technologies address an immediate system need: reliable power for new loads, lower energy costs, grid flexibility, or access to low-emission energy. Adoption is slower where technologies depend on new infrastructure build-outs, uncertain policy, or a lack of customers willing to pay large green premiums.

Underlying technologies

The technologies that power the future of energy and sustainability include the following:

Adaptation and resilience	This includes climate-risk analytics, wildfire monitoring, flood protection, and resilient infrastructure that help organizations prepare for and respond to physical climate risks.
Carbon management	This includes carbon capture and storage, direct air capture, and mineralization technologies that remove or permanently store carbon dioxide.
Clean firm power	Nuclear reactors, small modular reactors, next-generation geothermal, and hydropower provide low-emission electricity when variable renewable generation is limited.
Clean molecules and low-carbon fuels	Hydrogen, ammonia, methanol, synthetic fuels, electrolyzers, and fuel cells support decarbonization in sectors difficult to electrify directly.
Critical minerals and circularity	Extraction, processing, recycling, and substitution technologies support the supply of materials needed for batteries, grids, and renewables.
Digital energy systems	These are AI-enabled forecasting, grid digital twins, energy management software, automated dispatch, and energy trading platforms that improve decision-making across generation, storage, transmission, and demand.
Electrification technologies	Heat pumps, electric boilers, electric furnaces, and vehicle-charging infrastructure replace fossil-fuel-based processes with electricity-based alternatives.

---

[27] "Global Energy Perspective 2025," McKinsey, October 13, 2025.

Energy storage	Lithium-ion batteries, sodium-ion batteries, flow batteries, thermal storage, pumped hydro, and other long-duration technologies balance electricity supply and demand across varying time horizons.
Fusion energy	Magnetic-confinement, inertial-confinement, and alternative fusion approaches aim to generate power by replicating the nuclear process that powers stars.
Grid infrastructure and flexibility	Transmission lines, substations, transformers, grid-enhancing technologies, demand-response platforms, and virtual power plants move and balance electricity across power systems.
Measurement, reporting, and verification	Satellites, sensors, lidar, and data analytics platforms measure emissions, verify reductions, and identify operational risks.
Renewable-energy systems	Solar, wind, hydropower, geothermal, and marine technologies generate low-emission electricity from natural resources.

Key uncertainties

The major uncertainties affecting the future of energy and sustainability technologies include the following:

— *Grid and infrastructure execution.* Rising electricity demand, interconnection queues, transmission constraints, permitting delays, land access, and shortages of transformers, cables, and other equipment could slow the deployment of both new generation and large loads.

— *Demand growth, affordability, and reliability.* The scale and shape of future electricity demand remain uncertain, particularly as AI, data centers, electrification, and industrial growth increase pressure on power systems.

— *Commercial maturity of emerging low-emission technologies.* The ability of emerging technologies (including advanced nuclear, next-generation geothermal, green hydrogen, and e-fuels) to move from pilots to commercially relevant scale depends on several issues, such as cost, performance, permitting, customer offtake, and reliable deployment timelines.

— *Supply chains, materials, and workforce availability.* Scaling energy and sustainability technologies will depend on the availability of critical minerals, transformers, cables, electrolyzers, batteries, specialized equipment, and the workforce needed to build and operate large infrastructure systems.

— *Policy, market design, and social acceptance.* Regional policy divergence, trade friction, permitting rules, carbon border adjustment mechanisms, and evolving electricity market designs could either accelerate investment or increase fragmentation and cost. Public acceptance, community opposition, land use constraints, and trust in emissions measurement systems also may shape deployment.

Big questions about the future

Companies and leaders may want to consider a few questions when pursuing energy and sustainability technologies:

— How can energy systems scale fast enough to meet rising electricity demand from AI, data centers, electrification, and industrial growth while maintaining reliability and affordability?

— As “time to energy” becomes as important as cost, how should large energy users decide when to buy power, build dedicated infrastructure, or partner with utilities, developers, and technology providers?

- Which low-emission technologies are most likely to scale in the near term across power, industry, and transport, and which remain dependent on policy support, infrastructure build-outs, or long-term offtake?

- How can companies and governments accelerate grid modernization, permitting, interconnection, and supply chain readiness without increasing system costs or local opposition?

- How will regional policy, energy security priorities, carbon border adjustment mechanisms, and industrial strategy reshape the economics and geography of the energy transition?

Read more about the future of energy and sustainability technologies

Knowledge center

Insights on Sustainability

Energy and Materials Insights

Insights on Electric Power & Natural Gas

Related reading

Global Energy Perspective 2025

Software: The hidden catalyst for decarbonization

Charging ahead: Electrification equipment trends to watch in 2026

Renewables O&M reimagined: Boosting performance with AI and conventional levers

Solar power: From intermittent sunshine to reliable power

Nuclear power: A renaissance in the making

September 2026

Designed by McKinsey Global Publishing

McKinsey.com