



Technology Report 2025

AI leaders are extending their edge.

Authors and acknowledgments

David Crawford, chairman of Bain & Company's Global Technology, Media, and Telecommunications practice; Anne Hoecker, global leader of the TMT practice; and Dana Aulanier, practice vice president of the TMT practice, prepared this report.

Bain Partners Syed Ali, Laurent-Pierre Baculard, Bharat Bansal, Ann Bosche, Peter Bowen, Steven Breeden, Alessandro Cannarsi, Willy Chang, Matthew Crupi, Arjun Dutt, Greg Fiore, Jonathan Frick, Pascal Gautheron, Adam Haller, Peter Hanbury, Karen Harris, Jonny Holiday, Mark Kovac, Tamara Lewis, David Lipman, Neil Malik, Justin Murphy, Christopher Perry, Bill Radzevych, Paul Renno, Michael Schallehn, Jennifer Smith, Ravi Vijayaraghavan, Jue Wang, and Chuck Whitten; Associate Partners Gabe Dunn, Aaron Lewis, Ryan Petranovich, and Tanvee Rao; and Expert Partners Chris Bell, Purna Doddapaneni, Stephen Hardy, Chris McLaughlin, and Velu Sinha wrote its chapters.

The authors also wish to thank Partners David Cho and Maeghan Rouch; Associate Partner Tatum Quinn; Senior Managers Madeline Ayles, Shiv Bery, Mathieu Boutin-Delisle, Tobi Ogunsanya, Sushil Upadhyayula, and Alyssa Zhou; Consultants Jiachen Chu and Gaurav Srivastava; Senior Associate Consultants Margaret Chiang and Pooja Vettical; Associate Consultant Chimo Mgbeokwere; Expert Manager Alexandra Juegelt; and Practice Directors Tarun Gupta and Alex Smyth for their contributions. They would also like to thank these members of the Bain Capability Network: Senior Manager Eva Gupta; Managers Arpit Jain and Vaishali Sharma; Project Leader Anjali Mishra; Associates Aryan Gupta, Kiran Muchandimath, and Bhanu Singh; Analysts Sahil Garg, Dhiti M, Purnima Pant, and Tanishq Singh; and Senior Operations Analyst Raghav Sharma; as well as Jeff Bauter Engel, Mike Oneal, and David Sims for their editorial support.

This work is based on secondary market research, analysis of financial information available or provided to Bain & Company and a range of interviews with industry participants. Bain & Company has not independently verified any such information provided or available to Bain and makes no representation or warranty, express or implied, that such information is accurate or complete. Projected market and financial information, analyses and conclusions contained herein are based on the information described above and on Bain & Company's judgment, and should not be construed as definitive forecasts or guarantees of future performance or results. The information and analysis herein does not constitute advice of any kind, is not intended to be used for investment purposes, and neither Bain & Company nor any of its subsidiaries or their respective officers, directors, shareholders, employees or agents accept any responsibility or liability with respect to the use of or reliance on any information or analysis contained in this document. This work is copyright Bain & Company and may not be published, transmitted, broadcast, copied, reproduced or reprinted in whole or in part without the explicit written permission of Bain & Company.

Contents

AI Leaders Are Extending Their Edge	2
Value Evolution	3
Will AI Disrupt Tech's Most Valuable Companies?	4
Sovereign Tech, Fragmented World	9
Deals Rise in 2025, But Easy Wins May Be Over	15
Strategic Battlegrounds	21
Will Agentic AI Disrupt SaaS?	22
How Can We Meet AI's Insatiable Demand for Compute Power?	31
Humanoid Robots: From Demos to Deployment	35
Quantum Computing Moves from Theoretical to Inevitable	41
Operational Transformations	47
State of the Art of Agentic AI Transformation	48
AI Is Transforming Productivity, but Sales Remains a New Frontier	55
From Pilots to Payoff: Generative AI in Software Development	62
Building the Foundation for Agentic AI	68

AI Leaders Are Extending Their Edge

Halfway through the decade, it's clear that AI is the defining disrupter of our time. AI's reach is broader than any recent tech wave, reshaping business strategy, politics, trade, defense, and even social justice. Two years ago, we warned that it was already too late to wait and see. By then, leaders were using AI to improve EBITDA by 10% to 25%, while laggards fell further behind. Today, those leaders are compounding their gains and embracing agentic AI. If you're still piloting, you're dangerously behind.

Bain's sixth Technology Report delivers insights on this and other tech trends through the pragmatic lens of the real work we do with clients. The advance of AI agents is the immediate top-line story. At full potential, they'll run complete processes and workflows. Our report explores the implications for software strategy, enterprise deployment, humanoid robotics, and IT architecture, while also looking at broader implications for tech company valuation, trade relations, power production, and talent management.



David Crawford

Chairman of Bain's Global Technology, Media, and Telecommunications practice



Value Evolution



Will AI Disrupt Tech's Most Valuable Companies?	4
Sovereign Tech, Fragmented World	9
Deals Rise in 2025, But Easy Wins May Be Over	15



VALUE EVOLUTION

Will AI Disrupt Tech's Most Valuable Companies?

Hyperscalers and other market leaders have adapted well to technology shifts, but generative AI poses new challenges.

By David Crawford, Matthew Crupi, and Adam Haller

At a Glance

- ▶ Today's tech giants have proven unusually resistant, co-opting disruption through self-reinvention.
- ▶ AI could change that, as it introduces more layers of competition across infrastructure, models, applications, devices, search, and browsers.
- ▶ Incumbents are doubling down, investing heavily in AI to stay ahead, while fast-moving challengers gain traction and funding.
- ▶ Geopolitics, regulation, quantum, and agentic AI add uncertainty, making adaptability critical at every turn.

In Bain's *2024 Technology Report*, we noted the remarkable resilience of today's most valuable technology companies and their ability to co-opt disruption (see the chapter "How Tech Leaders Commercialize

Innovation”). Several tech giants with the largest market caps have maintained their lead for 15 years or more. That’s a big change from earlier eras when disruptive innovation regularly vaulted new companies to the top spots while moving old ones aside. The difference appears to be that today’s leaders are better at adapting to technology shifts due to their ability to self-disrupt and reinvent their businesses.

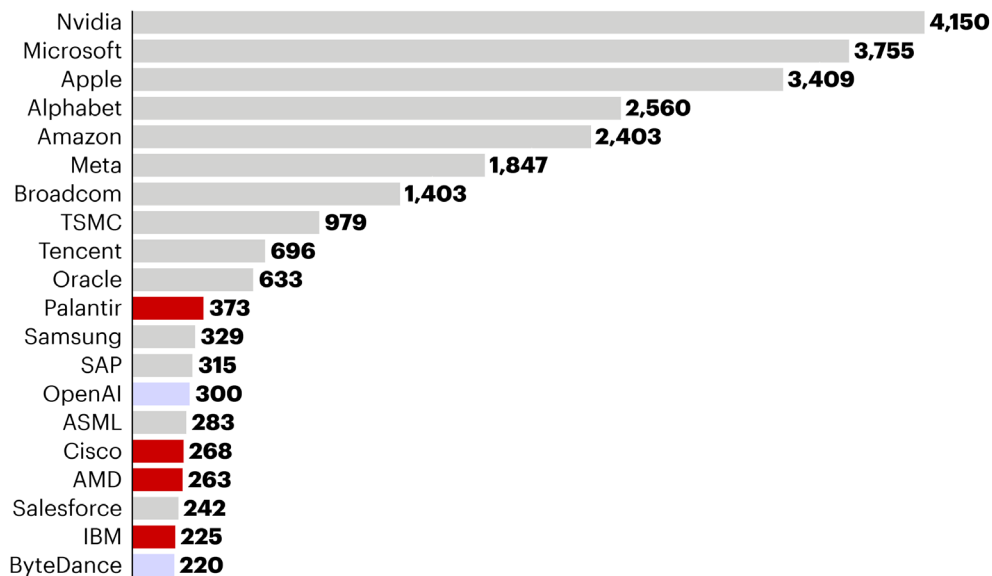
Will AI change that? AI (generative and agentic), with its ability to transform work processes and the unprecedented speed of its adoption, is this decade’s disruption in tech and beyond. Looking at today’s top 20 tech companies gives us insight into whether we expect the existing leaders to remain at the top, or if we may see significant shifts. At first glance, today’s value creation pattern mirrors what we saw during the shift to cloud computing: lots of value created by the incumbents (Amazon, Microsoft, Alphabet, Meta, Nvidia) and a few others, but also a set of vibrant entrants creating winning models, tools, and applications (see *Figure 1*).

But AI’s innovations range further and wider than those of the cloud. Aggressive challengers are gaining attention and funding that could put them in direct competition with today’s most valuable companies. While the leaders that emerged in the cloud era played mostly at the application layer, the AI era will see fierce competition at many levels, including infrastructure, models, and devices.

Figure 1: Market value remains concentrated in the tech incumbents, but new players are entering the top 20

Market valuations of the top 20 tech companies (\$B)

New to top 20 since December 31, 2019 ■ Public companies ■ Private companies



Notes: Private company valuations based on latest round of funding; public company valuations based on market capitalization on September 3, 2025
Sources: S&P Capital IQ; Crunchbase

Early winners in the AI era

The most valuable tech companies have emerged as early winners in the AI era, further concentrating value at the top. The five biggest companies account for more than 70% of the total market value of the top 20, up from 65% last year. Nvidia's market cap is up more than 800% since January 2023, and Microsoft, Amazon, Alphabet, Apple, and Meta are all valued above \$2 trillion. The hyperscalers are investing heavily in AI infrastructure, talent, models, and applications to protect their positions (see Figure 2).

At the same time, a new set of winners is emerging. Privately owned OpenAI is valued at around \$300 billion, which would place it among the top 15 companies if it were public. Anthropic is valued at more than \$60 billion. Other AI companies, including Glean, Anysphere, Mistral, and Figure, also have huge valuations, ranging from \$5 billion to nearly \$40 billion. According to CB Insights, there were about 20 times more tech unicorns (start-ups valued at more than \$1 billion) added in 2024 compared with 2014.

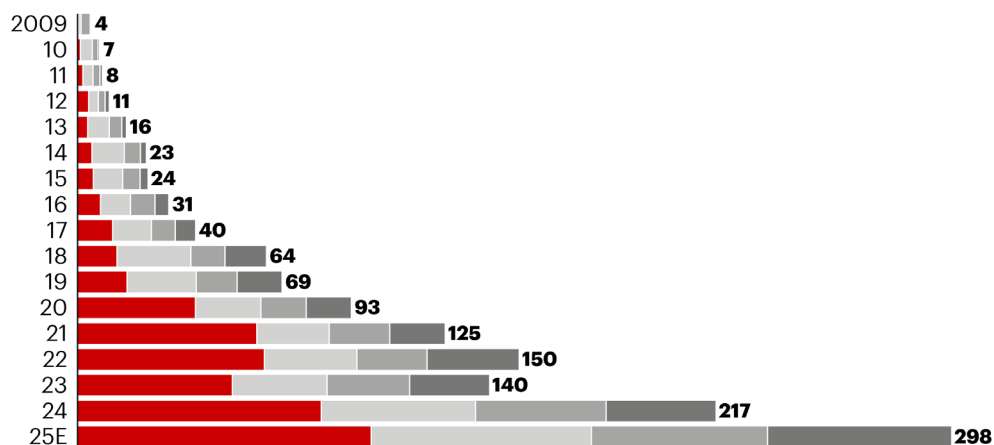
More layers of competition

Tech leaders can monetize their investments in a range of ways because they're competing at every level—owning the infrastructure, building their own models (for example, Google's Gemini or Amazon's Nova), defining platforms, and capturing disproportionate value at the application layer.

Figure 2: Hyperscalers are increasing capital investment in AI at the infrastructure, platform, and application levels

Capex spending (\$B)

■ Amazon ■ Alphabet ■ Microsoft ■ Meta



Notes: Company data based on fiscal years (ending December 31 for Amazon, Alphabet, and Meta; June 30 for Microsoft); forecast values are from earnings calls
Source: Company capex data from S&P Capital IQ and earnings calls

But in the age of AI, that dynamic may be shifting, because new players are encroaching on the leaders at every layer of the AI stack.

- **Infrastructure.** Start-ups like Coreweave offer specialized, high-performance GPU-as-a-service infrastructure optimized for AI and GPU-intensive workloads, often at lower latency and cost than the traditional cloud computing services. Hardware incumbents such as Nvidia are building AI factories—specialized, high-performance data centers built for the demands of the AI era—to strengthen their position, and governments are pushing for sovereign AI capabilities by investing in domestic infrastructure.
- **Models.** New players such as OpenAI, Anthropic, and Mistral have quickly gained ground, in many cases with early investment from the hyperscalers themselves. These start-ups are proving that foundation model innovation isn't limited to big tech.
- **Applications** will still be where we expect to see the most new value and potential disrupters. For example, Anysphere, which was founded in 2022 and is currently valued at \$9 billion, has rapidly gained traction among developers with its AI-powered code editor, Cursor—reinforcing how being best in class can still win against the large tech firms' products.

In addition to these layers, insurgents are competing with tech leaders in new areas:

- **Devices.** AI phones could shake up the established smartphone landscape. Consider the potential impact of a Gemini-enabled Android phone or a device resulting from the collaboration of OpenAI and Jony Ive.
- **Search.** Chatbots like ChatGPT, Claude, or Perplexity AI are already starting to displace search as the entry point to the Internet.
- **Browsers.** AI-based browsers like the ones developed by Perplexity and OpenAI could reignite the browser wars of the 1990s.

Each of these control points provides the incumbents with a wealth of data to improve their offerings, and they may need to self-disrupt as AI changes the playing field. But they'll need to move rapidly to compete: The newcomers are more agile and cost efficient, allowing them to move quickly. Mistral, for example, which is valued at more than \$6 billion and has fewer than 500 employees, can innovate and iterate quickly with its light hierarchy.

Uncertainties shaping the AI market

Beyond direct competition, multiple dynamics are unfolding, adding significant uncertainty to the environment.

- **Agentic AI.** These systems can perform complex workflows, make decisions without human prompts, and adapt dynamically, which could disrupt traditional software paradigms. Legacy players in the application layer (for example, enterprise SaaS) may find their models disrupted by competitors embedding agentic AI that delivers end-to-end outcomes. (For more, read “Will Agentic AI Disrupt SaaS?”)
- **US-China relations.** Geopolitical tensions are fracturing global technology supply chains, particularly in semiconductors and AI hardware. Export controls, investment restrictions, and sanctions are leading tech companies to reconfigure global strategies and confront uncertain access to key markets. As China’s firms accelerate domestic alternatives and US firms shift manufacturing, the resulting decoupling could reshuffle global tech leadership, with regional champions replacing global incumbents.
- **Regulatory pressure.** Governments are ramping up scrutiny on data privacy violations and AI safety at major tech firms. Many are pushing for their own sovereign AI to limit dependency on the US-based leaders, with some hoping to develop their own national champions.
- **Quantum computing.** Quantum represents a foundational shift in computation, threatening to make classical encryption obsolete and redefine problem-solving in materials science, logistics, and AI. As governments and companies race to develop quantum advantage, the first movers could leapfrog current tech giants in sectors from cloud to cybersecurity. The uncertainty around when scalable quantum breakthroughs will occur leaves incumbents exposed to disruption from start-ups or state-backed programs with early breakthroughs.

What does this mean for technology leaders?

- **Incumbent tech leaders.** Keep doing what you have been doing: Disrupt, innovate, attract talent, and acquire and partner for new capabilities. There is more competition at every layer (models, devices, browsers, GPU-as-a-service), and there are also more opportunities to scale.
- **Legacy technology.** Embrace and extend. Act quickly—start by cutting costs. Monitor the horizon and be willing to disrupt your business. Invest in innovation while enhancing capabilities through acquisitions and partnerships. Understand your customers’ needs and make the most of those valuable relationships.
- **Disrupters and start-ups.** Understand the scale of investment required to compete with the hyperscalers. Talent is scarce, so develop a strategy to acquire and retain the right people. Identify new ways to better serve your customers’ needs. Disruption brings opportunity to reshape the basis of competition.

The technology industry has been riding this wave of disruption since OpenAI released its first chatbot. Other industries have been slower to adopt, but AI’s disruption is likely to extend well beyond tech companies. Across sectors, early movers will have an advantage.

VALUE EVOLUTION

Sovereign Tech, Fragmented World

In a post-globalized era of tariffs and decoupling, semiconductors and sovereign AI are realigning global power.

By Anne Hoecker, Karen Harris, Jonathan Frick, and Ravi Vijayaraghavan

At a Glance

- ▶ Tariffs, export controls, and governments' sovereign AI push are accelerating the fragmentation of global tech supply chains and centers of influence.
- ▶ China and the US continue to compete across the full tech stack, from software to hardware.
- ▶ Sovereign AI definitions and strategies vary by country, but the commonality is investment to avoid being left behind.
- ▶ Leading companies will make decisions with optionality, moving boldly where confidence is high and prioritizing flexibility where uncertainty rules.

As geopolitical fragmentation accelerates in this post-globalization era, technology sits squarely at the fault line.

Key cutting-edge domains—semiconductors, AI, communications, quantum computing, and biotechnology—are no longer just catalysts for innovation and economic growth, but conduits for countries' political power, national security, and strategic advantage. Governments are stepping in

more forcefully, actively influencing and directing the flow of capital, talent, and intellectual property. Technological self-reliance (to the extent it's possible) is becoming a more urgent priority for nations worldwide, partly as a means of protecting themselves in case tech-leading countries wield their control over essential technologies—cloud computing, mission-critical business software, defense—as a geopolitical cudgel.

For several years, these dynamics have created constant, unpredictable challenges for technology executives. Two issues are rising fast on boardroom agendas: the near-term effects of tariffs on technology supply chains, and the longer-term business implications of governments' accelerating push for "sovereign" AI.

Two issues are rising fast on boardroom agendas: the near-term effects of tariffs on technology supply chains, and the longer-term business implications of governments' accelerating push for "sovereign" AI.

Navigating this more complex and fundamentally different tech environment will require updated strategies, new bets, and a high tolerance for ambiguity.

The decoupling core: Semiconductors and the electronics supply chain

Semiconductors are the pressure point at the epicenter of today's geopolitical tensions. Aiming to protect its advantage in leading-edge compute and the technologies it powers, the US has steadily tightened export controls on advanced chips, chipmaking tools, electronic design automation software, and high-bandwidth memory chips destined for China.

The restrictions and tariffs implemented by the US on China beginning in 2018 sparked a wave of supply chain diversification. Many companies adopted a "China Plus One" strategy that shifted manufacturing to countries such as Mexico and Vietnam. Now, the second Trump administration has proposed more extensive tariffs on a much wider set of countries.

For many tech executives, this raises significant supply chain questions, given the complex, global nature of the electronics value chain. No longer can the answer simply be China Plus One; a broader set of options must now be considered to ensure supply and cost stability. The only real hedge against unpredictable shocks to the system is continued regionalization or even nationalization; supply chains will become even more dispersed.

China, for its part, is racing toward self-reliance. Since 2019, it has invested more than \$250 billion in semiconductor manufacturing, tripling its domestic production capacity to a projected nearly 3 million wafers per month this year; that's roughly 20% of global capacity. While most of this growth is in mature, lagging-edge semiconductor nodes, China is also making progress in the production of more advanced chips smaller than 28 nanometers, now accounting for around one-fifth of global output of logic chips and a quarter of memory chips (see *Figure 1*).

China's strides are challenging the notion that market leadership in semiconductors is solely defined by the leading edge. While still critical, memories of the semiconductor shortage in 2021 and 2022 remind executives that a lack of lagging-edge chips can keep a company from shipping final products. China's strong position in less advanced chips, which have a larger global supplier base, initially led many to assume that customers would eagerly switch to vendors outside of China. But overcapacity in mature nodes and the perceived flexibility to change suppliers later have kept the status quo intact longer than expected.

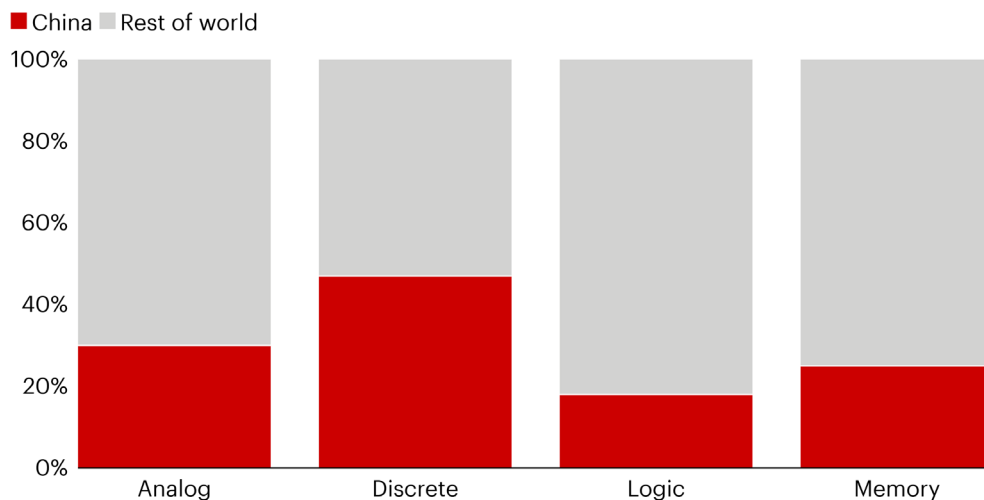
It's yet another signal that decoupling isn't linear, and it's far from over.

Sovereign AI: Leveling the playing field

The concept of sovereign AI has rapidly evolved from theory to geopolitical imperative. Sovereign AI systems are trained on domestic or culturally appropriate data, hosted by nationally or regionally controlled data centers (if possible), and increasingly rely on open-source foundation models developed

Figure 1: China now accounts for a large chunk of global semiconductor manufacturing capacity

Share of semiconductor fab capacity by chip type, 2024



Note: Fab capacity measured by wafer starts per month
Sources: Gartner; Bank of America; SEMI; Bain analysis

domestically, which allows governments and institutions to audit the systems for bias, transparency, and misuse.

This isn't just about privacy or control. It's about data security and aligning AI outputs with national values, regulatory standards, and strategic priorities, all while reducing dependence on foreign tech ecosystems. Sovereign AI capabilities are increasingly seen as a strategic advantage on par with economic and military strength.

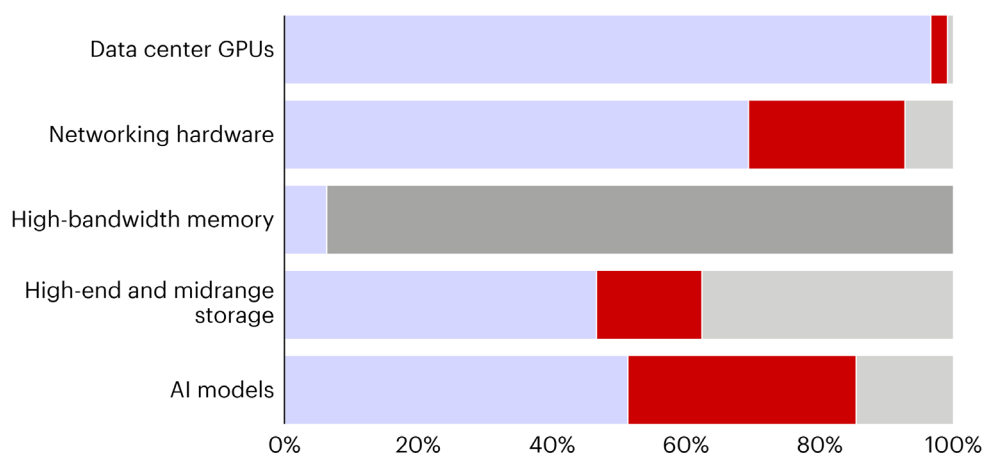
The race between the Chinese and American tech ecosystems is at the forefront of the decoupling movement. Both countries are advancing swiftly. The US leads in high-performance chips and foundation models, while China is expanding its AI capabilities through initiatives including the DeepSeek-R1 model and Huawei's Ascend 910C chip—all developed with minimal US tech. China is also investing heavily in physical AI, such as humanoid robotics. Meanwhile, South Korea dominates in high-bandwidth memory chips, illustrating the complexity of the global ecosystem and the fact that it remains difficult for any one country to become completely self-reliant (*see Figure 2*). While hardware, models, and apps are decoupling, open-source technologies and talent continue to cross-pollinate.

However, sovereign AI is a global priority. The EU's €200 billion InvestAI initiative, launched in February, includes €20 billion to build AI gigafactories—data centers equipped with at least 100,000 graphics

Figure 2: Chinese companies offer options in many of the critical AI tech components

Market share by company headquarters location

■ US ■ China ■ South Korea ■ Rest of world



Notes: Market share based on 2024 revenue, except for the AI models category, which is based on the number of large-scale models released since OpenAI's ChatGPT 3.5 in November 2022; large-scale models defined as those released after November 2022 that are estimated by EpochAI to have been trained on more than 10^{23} floating point operations (FLOP); GPUs stands for graphics processing units

Sources: New Street Research; OpenRouter Leaderboard; Bank of America; RampAI; IDC; Gartner; company websites; news reports; EpochAI; Bain analysis

processing units (GPUs) each. In a related project, Germany-based Deutsche Telekom has partnered with Nvidia to build an industrial AI cloud for European manufacturers. Saudi Arabia's new AI firm, Humain, plans to build domestic data centers with a combined capacity of 500 megawatts. It's starting with a 50-megawatt data center housing 18,000 Nvidia GPUs, slated to launch in 2026. Humain also aims to build one of the most powerful multimodal Arabic large language models.

AI goals vary. In China, it's about end-to-end control. In Europe, it's more about regulatory alignment and data localization. In the Middle East, it's participation in the global ecosystem. Practicality trumps purity: For most countries, it's simply not feasible to achieve full-stack independence, at least not today, given the realities of where semiconductor fabs are clustered and which countries control the best AI models.

That divergence will complicate everything for tech companies. As AI becomes embedded in business operations—from customer engagement to supply chain management—multinational firms will need to localize not just compliance, but technology architecture. A single AI workflow may need to be retooled for different markets, with varying models, training data, data usage practices, and infrastructure requirements.

And global AI standards? Unlikely. From content censorship to data labeling to acceptable uses, definitions of “responsible AI” differ widely and likely won't converge. AI systems are becoming more like national or regional products, shaped by the political and cultural norms in which they're developed.

Strategic implications: Rethinking how and where to compete

Executives are starting to recognize this isn't a passing phase. It's a new world order with profound consequences for everything a business does. To steer their organizations effectively through this new era, leading tech companies are focusing on four key principles.

To steer their organizations effectively through this new era, leading tech companies are focusing on four key principles.

Think in operating models, not just product lines. Some countries will build their own AI capacity or treat it as a strategic lever. Others will simply buy from the cheapest available source. That creates a patchwork of environments in which companies must tailor how they build, deploy, and monetize technology. Will local laws demand new infrastructure, retrained models, or local data partnerships? If so, what trade-offs are acceptable?

Don't assume the global tech race is over. It's very much still on and has a long way to go, as tariffs and export controls haven't slowed China as much as expected. The journey will be volatile and bumpy, with

big unknowns surrounding possible trade deals and the knock-on effects of China's accelerated investments in technology manufacturing capacity. Leading companies will closely follow developments, particularly in generative AI and humanoid robotics.

Don't mistake relocation for resilience. For multinational technology companies, moving production out of China is a start, but it's no longer enough. Supply chains must be regionalized, flexible, and built for continued political volatility. Establishing a stronger presence in the most important end markets is becoming more crucial, particularly for semiconductors and other strategic, capital-intensive sectors.

Make decisions with optionality. Executives won't get every bet right. Where confidence is high, move boldly. Where the future is murky, prioritize flexibility. For some companies, that might mean setting up neutral-region tech hubs. (Will Dubai become the next Singapore?) For others, it may mean delaying or skipping certain markets altogether because they're too costly or complex to serve.



VALUE EVOLUTION

Deals Rise in 2025, But Easy Wins May Be Over

Riding growth used to be easy in software. Now investors have to go out and find it efficiently.

By David Lipman, Christopher Perry, Jennifer Smith, and Jonny Holliday

At a Glance

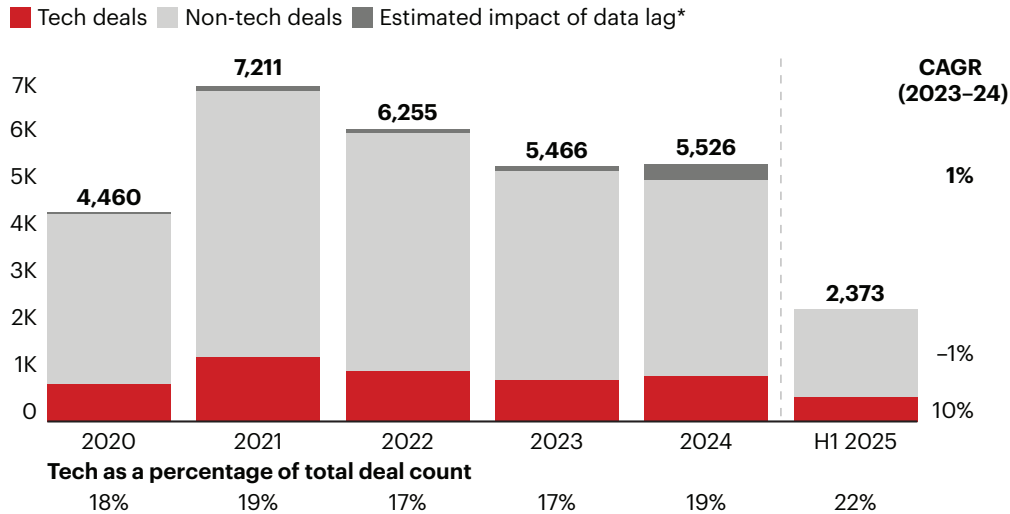
- ▶ Despite a difficult market, technology deals increased their share of all buyouts in the first half of 2025.
- ▶ Yet evidence continues to mount that the days of easy wins in software are receding into the rearview mirror.
- ▶ Investors generating top-tier results in the year ahead will be those working harder to find new sources of growth and those executing on those opportunities more efficiently.

After a fast start to 2025 (extending a strong finish in 2024), technology deal-making has not been immune to the tariff-related uncertainties and geopolitical tensions that have slowed the broader deal market since April.

But even as some deal processes have stretched out, tech investors remain upbeat going into the year's second half. The sector has held up better than most through the first half of 2025, lifting its share of all deals to 22% as of July, compared to 19% at the end of 2024 (see *Figure 1*).

Figure 1: Technology's share of North American private equity deals rose to 22% in the first half of 2025

Cross-sector PE deal volume by deal close date, North America



Notes: 1H 2025 represents deals closed January–July 2025; includes all deals from North America; excludes real estate; SPS data determines deal timing based on deal close date (not announced date); all deal types are included (buyouts, recapitalizations, add-ons, and minority deals); *Impact of SPS data lag estimated based on prior year actuals to account for delay in private market deal data appearing in SPS (excluded from 1H 2025 data); 2020–2024 as of December 31, 2024; 1H 2025 as of July 31, 2025
Source: SPS

Discussions earlier this year with 30-plus private equity tech specialists indicated they remain confident that technology—software in particular—is less exposed to tariff-related impacts than many sectors. And the pressure to move assets is only building. Not since 2012 has the backlog of technology companies held longer than four years been higher, and dry powder in tech-focused funds was sitting at \$476 billion globally at the end of 2024. There's a growing recognition that raising the next fund will likely depend on freeing up that capital sooner rather than later.

The shifting value equation

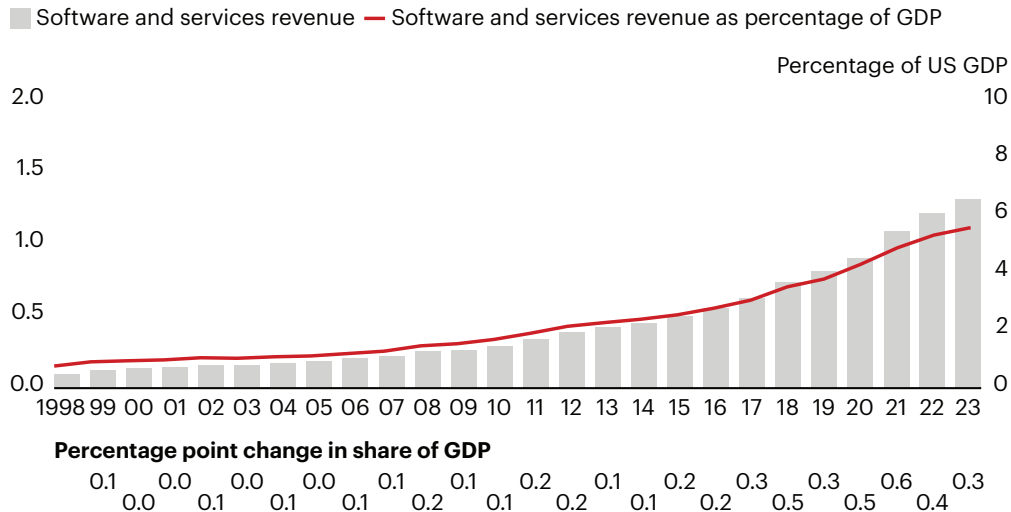
Amid these shortish-term market dynamics, however, tech investors face a larger, more complex question: As software markets mature, where will the next phase of growth and returns come from?

Many indicators suggest the nature of software investing is undergoing a fundamental shift. The “easy money” era of picking up a promising SaaS company and watching revenues (and multiples) explode is drawing to a close. Future returns will increasingly depend on finding new sources of revenue growth and expanding margins through operational excellence.

Revenue growth in the software sector has outpaced the broader market for so long it's almost taken for granted. But as penetration curves in many product areas begin to flatten, overall software spending is

Figure 2: Software spending continues to outpace overall GDP growth and expand its share of total output, but its relative growth is slowing

Total sector revenue, US (\$ trillions)



Note: Shows base year 2019 inflation-adjusted dollar values; includes exports
Sources: S&P Global; US Federal Reserve Bank

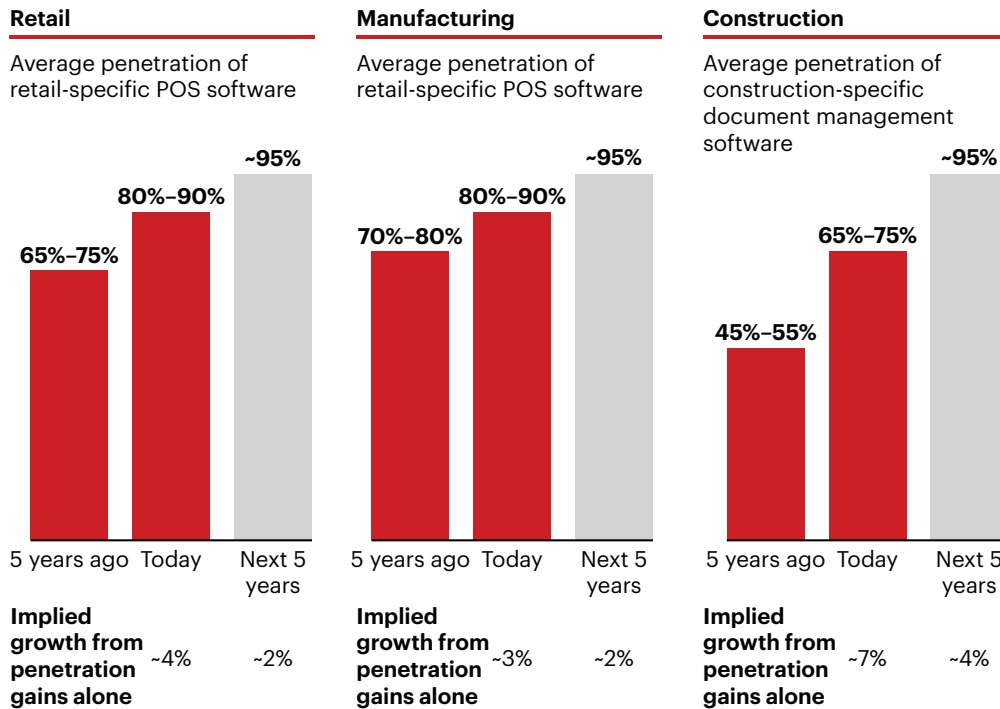
easing off, too. Although software continues to expand its share of US gross domestic product (GDP), its relative growth is starting to ebb (see Figure 2).

The amount of white space available varies by sector. But after adopting software at a breakneck pace for years, mature segments like retail and manufacturing have deeply embedded software solutions for nearly every workflow—not just the standard enterprise resource planning (ERP) or customer relationship (CRM) software but also specialty solutions for specific functions like contract management and employee engagement (see Figure 3). Relative digital laggard sectors like construction may have more room to run, but overall, growth derived by a company simply showing up in an underserved market will be harder to come by in the years ahead.

This is not to say that the outlook for software is dimming. We know for sure that analysts and pundits have consistently underestimated software growth as the industry continues to innovate and find new use cases and workflows to automate. The digital revolution in the workplace is ongoing and will continue to fuel spending. At the same time, however, tapping into that growth and turning it into investment returns will require different capabilities and approaches to value creation.

The easiest way to see this is to look at how PE funds have generated returns in the past (see Figure 4). Revenue growth has contributed 53% of the total value creation since 2010. Of that, the majority has come from traditional means: penetrating new markets and customers, upselling existing accounts, and eventually raising prices. Almost as much—43%—has come from multiple expansion as booming deal

Figure 3: Software penetration is topping out in major sectors like retail and manufacturing, though some sectors like construction offer more room to grow



Source: Historical Bain due-diligence data

markets pushed acquisition prices steadily higher. Margin improvement, meanwhile, has contributed just 4% of value.

In the current interest-rate environment, the same level of multiple expansion is unlikely, and those once-reliable revenue producers are losing steam. In addition to flattening penetration curves, traditional upsell/cross-sell motions are less effective and like-for-like pricing increases are harder to sustain amid pressure on customer budgets and heavy competition. That means investors will have to not only work harder to find each dollar of top-line growth but also build efficient new operations capable of executing on those opportunities profitably.

Top-tier firms seeking new revenue sources are focusing on several areas at once.

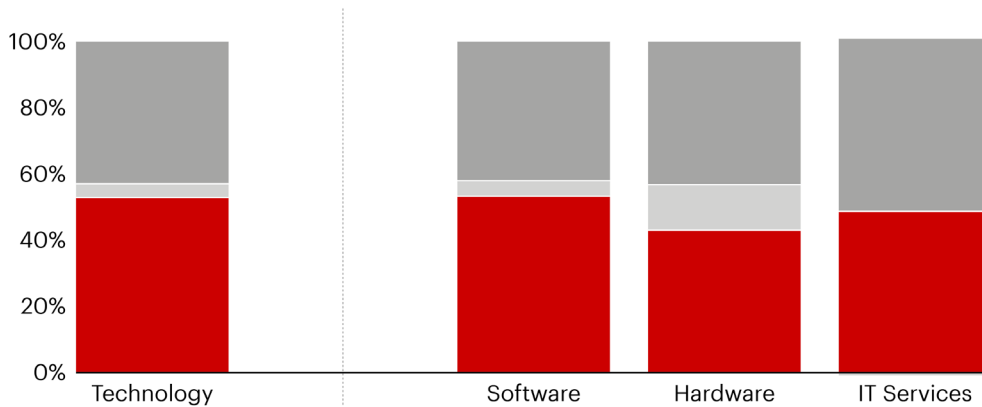
- Displacing competitors.** This is an imperative in any maturing industry. As penetration curves flatten, the companies maintaining growth rates are those adept at taking share from weaker competitors and incumbents: Think fighting for gray space, not opening new white space. This requires significant investments in new go-to-market capabilities to better quantify and segment market opportunities. These are motions many fast-growing software companies haven't had to worry about until now.

Figure 4: Tech has outperformed most other sectors for private investors largely because of revenue growth and multiple expansion—not margin improvement

Median indexed value creation

Buyout and growth deals in North America, entry years 2010–24

■ Revenue Growth ■ Margin Expansion ■ Multiple Expansion



Notes: All calculations in USD; deal universe includes fully and partially realized North American buyout and growth deals with initial investments made between 2010 and 2024; excludes real estate
Source: DealEdge powered by CEPRES data

- **Tapping AI.** It's no secret that artificial intelligence is changing the game in many product categories. It offers opportunities to both transform legacy products with new functionality and develop new offerings and use cases. Indeed, continued upsell and cross-sell will increasingly depend on AI-enabled innovation. That requires targeted investment in R&D and rapid testing aimed at aligning the success of the software provider with the success of the customer.
- **Deploying modern pricing models.** Waiting for a product to stick and then tapping incremental, seat-based price increases is less and less effective. Innovative commercial organizations are gaining traction with outcome- or value-based pricing models founded on the measurable value they deliver to the customer, not the number of users, provider costs, or other traditional factors.
- **Expanding geographically.** Even if a once-hot sector is slowing in the US, ample opportunity may exist in another region. But successful expansion into new geos requires a disciplined strategy often based on M&A capabilities (including successful integration and change management), new market development, and localization of products to adjust to different languages and workflow requirements.
- **Building in payments capability and/or monetizing data.** Especially in vertical software, or products aimed at a specific industry, customers are increasingly looking for solutions with integrated payments capabilities—built-in systems that smooth transactions and often capture valuable data that standard point-of-sale (POS) systems can't.

Tactics like these give software investors the means to maintain their growth focus. But generating top-tier returns will increasingly require concerted efforts to boost margins.

The traditional answer here, of course, is cost takeout—coming in and rationalizing a company’s cost structure, too often with a heavy hand that can end up being counterproductive. A more evolved approach to margin improvement is holistic: All costs get a thorough going over. But the real value comes from developing an improved strategy based on the right top-line considerations and matching it to improvements or investments in go-to-market capabilities, better cost-of-goods management, enhanced processes, and a more efficient, targeted R&D function.

AI can be an important tool to increase efficiency in all these areas as can a much tighter definition of who your most important customers are and how to serve them. Commercial excellence is eminently measurable, but so is R&D (surprise!) with tools like Faros AI or Jellyfish. Top-tier performers no longer treat the research function as a black box; tracking engineering productivity is increasingly critical.

For fund managers, a few key questions can help focus the new imperative for each portfolio company:

- Is our go-to-market approach the right one for future opportunities, or is it still calibrated to yesterday’s market?
- Are R&D and product development focused on the innovations that will stand out with today’s customers and power the next phase of growth?
- What are the specific future risks and opportunities AI presents for this business? How should we prepare right now?
- How can we use M&A to supercharge our strategy? Which opportunities are likely to be accessible?
- Is our current talent matched appropriately to the value creation strategy we’ll need in the future?

The opportunity in tech is no less vibrant than it ever was. But the best way of capitalizing on it is evolving rapidly. What worked so well in the past is unlikely to generate the same level of return in the future. It’s time to boost your value-creation game.



Strategic Battlegrounds

Will Agentic AI Disrupt SaaS?	22
How Can We Meet AI's Insatiable Demand for Compute Power? ...	31
Humanoid Robots: From Demos to Deployment.	35
Quantum Computing Moves from Theoretical to Inevitable	41



STRATEGIC BATTLEGROUND

Will Agentic AI Disrupt SaaS?

Disruption is mandatory. Obsolescence is optional.

By David Crawford, Chris McLaughlin, Purna Doddapaneni, and Greg Fiore

At a Glance

- ▶ Generative and agentic AI are disrupting software as a service (SaaS) by automating tasks and replicating workflows.
- ▶ SaaS leaders can manage the risks by identifying where AI can enhance their offerings and where it might replace them.
- ▶ To stay ahead, they must own the data, lead on standards, and price for outcomes, not log-ons, in an AI-first world.
- ▶ With the right playbook that includes deep AI integration, strong data moats, and leadership on standards, incumbents can shape, not just survive, the next wave of SaaS.

When software as a service (SaaS) first emerged 25 years ago, it revolutionized software by moving it to the cloud and speeding up feature delivery. Now, a fresh discontinuity is at hand. Generative and agentic AI—tools that can reason, decide, and act—are already:

- drafting code in Cursor’s AI code editor;
- handling support tickets in ServiceNow;
- preparing journal entries in Workday Financial Management; and
- writing ad copy in Adobe’s Experience Cloud.

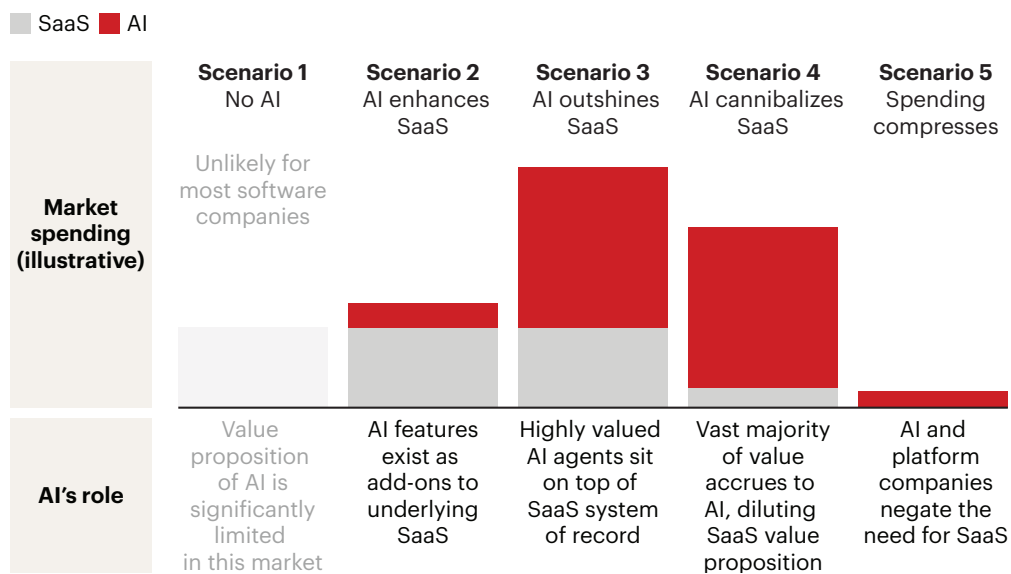
These aren't experimental one-offs. The cost curve trajectory of foundation models is accelerating downward even as accuracy improves. OpenAI's latest frontier reasoning model (o3) dropped 80% in just two months. In three years, any routine, rules-based digital task could move from "human plus app" to "AI agent plus application programming interface (API)."

SaaS providers know this strategic problem is urgent, but it's also addressable. Product leaders must answer several strategic questions:

- Which user workflows can AI and agents automate? To what extent?
- Which SaaS workflows can be handled by AI and agents?
- Where will AI increase the size of the software market, and where will it cannibalize?
- Where are incumbents and new entrants favored?
- What investment priorities will shape the outcome in their favor?

In our work with clients, we see five broad possibilities for any given SaaS workflow: No AI, AI enhances SaaS, spending compresses, AI outshines SaaS, and AI cannibalizes SaaS (see Figure 1).

Figure 1: Five broad scenarios illustrate how AI will affect software as a service (SaaS)



Source: Bain analysis

Potential for AI to automate tasks and penetrate workflows

To navigate these risks, executives should evaluate workflows according to two independent characteristics: the potential for AI to automate SaaS user tasks and the potential for AI to penetrate SaaS workflows. Mapping workflows against these characteristics can help identify value at risk and plans to capture it before it migrates elsewhere.

Six indicators can help companies understand the degree to which AI and agents can replace or further assist users: task structure and repetition, risk of error, contextual knowledge dependency, data availability and structure, process variability and exceptions, and human workflow and user interface dependency.

Where these indicators suggest a high potential to automate SaaS user activity, the AI disruption tends to expand the market, offering significant opportunity to capture top-line growth (see *Figure 2*).

Six additional indicators help identify which SaaS workflows are most easily replicated (and potentially captured) by AI and agents: external observability, industry standardization, proprietary data depth, switching and network friction, regulatory/certification barriers, and agent protocol maturity. The higher a workflow's AI penetration potential, the easier it is for a clever AI wrapper to siphon usage and margin (see *Figure 3*).

Four strategic scenarios

By plotting products and workflows, SaaS providers can estimate which scenario from Figure 1 most resembles the impact of AI. This maps out across four strategic scenarios (see *Figure 4*).

- **AI enhances SaaS (low user automation, low AI penetration):** These workflows are **core strongholds** for incumbents. These still rely on human judgment, and rivals struggle to mimic the logic behind them. Think of Procore's project cost accounting or Medidata's clinical-trial randomization—both require deep domain knowledge, strict oversight, and regulated data flows. Incumbents should use AI to boost productivity while protecting the unique data that sets them apart. Price the time savings at a premium.
- **Spending compresses (low user automation, high AI penetration):** These workflows are like **open doors** that expose incumbents to new risks. People still play a role, but third-party agents can hook into exposed APIs and siphon value. Examples include HubSpot's list building or the task boards on Monday.com. To defend and salvage value and customer influence, incumbents must launch their own agents fast, deepen partner integrations to raise switching costs, and limit access to critical end points.
- **AI outshines SaaS (high user automation, low AI penetration):** These workflows will be an incumbent's growth **gold mines**. Here, companies hold exclusive data and rules, giving them a head start on full automation. Cursor's AI code editor and Guidewire's claims adjudication are good examples. Leaders should build solutions with end-to-end agents, shift pricing from seat-based to outcome-based, and train sales teams to sell business results, not just features.

Technology Report 2025

Figure 2: Key indicators to assess AI's potential to automate software-as-a-service (SaaS) user tasks

Indicators of potential for user task automation	Why it matters	Examples in which AI replaces humans	Examples in which AI assists humans
Task structure and repetition	Highly repetitive sequences can be templated; fragmented, ad-hoc tasks cannot	High-volume invoice coding with Tipalti	Project-based billing with Procore
Risk of error	The greater the regulatory, safety, or financial liability, the harder it is to cede human oversight	A/B-testing subject lines with HubSpot	Oncology drug-trial randomization with Medidata
Contextual knowledge dependency	If critical facts live outside documented data, humans stay involved	Standard leave-balance checks with BambooHR	Executive compensation design with Workday
Data availability and structure	Agents need well-labeled, machine-readable inputs at run-time—fine-tuned model or not	Git commit histories with GitHub	Supplier emails stored as images with legacy enterprise resource programming systems
Process variability and exceptions	Straight-through flows are computer friendly; high exception rates force escalation to human	Auto-renewing SaaS subscriptions with Zuora	One-off revenue recognition with SAP S/4HANA
Human workflow and user interface dependency	When users must inspect and adjust each step, a hybrid human-agent model persists	Batch payroll calculation with ADP	Field-service dispatch with ServiceTitan in which techs juggle photos, notes, and spare parts

Source: Bain analysis

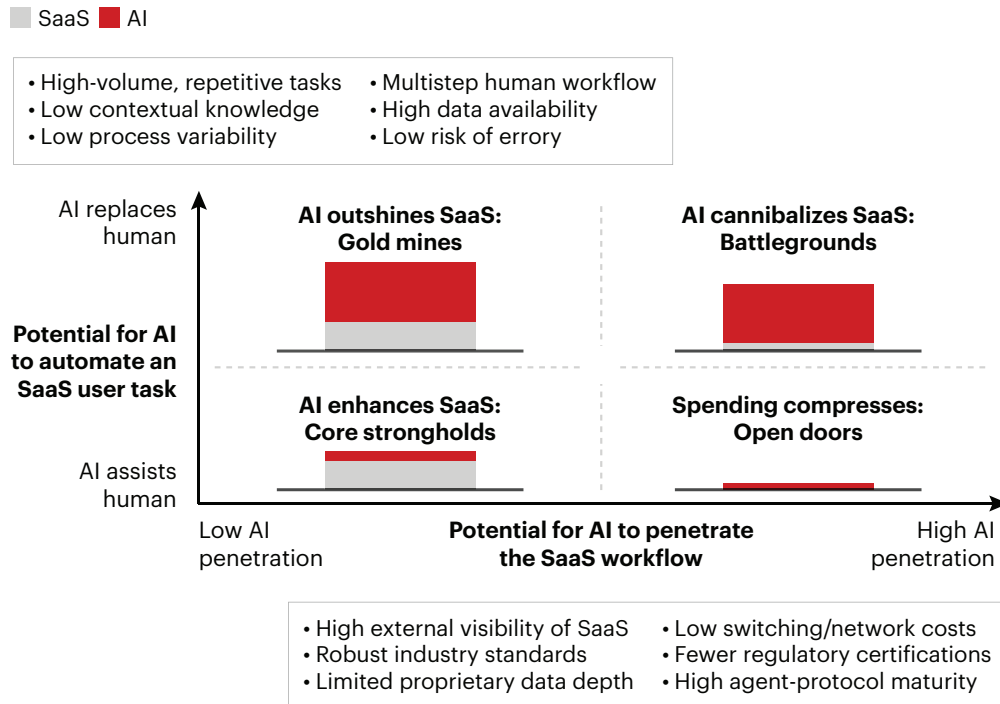
Technology Report 2025

Figure 3: Key indicators of AI's potential to penetrate a software-as-a-service (SaaS) workflow

Indicators of potential for SaaS workflow penetration	Why it matters	Examples of high AI penetration	Examples of low AI penetration
External observability	Process logic is visible through open application programming interfaces (API) or exported reports	Intercom's REST end points for every ticket event	Epic's controlled HL7 interfaces
Industry standardization	Common schemas ease replication of SaaS activities and roles	E-commerce order objects in Shopify's GraphQL	Custom tax codes in Vertex
Proprietary data depth	Large bodies of unique history or sensor data resist AI penetration	Generic customer relationship management deals in many small and medium-sized business tools	ServiceTitan's millions of field service jobs
Switching and network friction	Low setup cost, minor partner ecosystem, and installed base of connectors	Standalone survey forms in Typeform	Salesforce's AppExchange universe
Regulatory/certification barriers	Few mandatory audits or licenses	Internal knowledge-base search	HIPAA-compliant patient scheduling
Agent protocol maturity	Widely adopted syntactic standards, such as Anthropic's Model Context Protocol (MCP), and emerging semantic ontologies make orchestration trivial	A Shopify-storefront agent that drafts and publishes product descriptions by chaining Shopify's Admin API with a large language model through MCP-style calls	Custom actuarial modeling in Milliman's Integrate

Source: Bain analysis

Figure 4: Mapping products and workflows into four strategic scenarios helps software-as-a-service (SaaS) executives set offensive and defensive priorities



Source: Bain analysis

- **AI cannibalizes SaaS (high user automation, high AI penetration):** These workflows will be **battlegrounds**. Incumbents should have the advantage, but to keep it they will need to proactively replace SaaS activity with AI. Incumbents that fail to do this risk disruption, obsolescence, and losing out to entrants. Tasks such as Intercom’s Tier 1 support, Tipalti’s invoice processing, or ADP’s time-entry approvals are easy to automate—and just as easy for others to copy. Winners will be the organizations that scale agent orchestration best. Most companies must pick a lane: Either become the neutral agent platform or supply the unique data that powers it. Only a few giants (Salesforce, for example) can realistically do both.

Bottlenecks and the need for common syntax

SaaS unbundled suites of apps and services. Agentic AI is now rebundling control on a three-layer stack (bottom to top): systems of record, agent operating systems, and outcome interfaces.

- **Systems of record** sit at the base of the stack—the source of truth. They store core business data, manage who can access it, and enforce rules that keep everything consistent—from approvals to compliance checks. Their edge lies in unique data structures, long histories of activity, and built-in regulatory logic that would be costly for others to replicate.

- **Agent operating systems** sit in the middle tier, orchestrating the actual work. They plan tasks, remember context, and invoke the appropriate tools for users and agents. Early versions include Microsoft’s Azure AI Foundry, Google’s Vertex AI Agent Builder, and Amazon Bedrock Agents. Today’s advantage hinges on GPU scarcity, proprietary AI models, and tightly integrated toolchains that speed up deployment.
- **Outcome interfaces** form the top layer. These translate plain language requests—such as “close my books” or “replace pump 17”—into agent actions and share updates back through tools such as Teams, Slack, or custom mobile apps. Their power comes from being woven into daily routines and delivering a trusted, intuitive user experience.

As models have become more powerful, communication across layers and across vendors has become the bottleneck. Vendors have stepped into this void to improve syntax. Anthropic’s Model Context Protocol (MCP) and Google’s Agent2Agent (A2A) standardize the way agents package tool calls, security tokens, and results as they move among layers. But they don’t provide a shared vocabulary (that would define terms such as *invoice*), policy, or work order—nor do they show how those concepts map to APIs, tables, and approval gates.

As models have become more powerful, communication across layers and across vendors has become the bottleneck. Vendors have stepped into this void to improve syntax.

The emergence of these standards (MCP and A2A) has shown strong network-effect dynamics—for instance, lightning-fast tipping points, winner takes most. We expect that the standard for this semantic layer will be similar. In other words, the first semantic layer that creates an industry-wide standard to enable an *invoice.bot* to talk to a *payment.bot*, for example, will reshape the AI ecosystem and direct a large next wave of value.

SaaS incumbents are well-positioned to lead—if they move fast. This will require high-stakes strategic bets—such as selective open-sourcing or a shift in the monetization model—and will yield a unique, durable industry influence position. Win here, and your platform becomes a marketplace, earning revenue even when someone else’s agent takes the action. Miss it, and you risk exposing your IP and becoming a silent back end while the semantic gatekeepers harvest the margin (see *Figure 5*).

Strategic priorities for SaaS leaders

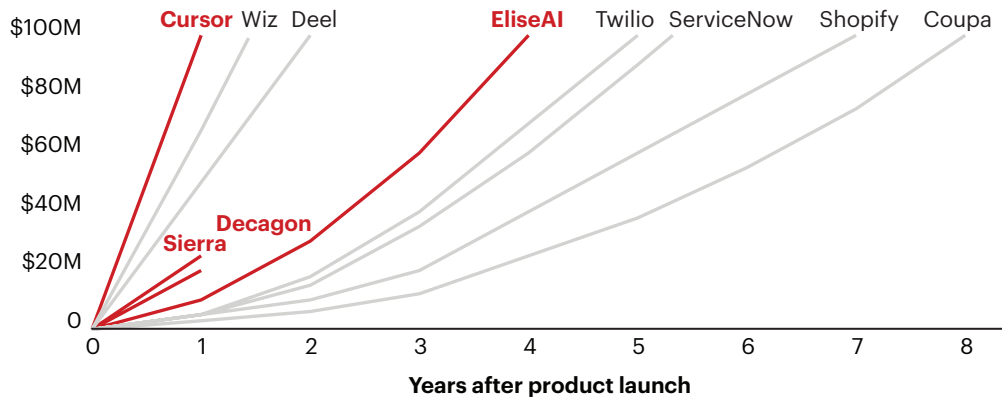
Will AI and agents disrupt SaaS? Yes. In some cases, that disruption will grow the market; in others, it will commoditize the market. In some cases, the disruption will favor incumbents; in other cases, it will

Figure 5: AI-native company growth outpaces comparable software-as-a-service (SaaS) providers

Years to \$100 million annual recurring revenue (ARR) after product launch

— AI agent — SaaS

ARR (in millions of US dollars)



Source: Bain analysis

favor new entrants. Disruption is mandatory, but obsolescence is optional. What can SaaS executives do to navigate this opportunity?

- Make AI central to your roadmap.** Look for the key jobs that your software helps users accomplish, and deploy AI to automate and speed them up. Take a customer-centric view: Identify repeatable tasks that smart agents can handle, and implement those before your customers look elsewhere. This could mean integrating off-the-shelf models or training your own model with your data. Turn your product into a “do it for me” experience, and help customers see the ROI. Embed AI deeply, stay at the center of the workflow, and deliver more value.
- Turn unique data into your edge.** Your data is your moat. While models such as GPT-4o are everywhere, the real value lies in the proprietary data you own—usage patterns, domain-specific content, and transaction history, for example. Double down on capturing and using this data to deliver results no outsider can match. And protect it. If you connect with other AI platforms, make sure your terms stop them from learning from your data and cutting you out. The aim: Become the best source of truth for a key process or data set. Workday’s positioning as a secure hub for managing both human and AI workflows is a good model.
- Shape investment and competitive plans across the four strategic scenarios:** core strongholds in which AI enhances SaaS, open doors in which spending compresses, gold mines in which AI outshines SaaS, and battlegrounds in which AI cannibalizes SaaS.

- **Decide your strategy for addressing the semantic gap for your industry.**
 - **Get your house in order:** Standardize how you define key objects within your own platform. This sets the foundation to either join or lead the next generation of industry-wide agent platforms.
 - **Open source early, selectively:** Publish schemas in which you already lead—as ServiceTitan and Guidewire do. Doing nothing cedes definition power to others; giving away too much puts competitors on a fast track. In standards wars, early movers with a practical solution often win.
 - **Make it hard to copy:** Build unique constraints—for instance, approval flows, state transitions, and compliance rules—right into your data model. Any external agent should have to validate through your system of record.
 - **Rally the ecosystem:** Standards stick when vendors, customers, and cloud platforms align. Bring the group together, shape the agenda, and offer real code to become the default leader.
- **Rethink pricing for an AI-first world.** Seat-based pricing may not fit when AI is doing the work. If an agent replaces a human task, customers will expect to pay based on outcomes, not log-ons. Start experimenting with pricing tied to results: tasks completed, tickets resolved, AI outputs generated. Leaders, such as Intercom and Salesforce, are already shifting in this direction. The fundamental shift is to stop charging for access and start charging for work done. Stay flexible as you learn what your customers value most.
- **Build AI fluency across the business.** AI needs to be a core capability, not a side project. That means helping your teams understand what AI can and can't do, hiring or training the right talent (from machine learning engineers to prompt designers), and building a culture that's excited about innovation. Everyone—from product to sales—should be able to explain how your AI features work and what value they deliver. And that fluency should extend to customers, too. Help them understand and get the most out of what you've built. In the end, your organization should be as comfortable using AI as a new hire is with a browser.

Write the next chapter before your competitors do

AI is disrupting SaaS, creating upsides and downsides. By tailoring investments and strategic plans to each workflow's strategic context, anchoring to the new platform layers, and investing in semantic gaps that affect your developers, today's leaders can shape the future—not chase it.



STRATEGIC BATTLEGROUND

How Can We Meet AI's Insatiable Demand for Compute Power?

Technological innovation, new revenue, and public support may be needed to fund and supply enough electricity.

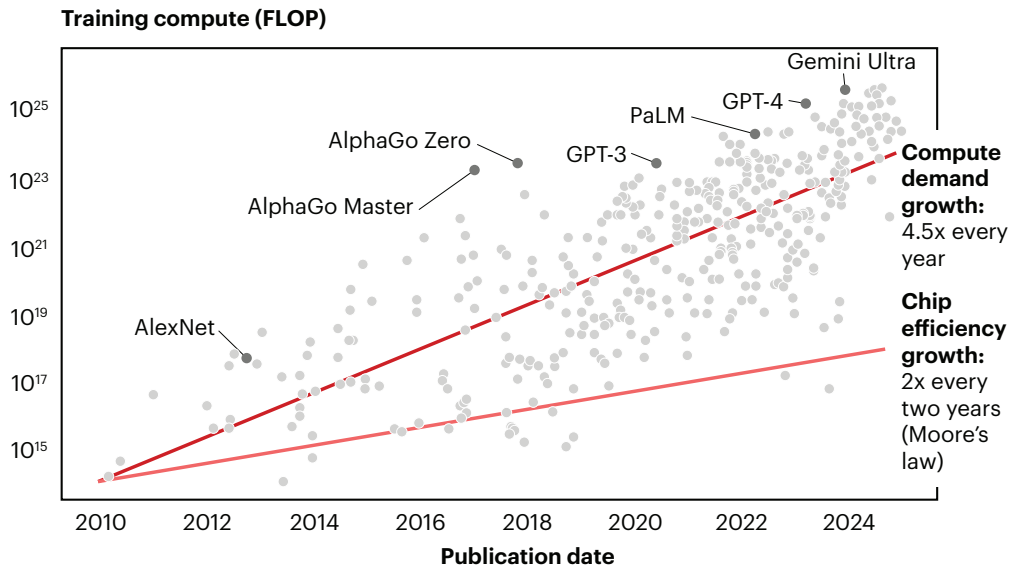
By David Crawford, Michael Schallehn, Paul Renno, Peter Hanbury, and Alessandro Cannarsi

At a Glance

- ▶ AI's computational needs are growing more than twice as fast as Moore's law, pushing toward 100 gigawatts of new demand in the US by 2030.
- ▶ Meeting this demand could require \$500 billion in annual spending on new data centers.
- ▶ But even an aggressive reinvestment of IT and AI savings leaves a big gap.
- ▶ Technological and algorithmic breakthroughs could help, but supply chain shortages or insufficient power supply could also thwart progress.

Of all the breathtaking observations about AI, few exceed this one: The growth rate for AI's compute demand is more than twice the rate of Moore's law (*see Figure 1*).

For decades, Moore's law—that is, the number of transistors on an integrated circuit doubles about every two years—has been the unrivaled measure of technological progress. If you apply its rate of progress to other fields of endeavor, the results are equally mind-blowing and amusing. For example, an automobile

Figure 1: Compute demand grows twice as fast as chip efficiency

with Moore's law applied would travel at 300,000 miles per hour, achieve 2 million miles per gallon of gas, and cost four cents.

Now, AI's compute demand—that is, the number of computations that must be performed to support evolving models—has grown at twice that rate over the past decade. With continued growth of these models and more adoption of AI by enterprises, Bain's analysis suggests that the total global compute requirements could reach 200 gigawatts by 2030. In the US alone, total demand could reach 100 gigawatts by that time, which would increase new electricity demand on a grid that has seen relatively flat load growth for the past 20 years.

What are the implications of this for technology executives charged with allocating capital and managing investments? If you bet on continued growth and add lots of power generation or compute capacity while the trend slows down, you could be stuck with catastrophic unutilized power and compute capacity. If you bet that the trend will slow while it turns out to be durable, you may find yourself with insufficient capacity to capture a wave of growth and market share.

What could change the trajectory?

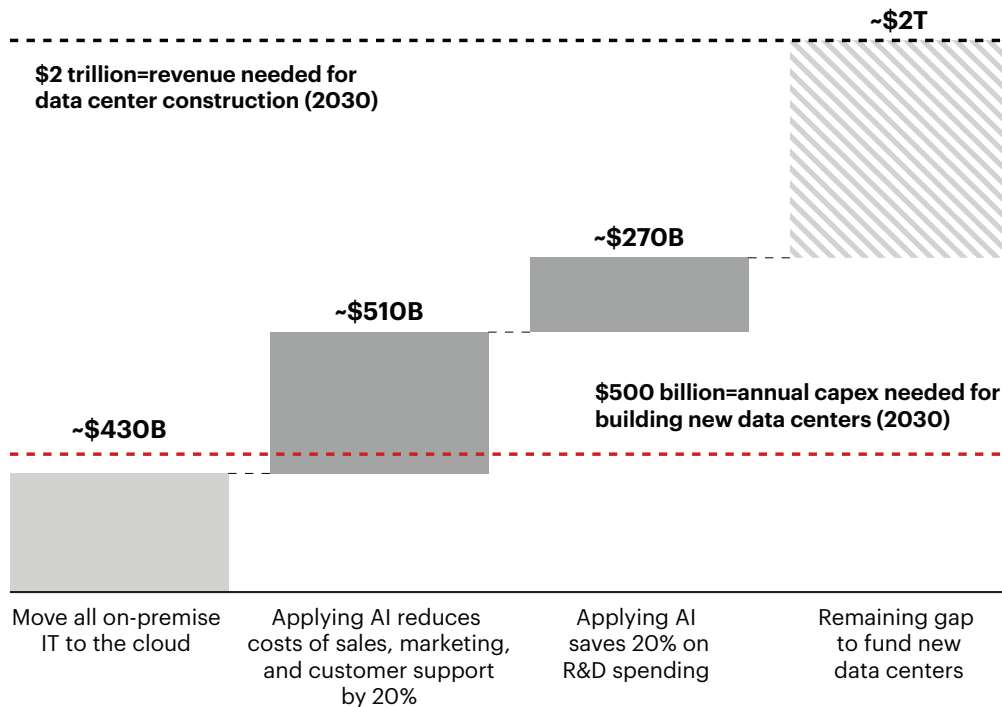
As companies form their strategic plans and determine investments, they should be considering four critical factors that could either impede or accelerate this growth: unaffordable economics, better graphics, technological breakthroughs, and supply chain shortages.

The economics become unaffordable. Bain's research suggests that building the data centers with the computing power needed to meet that anticipated demand would require about \$500 billion of capital investment each year, a staggering sum that far exceeds any anticipated or imagined government subsidies. This suggests that the private sector would need to generate enough new revenue to fund the power upgrade. How much is that? Bain's analysis of sustainable ratios of capex to revenue for cloud service providers suggests that \$500 billion of annual capex corresponds to \$2 trillion in annual revenue.

What could fund this \$2 trillion every year? If companies shifted all of their on-premise IT budgets to cloud and also reinvested the savings anticipated from applying AI in sales, marketing, customer support, and R&D (estimated at about 20% of those budgets) into capital spending on new data centers, the amount would still fall \$800 billion short of the revenue needed to fund the full investment (*see Figure 2*).

The algorithms get better. When mathematical problems grow too complex or costly, step-change progress often comes not from incremental tuning but rather from entirely new algorithmic approaches. Innovations such as MapReduce (which popularized distributed data processing) and the Transformer architecture (which unlocked a more efficient way to process sequential data) exemplify this kind of breakthrough. Even at smaller scales, algorithmic innovation can unlock meaningful gains. In foundation models, techniques such as mixed-precision matrix computation improve training and inference

Figure 2: Even with AI-related savings, \$800 billion in additional revenue would need to be generated to fund the necessary data centers



Sources: OECD; S&P Global; Bain Cloud CIO Survey; IDC; Fortune; Gartner

efficiency. Logical methods, such as chain-of-thought prompting or large model distillation, boost performance while lowering computational load. DeepSeek is a recent example, pushing the compute efficiency frontier through smarter algorithmic design. Still, even with these innovations, the path forward requires a significant increase in infrastructure to reach the 100 gigawatts of additional compute power that will be needed.

Technological breakthroughs change the landscape. History is replete with unexpected leaps of progress in computational power. Sixty years of Moore's law progress in semiconductors has given us handheld devices that far outperform the most powerful computers of the 1970s. Many speculate that quantum computing, for example, could displace the favored semiconductors trajectory of today, reducing the compute and power demands of tomorrow's systems. Bain's research suggests we are at least 10 to 15 years away from quantum computers stable enough to replace generative AI training and inference workloads. Other technological breakthroughs could include specially designed training and inference application-specific integrated circuits (ASICs), which could be more efficient than general purpose graphics processing units (GPUs), or new forms of memory or advanced packaging to improve power efficiency.

Bain's research suggests we are at least 10 to 15 years away from quantum computers stable enough to replace generative AI training and inference workloads.

Shortages hit the supply chain. It will be difficult to build data centers fast enough to meet rising demand given constraints in four areas: power supply, construction services, compute enablers (such as GPUs), and the limited supply of data center equipment (including electrical switchgear and advanced cooling). Of these, increasing the supply of electricity may be the most challenging as bringing new power generation, transmission, and distribution online in a highly regulated industry can take four years or longer.

While no single issue will solve this deep challenge, innovation, government support, and efficient markets are all factors that could help close the gap. AI has the potential not only to improve productivity but also to spur the development of new businesses and revolutionary technological advancements in areas as diverse as drug discovery, autonomous vehicles, and logistics. Such large shifts frequently unlock new value in the economy, and they could help produce the revenue needed for the necessary capex. However, without such innovations or breakthroughs, general progress could slow, and the field could be left to only those players in markets with adequate public funding.



STRATEGIC BATTLEGROUND

Humanoid Robots: From Demos to Deployment

With capabilities evolving rapidly, now is the time to begin assessing when and how bipedal robots may change industries.

By Peter Hanbury, Arjun Dutt, and Neil Malik

At a Glance

- ▶ Humanoid robotics are drawing capital and headlines, but early deployments are mostly limited to highly structured environments.
- ▶ Tech leaders should look beyond the hype and track capability trajectories: Intelligence and perception are nearing parity with humans, while handling and battery life remain gating factors.
- ▶ Commercial success will hinge on ecosystem readiness; companies that pilot early, invest in infrastructure, and build workforce trust will be well positioned when the robots are truly ready.

Humanoid robots are having a moment—from viral videos to billion-dollar valuations. The reality behind the headlines is more complex. While demonstrations dazzle, most deployments remain early-stage, with heavy reliance on human supervision.

These robots, typically bipedal with dexterous movement, advanced sensing and vision, and AI-powered reasoning, drew about \$2.5 billion in venture capital investment in 2024. Expectations for their deployment are partly driven by demand: Demographic changes in some advanced economies could lead to labor

shortages as working-age populations decline by up to 25%. (For more on the underlying economics, read the Bain Brief “Humanoid Robots at Work: What Executives Need to Know.”)

Humanoids, along with other types of robots (industrial, mobile, collaborative with humans), are part of an expanding automation toolkit to address workforce gaps and productivity challenges. For executives navigating automation strategies, understanding the real technology trajectory is critical. Companies making investment decisions need to understand which capabilities are advancing fastest, and what realistic adoption timelines look like across industries.

Reality check: Humanoid robots aren’t ready for prime time yet

Most humanoid robots today remain in pilot phases, heavily dependent on human input for navigation, dexterity, or task switching. This “autonomy gap” is real: Current demos often mask technical constraints through staged environments or remote supervision. Lessons from the autonomous vehicle sector suggest a phased approach: safe environments first, building trust through performance, then scaled deployment.

Controlled environments such as industrial, portions of retail, and select service environments are likely to be where humanoid robots are deployed first—places where the layout and environment are well known and closely controlled, and where tasks are likely to fall within a limited subset. More variable environments with greater potential for direct human interaction, such as homes, cities, or the outdoors, will take longer, especially given the capability advances that will be needed for true autonomy in unconstrained settings.

Four capabilities will determine progress

Today, core technologies in humanoid robotics remain below human capabilities. Intelligence and perception, however, are advancing rapidly and are likely to be the first to reach human-level performance (see Figure 1).

- **Intelligence:** Generative AI is advancing rapidly, enabling high-level reasoning, planning, and spatial awareness. These capabilities are likely to surpass human performance in many tasks within the next two to three years. In physical work settings, specialized knowledge is often needed, raising the bar on what data the robots must be trained on and what situations they must be able to think through.
- **Perception:** Sensors, especially vision, are at a similar stage to generative AI, catching up with human capabilities powered by advances across LiDAR and other technologies. However, vision sensors still lag the human eye in dynamic range—particularly in low-light conditions—and in identifying reflective or transparent objects, such as shiny surfaces and clear plastics.
- **Handling:** Despite advances, dexterity and fine-motor control are still in relatively earlier stages, with real gaps in tactile sensitivity and precision. But not all jobs require human-level dexterity. Tasks such as warehouse sorting or tray delivery can be executed with current levels of mechanical reach and

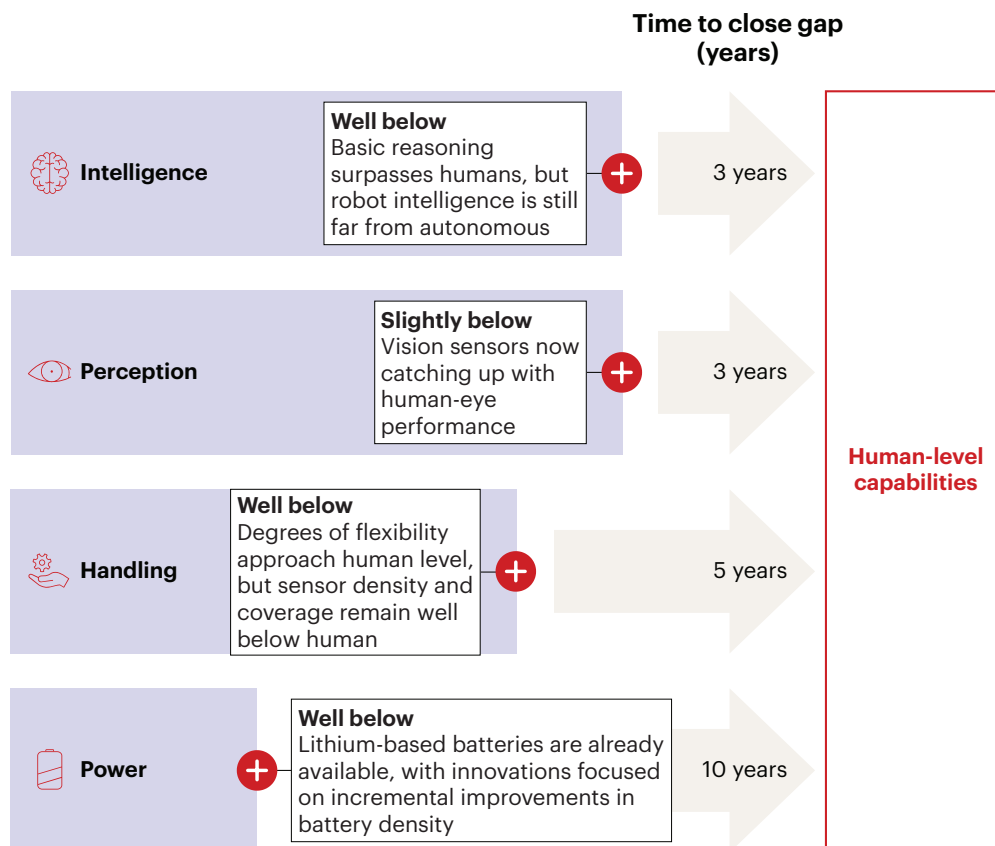
grip. Tasks such as precision manufacturing or lab work will require further advances and potentially significantly higher cost to be addressable.

- **Power:** Battery performance is improving, but slowly. Most humanoids today operate for only about two hours. Achieving a full eight-hour shift without recharging could take up to 10 years or even longer, as energy density improves and costs decline (*see Figure 2*). Until then, operators will need to rely on operational innovations such as swappable batteries and fast charging, or limit operations to environments where robots can remain continuously plugged in.

Managing expectations: Where humanoids will deploy first

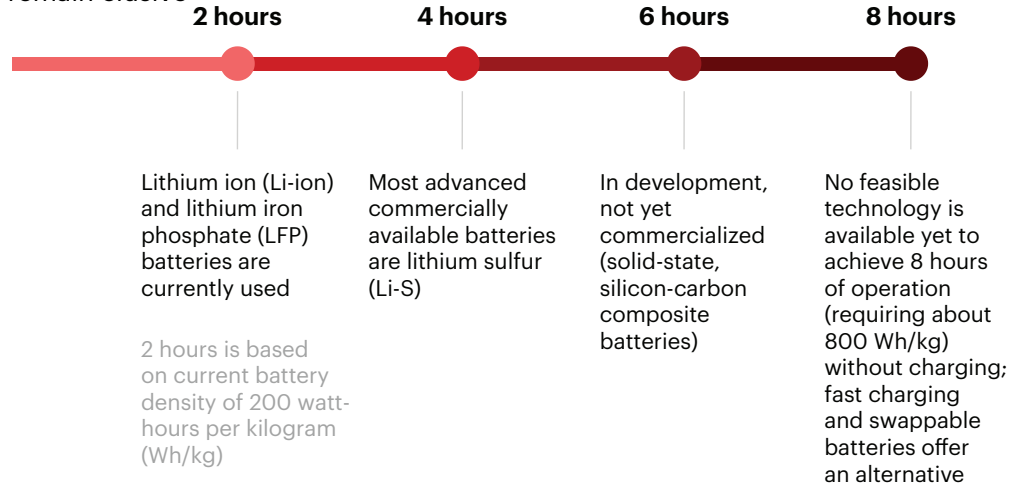
The most promising short-term value for humanoids lies not in general-purpose humanoids, but in hybrids that combine human-like perception with wheeled or static platforms or limited dexterity. For example, some companies are developing humanoids with a two-arm torso on a wheeled base to perform warehouse logistics.

Figure 1: Enabling technologies are still coming online, with intelligence and perception catching up first



Source: Bain & Company

Figure 2: By 2030, improvements in battery technology could provide robots with six hours of operation on a single charge—but a full eight-hour shift could remain elusive



Sources: Bain & Company; P3; S&P Global; market participant interviews; Bloomberg New Energy Finance

In the next three years, the first commercial applications will come from semi-structured tasks such as tote picking, palletizing, or line feeding inside durable goods factories, warehouses, and even transportation settings, where humanoids can leverage existing automation infrastructure and workflows. Early deployments will remain in closed environments where traffic is limited and predictable. While industrial robots are already common in these sectors, there remain many areas where automation is still limited, often due to variability or cost, making them prime opportunities for humanoid deployment (see Figure 3).

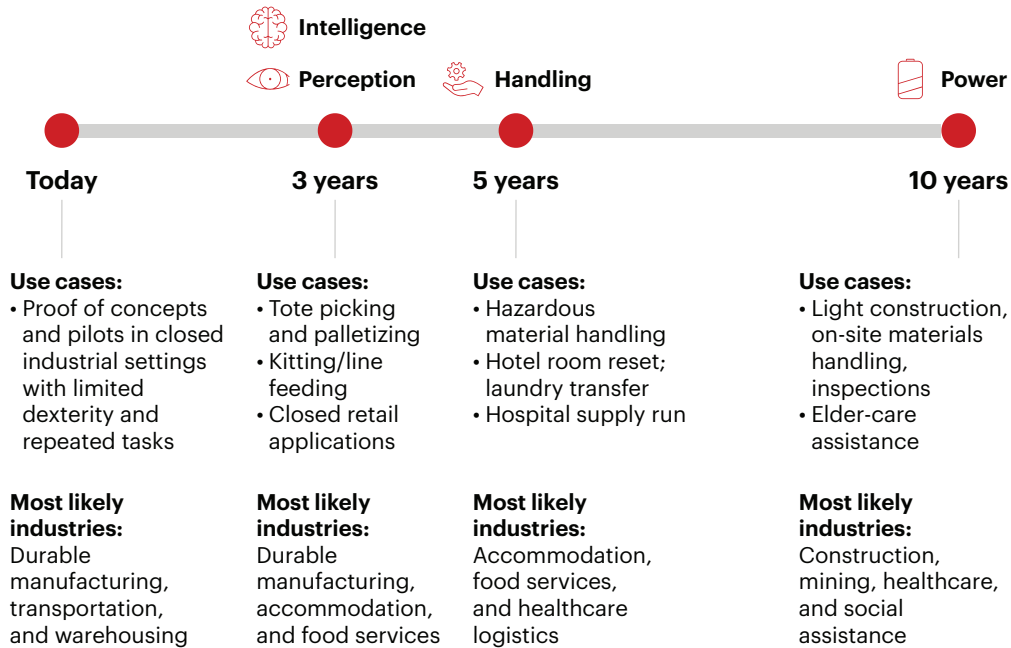
In five years, improved dexterity and battery modules will likely support robots' move into semi-structured service settings, where they'll perform tasks such as cleaning and preparing hotel rooms, hauling laundry, running hospital supplies, or shuttling hazardous materials. Jobs requiring eight-hour shifts will likely be enabled with modular battery "hot swaps" or fast charging. Safety will remain paramount, and use cases will expand into "open," guest-facing areas only as certification and human-acceptance thresholds are met.

Within the next decade, we expect physical intelligence—the ability of autonomous systems such as robots, self-driving cars, and smart spaces to perceive, understand, and act in the real world—to reach cross-domain capabilities. Once that happens, and battery power can support a full shift without intervention, real open-ended use cases will start to emerge in applications as diverse as elder-care assistance, light construction, or materials handling in mining and other remote environments.

In short, capabilities will unfold in waves: industrial workflows in controlled environments first, variable service environments next, and finally the messy, open real world, once dexterity and energy density catch up.

Figure 3: Within five years, robots could match human capabilities in intelligence, perception, and handling—though battery life could remain the limiting factor

Estimated time to match human capabilities



Source: Bain & Company

Technological readiness is only a part of the story. Turning capability into commercial traction will also require clear regulatory pathways, rigorous safety and certification regimes, workforce acceptance, and perhaps most important, public trust in machines that look and move like us. Progress on those fronts will ultimately determine how quickly humanoid robots shift from headline grabbers to everyday coworkers.

Strategic moves in the humanoid robot ecosystem

Humanoid robots are still in the early stages of development and trial, and the actions that matter most vary depending on where companies play in the emerging and potentially disruptive value chain.

- **Technology providers.** Identify key control points (e.g., AI, software stack, semiconductors, vision), differentiate offerings, and pilot vertical solutions to gather the industry learnings and data needed to build out physical intelligence. Evaluate monetization strategies, assess operational implications (e.g., over-the-air upgrades and safety certifications), and engage with regulatory bodies.
- **Equipment and component manufacturers (motors, batteries, gearboxes, materials).** Innovate and scale batteries, build mechatronics (e.g., motors, tactile sensors, and actuators) that meet humanoid specs and safety requirements, codevelop with original equipment manufacturers (OEMs),

and focus on design wins. Support multiple modalities (e.g., high dexterity through limited dexterity). Prioritize lighter and advanced materials to improve efficiency, performance, and safety.

- **Humanoid robot integrators.** Define resilient sourcing strategies across mechatronics and technology to navigate the fast-evolving landscape. Combine robots, AI, fleet management, and workflow redesign into turnkey offerings and solutions. Develop a hybrid strategy and industry-specific playbooks, service contracts, change management approaches, and safety certification pathways to accelerate customer uptake. Establish an operating ecosystem for early adopters, covering spares, services, battery swap infrastructure, and ongoing maintenance to reduce deployment risk and ensure uptime.
- **Humanoid robot adopters.** Identify addressable workflows (e.g., tote handling, palletizing, line feeding) and assess where humanoid and non-humanoid automation could generate value across tasks and geos. Run pilots to evaluate potential ROI, data readiness, and technical hurdles, and begin training the workforce. Upgrade IT/OT capabilities, invest in data infrastructure and safety standards, and mobilize broader automation adoption as familiarity with humanoid capabilities grows.

Executives who start learning early, identify opportunities across specific tasks, and develop a solid understanding of technology, data, and safety readiness will be well positioned to capture value as soon as the hardware is ready to walk through the door.

Preparing for the commercial rise of humanoid robots

Humanoid robots will not replace broad swaths of labor overnight, but they will arrive in waves and deliver clear commercial value as part of a broader automation journey across enterprises. Executives who start learning early, identify opportunities across specific tasks, and develop a solid understanding of technology, data, and safety readiness will be well positioned to capture value as soon as the hardware is ready to walk through the door.



STRATEGIC BATTLEGROUND

Quantum Computing Moves from Theoretical to Inevitable

Quantum will likely become part of a mosaic, working with classical computing to solve big problems.

By Gabe Dunn, Velu Sinha, Laurent-Pierre Baculard, Syed Ali, and Willy Chang

At a Glance

- ▶ Quantum computing is advancing, with up to \$250 billion impact possible. But full potential isn't guaranteed and may be gradual.
- ▶ Quantum is poised to augment, not replace, classical computing, each applied where most appropriate to provide solutions.
- ▶ Cybersecurity is the most pressing concern. Deploying post-quantum cryptography (PQC) can protect data from decryption.
- ▶ In industries where quantum hits first, talent gaps and long lead times mean leaders should start planning now.

Over the past two years, quantum computing has moved closer to practical, real-world applications. Breakthroughs in fidelity, error correction, and scaling qubits (the basic units of quantum computing, like the 0's and 1's bits in classical computing) across platforms signal that it's not a question of *if* but *when*.

Investment is following suit. Tech giants like Alphabet, IBM, and Microsoft are doubling down, while governments are scaling national quantum strategies. And it's not just computing: Quantum sensing, communication, and annealing (a technique for solving optimization problems) are already at work.

Given the early state of commercialization, expansive and open-minded approaches are critical for the development of specific types of qubits; the infrastructure necessary to scale and manage quantum components that will run alongside the host classical systems; and algorithms and middleware tools for connecting with data sets and sharing results.

IBM has taken this broad view, developing several generations of quantum-related technologies over the past 20 to 30 years. The company has generated interest in its quantum computing systems among academic and industrial users, in some cases by supporting solution providers as they explore the market.

But the field remains open. No single technology or vendor has pulled ahead, and many technical hurdles remain. Experimentation costs have fallen, and companies can now embark on exploring quantum with relatively modest entry costs. The opportunities and uncertainties make preparation and agility key.

Quantum's big market potential: Big but uncertain

Quantum could unlock as much as \$250 billion of market value across industries like pharmaceuticals, finance, logistics, and materials science (see *Figure 1*). While the full potential is immense, the pace of progress and extent of realization are uncertain, and many advances will need to be made beyond qubit scaling. To reach full market potential, a fully capable, fault tolerant computer at scale will be needed—and that's still years away.

At least four major barriers stand in the way:









- **Hardware maturity:** Quantum computing faces steep technological hurdles before it can reach full potential, many having to do with the need to convert or hold information in the fragile quantum state. These include physical scaling, fidelity and error correction, coherence times (that is, how long a qubit can retain its quantum state), quantum memory (similarly, the ability to store quantum information reliably over time), data loading (the process of converting traditional data to quantum information), and qubit control bottlenecks (the ability to manipulate qubits without losing fidelity or picking up cross-talk). Although some might hope for a Moore's law relative to qubit scaling, the nature of quantum devices and the challenges of scaling here are quite different, and their difficulty increases exponentially with qubit count.
- **Algorithm maturity:** While quantum computing hardware garners most of the headlines, for many use cases, practical application will require major advances in quantum algorithms (QA). Research is ongoing and major progress has been made in optimizing existing quantum algorithms, but the pace of new QA development has slowed.

- **Quantum machine learning (QML):** Over half the projected market value (about \$150 billion) sits here, but it's still mostly theoretical. Key algorithmic and data-loading bottlenecks suggest this could be among the later use cases realized, and the applications for the highest-value machine-learning cases (generative AI, LLMs) remain even more speculative.
- **Practical ROI:** Many current quantum targets, including simulation and optimization, are already being tackled with “good enough” classical computing. To justify using quantum computing instead, it would need to deliver real, sustained performance and cost advantages in places where classical computing approaches fall short, even as classical computing continues to advance.

Some expect a single “quantum breakthrough,” but we anticipate more of a gradual curve: early wins in narrow domains within five to ten years with broader adoption unfolding over time.

The market today for quantum computing hardware and services is less than \$1 billion a year. Barring a major breakthrough, roadmaps suggest that over the next five to ten years we'll see initial examples of quantum supremacy—quantum computers outperforming classical approaches in practical and useful applications—although the scope of application will be limited. These earliest practical applications in simulation (for example, metallodrug- and metalloprotein-binding affinity, battery and solar material

Figure 1: Quantum computing's market potential could be between \$100 billion and \$250 billion

	Quantum computing applications			
	Simulation \$30B–\$50B	Optimization \$20B–\$40B	Machine learning \$50B–\$150B	Cryptography \$1B–\$10B
Quantum industry uses	 Drug discovery in silico platforms \$15B–\$25B	 Finance: Portfolio optimization \$1B–\$5B	 ML platforms and services \$50B–\$150B	 Government \$1B–\$10B
	 Finance: Risk management \$5B–\$10B	 Logistics: Network optimization \$20B–\$35B		
	 Material design: Energy & chemistry \$5B–\$15B			
	 Academic research \$1B–\$5B			

Source: Bain analysis

research, or credit derivative pricing) and optimization (logistics, portfolio analysis) will boost the quantum computing market to between \$5 billion and \$15 billion by 2035—still a far cry from the \$250 billion a fully capable, full-potential quantum computing could unlock.

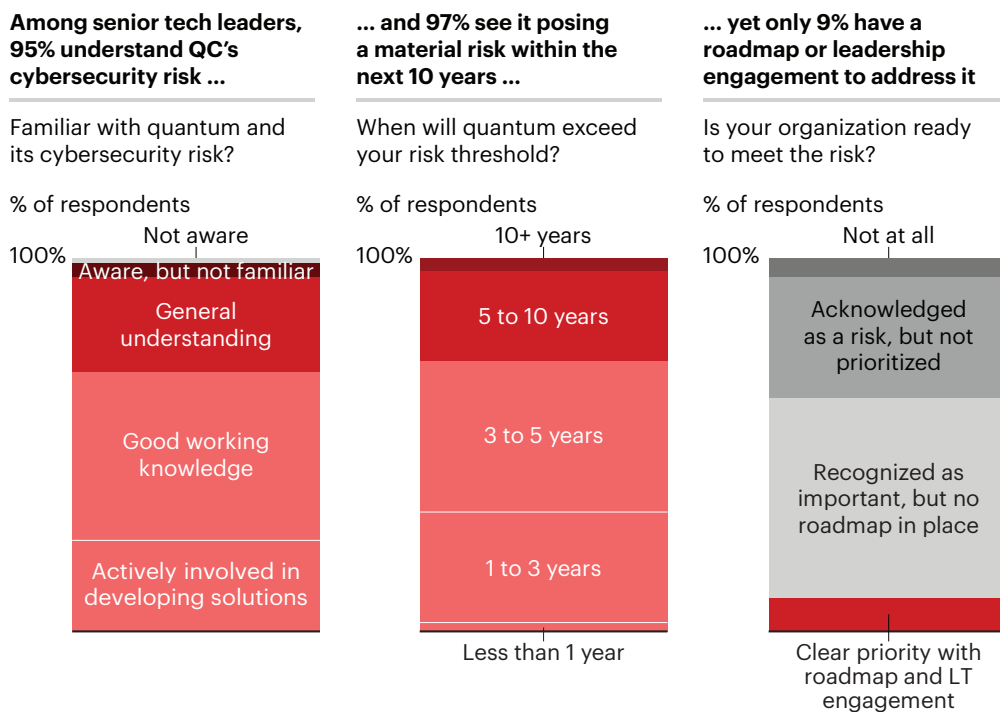
Getting ahead means choosing the right pilot use cases and investing in talent and technical readiness now.

Cybersecurity: Quantum’s immediate implication

The potential for quantum computing to overcome today’s best encryption is real. Although today’s quantum computing can’t break state-of-the-art encryption yet, some malicious organizations have embarked on a strategy of “harvest now, decrypt later,” intending to store sensitive data for five years or more, until the evolution of quantum empowers them to break the encryption.

Bain’s recent survey on the implications of post-quantum cryptography (PQC) on cybersecurity found that 73% of IT security professionals expect this to be a material risk within five years, and 32% expect it within three years, though some expect it will take longer (see *Figure 2*). However, even if it takes years

Figure 2: Most tech leaders recognize quantum’s cybersecurity risk, but very few have a plan to address it



Note: The survey population varies by question because respondents were screened out based on certain criteria, ranging from 182 to 226 participants
Source: Bain Post Quantum Cryptography (PQC) Survey, May 2025

for quantum to crack today's levels of encryption, PQC is becoming a necessity. Still, only 9% of tech leaders surveyed said they have a roadmap in place for dealing with it.

The transition isn't easy; it requires mapping the company's cryptographic landscape, updating protocols, and ensuring compliance. Companies will need to update their IT and cyber technology stacks with PQC-enabled solutions. For large or legacy-heavy organizations, this process could take years.

A future of hybrid solutions

Quantum computing will not replace classical computing—it will complement it, becoming an important part of a broad mosaic of solutions. Quantum computing will play a targeted role, solving specific problems where classical systems fall short. Quantum computing is likely to replace supercomputing tasks in initial applications, where it won't compete with high-performance data centers. As it evolves, it's likely to take on a wider range of activity, but still in a hybrid fashion where it will work with classical computing to form complete solutions. Already, much of the focus is on developing architectures in which quantum and classical computing work well together.

Quantum computing will play a targeted role, solving specific problems where classical systems fall short.

The future compute stack will be a mosaic, with quantum processors alongside CPUs, GPUs, and other accelerators optimized for specific functions. Winning companies will be those who can knit these together into a unified, high-performing system.

Symmetrically, the future of analytics will be a mosaic, too: with capabilities ranging from simple regressions to AI, agentic, high-performance compute capabilities and quantum algorithms, each evolving in complementary fashion and directed at specific business problems.

Mobilize today; lead tomorrow

For many companies, the most pressing need is securing data for a post-quantum world. In sectors where quantum is likely to have a near-term impact, developing readiness is more operational than technical. That means defining target use cases, building capabilities, forming partnerships, and scanning for signals in a fast-moving space.

Working with companies to assess their quantum readiness, we've seen it takes three to four years on average to go from awareness to a structured approach that includes a strategic roadmap, an ecosystem of partnerships, and pilot programs. Moving quantum use cases from R&D out to business units and

functions, including the time needed to experiment and climb learning curves, can take between six and nine months. This reflects the time needed for mathematical modeling, algorithm tuning, formatting incoming and outgoing data, deployment, interpretation, impact on competitiveness, as well as recruiting and training the right people. The challenge today is navigating between moving too quickly (including overinvesting) in a not-yet-mature technology and moving too slowly, which could leave a company struggling to keep pace with competitors.

Most companies are still in the early stages. With talent scarce and the learning curve steep, those who move now will shape the quantum landscape later.



Operational Transformations

State of the Art of Agentic AI Transformation	48
AI Is Transforming Productivity, but Sales Remains a New Frontier. . .	55
From Pilots to Payoff: Generative AI in Software Development . . .	62
Building the Foundation for Agentic AI	68



OPERATIONAL TRANSFORMATIONS

State of the Art of Agentic AI Transformation

Tech-forward enterprises have cracked the code on ROI for AI. Falling behind is riskier than ever as the next wave of agentic AI raises the stakes.

By David Crawford, Anne Hoecker, Chuck Whitten, Jue Wang, Aaron Lewis, and Ryan Petranovich

At a Glance

- ▶ AI leaders have moved from pilots to profits, delivering 10% to 25% EBITDA gains by scaling AI across core workflows.
- ▶ Falling behind is increasingly risky: Companies still experimenting should follow the proven playbook for AI transformation.
- ▶ Agentic AI offers another round of gains but requires more technology for agents to interact and operate across silos.
- ▶ The pace of innovation demands pragmatism, not purity; winners will build momentum with fit-for-purpose solutions.

Artificial intelligence continues to surge ahead at an unprecedented pace—catapulting some companies forward and leaving others far behind.

In 2023 and 2024, tech-forward enterprises broke through the pilot phase, achieving 10% to 25% EBITDA gains by scaling information retrieval and single-task AI. In doing so, they established a repeatable

playbook that others can now follow—grounded in robust methodology, analytic tools, and clear benchmarks.

Yet a year on, most organizations remain stuck in experimentation mode, satisfied with minor productivity gains that haven't delivered significant value. But as the leaders move forward, falling behind risks ceding competitive ground that may be difficult to recover.

And things are about to get even more interesting: In the first half of 2025, major players—including Anthropic, Alphabet, Microsoft, OpenAI, Salesforce, and others—debuted their visions of agentic AI. Tech-forward enterprises are already turning their focus from automating tasks to redesigning entire workflows. Early adopters are figuring out how agents will coexist safely, discover each other, gather the context and data they need, and collaborate productively. As the vision meets reality, they'll also wrestle with data silos, informal context, intellectual property, privacy, security, and vendor profit motives.

A purist view of architecture won't meet the moment. Given the current pace of AI innovation, the most likely scenario is that we'll see rapid, fitful, and hard-to-predict progress. Companies should maintain an architectural North Star but sustain progress with fit-for-purpose, domain-specific, and human-in-the-loop builds for the foreseeable future.

Maximizing value demands focused attention on a few key priorities:

- **Keep up the pace.** The most important tasks remain redesigning processes and workflows while cleaning up data. Falling further behind is dangerous.
- **Follow the taillights of the enterprise leaders.** The playbook is established, with methodology, analytic tools, and benchmarks available.
- **Take a principled but flexible view of architecture.** Balance long-term vision with flexible, domain-specific solutions to keep pace with AI's progress.

Leaders cracked the code

In our *2023 Tech Report*, we wrote about how some leaders were beginning to unlock the secrets of AI productivity (see “You’re Out of Time to Wait and See on AI”). In 2024, we could already see clear patterns about where AI ROI could reliably be captured (see “Five Functions Where AI Is Already Delivering”). While transformation styles varied, these early adopters have delivered a roadmap with proven methodology, analytic tools, and benchmarks. We can summarize this roadmap with five critical actions:

- Set ambitious goals based on top-down diagnostics, not trials and pilots.
- Charge general managers with meeting these targets, not the CIO or CTO.
- Redesign entire workflows, not siloed activities or use cases.

- Curate and clean the data and application environment as needed, not holistically.
- Make, buy, or partner to build capabilities for each major workflow, rather than waiting for enterprise-wide solutions.

The biggest insight from these transformations is that the most important aspects of the transformation are process redesign and cleaning up the data and application environment. Because of this, it doesn't make sense to wait for the dust to settle on technology. There's no way to cut corners on process, data, and application cleanup. Every day a company waits is another day it's left behind.

AI leaders pursue agentic AI

AI innovation increasingly focuses on enabling models to work with much more complex reasoning, context, and unstructured data while communicating with SaaS applications and other agents (*see Figure 1*).

As noted earlier, many tech giants debuted their visions of agentic AI in the first half of 2025. While flavors vary, the underlying progression of capabilities crystallizes into four levels.

- **Level 1: LLM-powered information retrieval agents** (e.g., knowledge assistants, copilots)
- **Level 2: Single-task agentic workflows** (e.g., task-doers with self-contained action loops)
- **Level 3: Cross-system agentic workflow orchestration** (e.g., complex workflow execution, supervised agents)
- **Level 4: Multi-agent constellations** (e.g., any-to-any agent, loosely coupled collaborative agents)

Tech-forward companies scaled Level 1 tools in 2023 and 2024, with varying degrees of success. When deployed in a diffuse way, they delivered microproductivity—what we might call “grab-a-coffee” time savers. But when deeply embedded in functional workflows in areas such as sales, development, and product management, the gains compounded—especially after heavy data cleaning and curation as well as continuous high-quality governance.

Levels 2 and 3 are now where capital, innovation, and deployment velocity are converging (*see Figure 2*). Level 4 is on the whiteboard, held back by several practical realities outlined in the next section.

Practical, not purist architecture

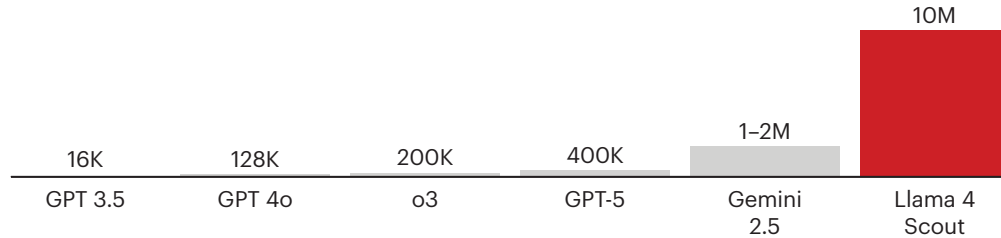
In 2025, leading tech companies turned their attention to making single- and multi-system workflows smarter, powered by agents.

As vendors race to bring their agentic visions to life, enterprise teams face new challenges: How will these agents operate safely, find and connect with each other, gather the context and data they need, and collaborate productively as the vision meets the reality of vendor profit motives, data silos, informal context, enterprise data, IP, and security?

Figure 1: AI innovation continues at an unprecedented pace

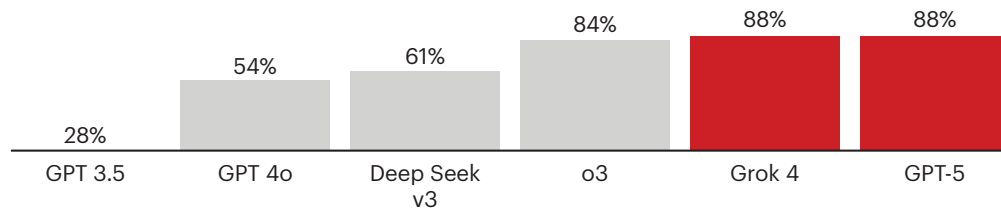
**Context windows expanded more than 10x,
allowing fuller-context reasoning and persistent memory across tasks**

Context window growth, 2023–25



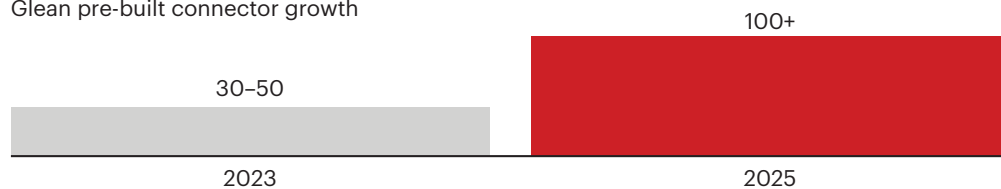
**~3x boost in multi-hop reasoning benchmarks,
pushing agents from task completion to problem solving**

GPQA accuracy



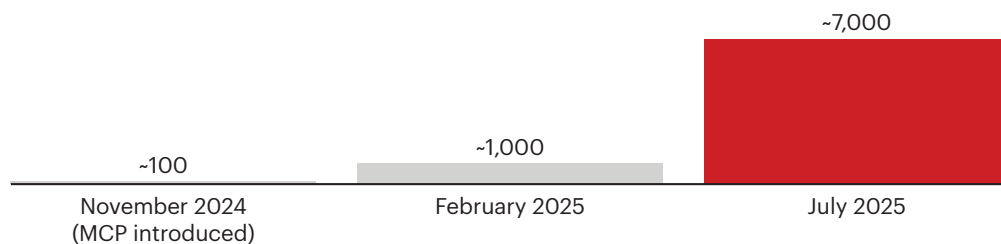
**Tool and connector ecosystems grew 3x,
with vendors rapidly expanding the number of pre-built connectors to other SaaS
in order to expand workflow reach**

Glean pre-built connector growth



**Agent-to-agent protocols progress grew 7x,
since February**

Number of MCP servers published and made available for developers and enterprise teams



Notes: GPQA stands for Graduate-Level, Google-Proof Q&A, a benchmark for evaluating reasoning in large language models (LLMs), based on answering several hundred difficult science questions; SaaS stands for software as a service; MCP stands for Model Context Protocol, a specification that allows AI to communicate with other tools and data sets; an MCP server is the component on the tool or data side that provides information for the querying AI
Sources: LLM websites; LLM benchmark websites

Technology Report 2025

Figure 2: AI companies are pushing development into Levels 2 and 3

Company	ARR	Agentic development
Anysphere	\$500M	Cursor, a software development agent, turns a GitHub issue (e.g., a bug fix or new feature) into a fully tested request and then writes code, refactors, queues tasks, and resolves conflicts
Windsurf	\$80M	Queues multi-step code editing flows inside an integrated developer environment (IDE), applies fixes, and can often merge branches with no manual clicks
Lovable	\$100M	Agent with a natural-language chat interface that designs, codes, and deploys a production web app from prompt to live URL in minutes
Crescendo	\$90M	Automates the customer experience (CX) across all channels (web, email, phone, social media, etc.), from pre-purchase to post-sale, using accurate data analytics
Decagon	\$10M	Omnichannel CX agents span chat, email, and voice, resolving issues and syncing backend tickets automatically
Glean	\$100M	Agents search every app in a company, query live data, and deliver answers or actions into the workflow
Clay	\$30M	Connects to existing CRM or productivity tools, enabling teams to build task-specific agents quickly
Perplexity	\$150M	Agent researches products, curates options, and links directly to checkout
Mercor	\$450M	Performs candidate screening, matching, and automated interviewing, and supports other self-contained tasks
Harvey	\$100M	Drafts, cites, red-lines, and files litigation documents, creating a complete production line for legal matters

Notes: ARR stands for annual recurring revenue, a measure of predictable subscription-based revenue; CRM stands for customer relationship management, a software tool for tracking customers and sales
Source: Company websites; Mercor CEO interview in Tech Crunch; CB Insights

As higher levels of agentic autonomy are pursued, enterprises are encountering a number of thorny issues.

- **Human work:** Most work happens across multiple systems and organizations, with context and informal processes.
- **Technology gaps:** These include a lack of communication standards (Model Context Protocol, or MCP, isn't USB) and compounding errors in multistep tasks.
- **Enterprise reality:** Data isn't clean, and privacy, security, and intellectual property are real concerns.
- **Vendor motives:** These run counter to open standards, shared IP, workflows, and data, leading to battles and walled gardens.

Several architectural visions exist for how higher levels of agentic autonomy can be enabled. Most call for an interconnected fabric or mesh that will register, distribute, and allow communication between agents to enable secure collaboration. Some visions are more libertarian than others, but most resemble Web 3.0: a logical vision for how things should work if no one were greedy and governance and accountability were not thorny issues. Much like Web 3.0, we expect these visions to serve as a useful aspiration, but we don't expect them to survive contact with enterprise reality unchanged.

With AI moving at breakneck speed, progress is likely to be rapid, uneven, and tough to forecast. Walled gardens will take the lead.

For this reason, a rigid approach to architecture falls short of what's needed. With AI moving at breakneck speed, progress is likely to be rapid, uneven, and tough to forecast. Walled gardens will take the lead. Fit-for-purpose custom builds will dominate enterprise-wide architectures for some time. Human-in-the-loop applications are likely the pragmatic reality for years. Context and graph analytics will remain closely guarded assets. Standards battles will play out at lightning pace (witness the MCP and Agent-to-Agent, or A2A, adoption tipping points); incumbents may try to selectively open source their IP; and domain-specific leaders will emerge.

Now is the moment to act

Maximizing value in this next phase demands disciplined focus on a small set of high-impact priorities. The organizations that move decisively will extend their lead; those that hesitate risk being left behind.

- **Keep up the pace.** The critical work remains redesigning processes and workflows while cleaning and standardizing data. Any further delay compounds technical debt and makes catching up much harder.
- **Follow the taillights of enterprise leaders.** The path forward is well-mapped. Proven playbooks, tested methodologies, advanced analytic tools, and benchmark data are already available. Focus on them to accelerate progress rather than reinventing the wheel.
- **Take a principled but flexible view of architecture.** Expect domain-specific platforms, not one-size-fits-all enterprise systems—for example, tailored solutions for supply chain, sales, and other key domains. Plan for human-in-the-loop oversight for now—think Iron Man suits, not fully autonomous Iron Man robots. Finally, select vendors strategically to limit (or at least balance) agent lock-in and preserve optionality for future evolution.



OPERATIONAL TRANSFORMATIONS

AI Is Transforming Productivity, but Sales Remains a New Frontier

Potential applications of generative and agentic AI could free up more selling time and boost conversion rates.

By Ann Bosche, Jue Wang, Peter Bowen, Tamara Lewis, Justin Murphy, and Mark Kovac

At a Glance

- ▶ Sales teams have trailed other functions in adopting and benefiting from AI, but the potential is too great to ignore.
- ▶ AI can handle tasks that free up sellers to spend more time with customers, and early successes show 30% or better improvement in win rates.
- ▶ As elsewhere, the secret to significant gains lies in reimagining sales processes rather than just automating existing ones.
- ▶ Identifying high-potential areas and deciding where to start are important first steps, along with securing C-level sponsorship.

Over the past two years, generative AI has taken center stage with promises to improve productivity by accelerating software development, streamlining marketing content, enhancing support solutions, and reducing administrative burdens. Despite the enthusiasm, most companies haven't unlocked these benefits at scale or seen meaningful gains in cost efficiency or revenue growth.

Now, agentic AI is stepping in with self-directed agents that can follow a complex workflow, set goals, plan, execute, and learn on the fly—all with minimal human input. The potential? Smarter systems, faster outcomes, and more room for people to focus on what really matters.

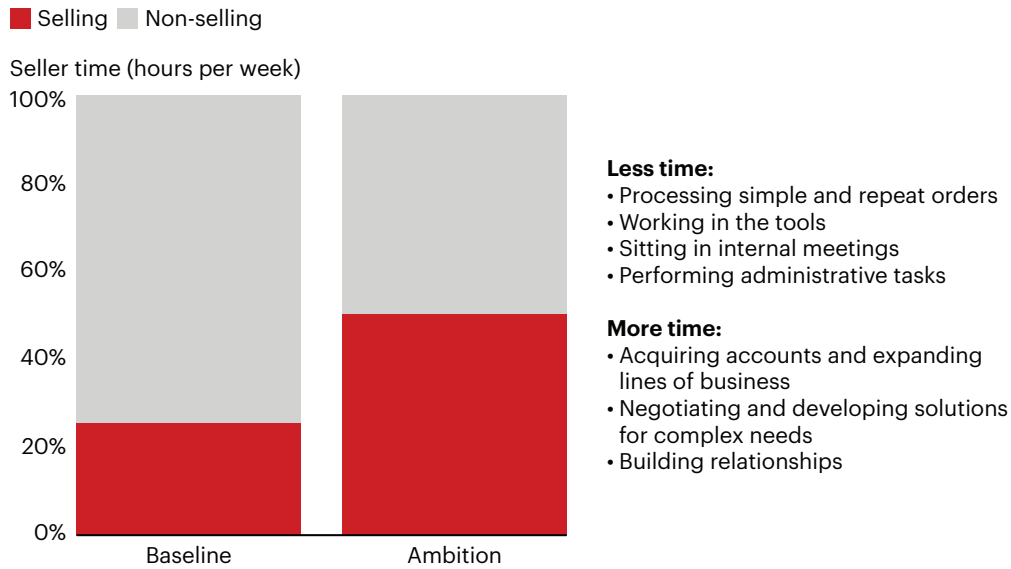
But truly successful results remain rare. Many companies are logging small productivity improvements in a few areas such as software development, but only a few can measure their success in double digits.

But truly successful results remain rare. Many companies are logging small productivity improvements in a few areas such as software development, but only a few can measure their success in double digits.

That's because most companies haven't cracked the formula yet on implementing AI at scale—and sales represents a more difficult challenge than most activities for a handful of reasons:

- **One use case rarely moves the needle** because a seller's day is fragmented across dozens of tasks. Most companies haven't stepped back to map the end-to-end selling journey, so efforts remain piecemeal.
- **Bottom-up experimentation doesn't work** because the objectives are inherently unclear.
- **Applying AI to existing processes often results in only small productivity gains** (micro-productivity) because new bottlenecks emerge. Without process redesign, companies end up automating inefficiencies instead of removing them.
- **AI needs massive data context and cleanliness** but sales and go-to-market data are spread across many systems with little quality control or governance.
- **Sales teams are stretched and distracted**, and this is one more tool in a long parade of tech promises. Unlike functions such as engineering, in which workflows are relatively standardized, sales processes vary wildly by team, region, and individual.
- **Frontline teams are often reluctant to change** their behavior. Making quota is seen as "good enough," and AI training is typically static.

The upside, however, is too promising to ignore. Sellers may spend only about 25% of their time actually selling to customers. AI could double that by taking on much of the work that surrounds selling but doesn't add much value, leaving more time for customer service (see *Figure 1*). And that's only half the

Figure 1: AI could free up more selling time and boost conversion rates

Source: Bain & Company

picture: AI also helps teams improve conversion rates at every step in the selling funnel—step-change improvements that add up to more than a 30% increase in win rates.

Mapping AI across the sales life cycle

Sales teams looking at this potential from AI need to determine where AI can deliver the biggest gains and where to start. Bain's work with business-to-business and business-to-consumer technology and consumer companies deploying AI in sales has identified 25 use cases across the various steps in the sales life cycle that leaders should explore to capture maximum benefits from deploying AI (see *Figure 2*). Some of these started as more traditional software automation and were enhanced by AI/machine learning. Many of them have been further enhanced by generative AI, and now we're seeing agentic AI deployed in several use cases.

Realizing agentic AI's potential

The deployment of agentic AI promises to unlock even more value in sales. The technology is moving quickly, but most companies are still crawling. Vendors are likely to deliver more capable applications over the next 6 to 18 months, but already we're seeing targeted results at scale—for example, among companies using no-code workflows (see *Figure 3*). The biggest hurdles remain cleaning the data, standardizing the process, making difficult governance decisions, and changing the way work gets done (which must include shutting down the old ways of working as well as access to old tools/data).

Figure 2a: Across the sales life cycle, 25 use cases are good candidates for AI

Sales life cycle

■ Prospecting and lead generation ■ Discovery and engagement ■ Lead qualification
■ Solutioning/demos ■ Quoting and closing ■ Post-sales support



Lead generation and prospecting

Lead generation and contact enrichment
Automated account and lead research
Account-needs prediction
AI business and sales development (BDR/SDR) automated outreach



High-velocity guided selling or self-serve

Inbound seller copilot
Intent-based routing and matching
Customer self-service
Shopping agent optimization



Know the customer, know the brief

Seller content search, knowledge assistant
Content curation, recommendation
Content generation, customization
Next best action, what to pitch, guided selling
Pre-call prep, near-real-time coaching, sales copilot
Post-call summarization and follow-ups
True account 360, customer 360
Interaction capture, conversation intelligence

Source: Bain & Company

Figure 2b: Across the sales life cycle, 25 use cases are good candidates for AI

Sales life cycle

■ Prospecting and lead generation ■ Discovery and engagement ■ Lead qualification
■ Solutioning/demos ■ Quoting and closing ■ Post-sales support



Automated data and artifacts

	Prospecting and lead generation	Discovery and engagement	Lead qualification	Solutioning/demos	Quoting and closing	Post-sales support
CRM automation, data cleanup, automated activity capture						
Automated configuration, price quote, demo						
Automated proposals, solutioning artifacts, standardization						
Ticket resolution, escalation, routing, automated context capture						



Operational planning and visibility

	Prospecting and lead generation	Discovery and engagement	Lead qualification	Solutioning/demos	Quoting and closing	Post-sales support
Pipeline and revenue intelligence						
Territory, account segmentation, account planning, forecasting, quota setting						
Sales manager assistant (visibility and insights, where to lean in)						



Teaming, learning, development

	Prospecting and lead generation	Discovery and engagement	Lead qualification	Solutioning/demos	Quoting and closing	Post-sales support
Modern teaming, digital sales room, partner enablement						
Adaptive learning; coach, learn—in the moment						

Figure 3: The evolution of one sample use case—lead management—shows the rapid progress delivered by AI over the past few years

Pre-2022	2023 to 2024	Now	Next 6 to 18 months
Traditional AI/machine learning	Generative AI and large language models (LLM)	Agentic: “crawl” phase	Agentic: “walk and run” phase
Definition: Systems that learn from structured data to make decisions	Definition: Mix of expert models trained on unstructured data sets to generate human-like responses	Definition: Goal-oriented systems that carry out step-by-step workflows, using reasoning plus tool integration	Definition: Self-directed AI agents that use advanced reasoning to autonomously set goals, plan, act, and learn across multistep workflows
Example: lead generation and prospecting <ul style="list-style-type: none"> Traditional propensity for model-based or rule-based scoring 	<ul style="list-style-type: none"> Lead and contact enrichment using LLMs and a wide range of first-party and third-party intent signals LLM-powered personalized outreach—why and how to engage 	<ul style="list-style-type: none"> Multistep (on-demand) lead research and outreach Real-time contact validation Automated lead research, lead segmentation, and qualification Reasons to engage What to pitch How to pitch Email outreach 	<ul style="list-style-type: none"> Self-directed research, lead creation, pipeline generation, and validation

Source: Bain & Company

Identifying where to get started

Many companies struggle with where to begin given the wide range of viable AI applications. The domains in Figure 2 illustrate use cases that are often interdependent, making it hard to move forward without first addressing foundational elements such as data architecture and business alignment.

Take lead generation and prospecting. Without clean, connected data, sellers don’t know why an account is hot, who to engage, what to pitch, or how to tailor the message. While many firms jump ahead to guided selling, reps first need insights that are trustworthy, easy to act on, and genuinely new.

The most effective pilots focus on one or two domains at the front end of the sales life cycle, in which sellers need the most help identifying, informing, and acting on leads. Leading companies build from there, prioritizing use cases based on business value and process readiness. That approach lays the groundwork for lasting gains in sales efficiency, stronger customer engagement, and seller confidence in AI tools.

Landing the full potential of AI in sales

In our work helping companies experiment with AI in sales, we've seen a consistent set of lessons emerge that separate the pilots that fizzle from those that scale.

- **Adopt an end-to-end view of a process.** Generative or agentic AI may be the headline, but the real value lies in the combination of agentic and generative AI with traditional AI and automation, process redesign, data cleanup, top-down target setting, and focus of execution.
- **Reimagine processes.** Automating mediocre processes only accelerates mediocre outcomes. Rethink selling activities and develop best-practice workflows.
- **Narrow the scope to scale.** Trying to do everything at once slows momentum. Start with high-impact slices of the sales process (for example, one or two domains out of the six in Figure 2) and build a roadmap that reflects your commercial motion.
- **Focus on the data,** with a bias toward speed over perfection. Data matters, but perfection isn't required. Focus on what's good enough to move fast and what's needed to clean up the data to reach that point. The first step is eliminating old, inaccurate, or confusing data and content—sometimes as much as 80%. It takes time and resources; don't underinvest here.
- **Test, learn, iterate.** Rapid proofs of concept are your best tool to identify where value exists. They also build conviction around the vision and the steps to get there.
- **C-level sponsorship and execution.** Solid change management is table stakes; a true AI transformation will also require sustained focus from the executive suite. A dedicated implementation team with real capabilities should be given accountability for setting targets and reaching goals.

AI has huge potential to transform sales, but most companies aren't seeing meaningful results yet. To turn promise into performance, teams need to identify and prioritize high-value use cases, reimagine critical processes, and clean up their data. It all hinges on a clear, top-down commitment to deploy AI at scale. When done right, leaders can dramatically improve life for frontline sellers and build a durable edge over competitors still stuck in wait-and-see mode.



OPERATIONAL TRANSFORMATIONS

From Pilots to Payoff: Generative AI in Software Development

AI tools improve productivity, but process changes are necessary to generate real value.

By Purna Doddapaneni, Bill Radzevych, Steven Breeden, Bharat Bansal, and Tanvee Rao

At a Glance

- ▶ Software coding was one of the first areas to deploy generative AI, but the savings have been unremarkable.
- ▶ Real gains come from applying AI across the software life cycle—not just code but product requirements, planning, test, and maintenance.
- ▶ Redesigning processes and applying time saved to other work are among the best ways to improve returns on AI investments.
- ▶ The next wave of autonomous agents raises the stakes, allowing companies to redesign entire workflows to gain a competitive edge.

Generative AI arrived on the scene with sky-high expectations, and many companies rushed into pilot projects. Yet the results haven't lived up to the hype. Two out of three software firms have rolled out generative AI tools, and among those, developer adoption is low. Teams using AI assistants see 10% to 15% productivity boosts, but often the time saved isn't redirected toward higher-value work. So even those

modest gains don't translate into positive returns. Without a plan to turn interest into habit, initial gains quickly evaporate, leaving leaders asking, "Where's the payoff?"


Beyond code completion: Generative AI for the entire life cycle

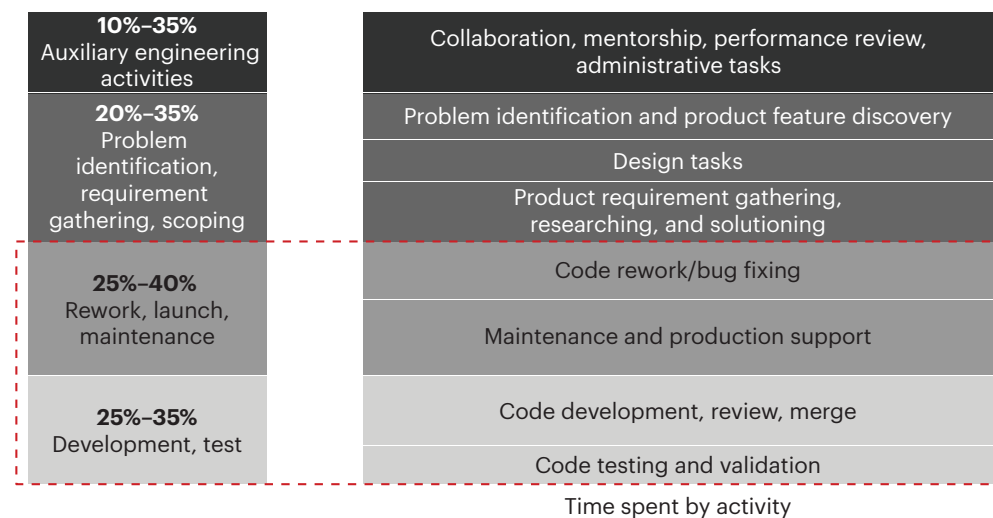
Early initiatives often fixate on code generation—that is, using generative AI to write code faster. But writing and testing code only accounts for about 25% to 35% of the time from initial idea to product launch (see Figure 1). Speeding up these steps does little to reduce time to market if others remain bottlenecked.

Real value comes from applying generative AI across the entire software development life cycle, not just coding. Nearly every phase can benefit, from the earlier discovery and requirements stages, through planning and design, to testing, deployment, and maintenance. Broad adoption, however, requires process changes. If AI speeds up coding, then code review, integration, and release must speed up as well to avoid bottlenecks. Leading companies such as Netflix recognized this and shifted testing and quality checks earlier (the "shifting left" approach) to ensure that rapidly generated code isn't stuck waiting on slow tests.

So far, generative AI has served as a smart assistant, a copilot with a human in control. Agentic AI will usher in a more autonomous wave—namely, agents that can manage multiple steps of development with little to no human intervention. For example, start-up Cognition introduced an AI "software engineer" (named Devin) in 2024 that can build and troubleshoot applications from natural language prompts.

Figure 1: AI coding assistants may be able to take on as much as 40% of the work that coders do

 Partially or fully addressable with coding assistants



Note: Industry experience is based on software-as-a-service developer team surveys, with developer teams ranging from around 2,000 to 20,000 full-time-equivalent employees
Source: Bain & Company

How leaders scale generative AI

Leading adopters treat generative AI as a fundamental transformation of their software development life cycle rather than a one-off project. They take a future-back approach to rearchitect their end-to-end software development life cycle around generative AI, embedding it deeply into workflows and scaling it enterprise-wide. They weave it into development workflows and scale it across use cases.

Leading adopters treat generative AI as a fundamental transformation of their software development life cycle rather than a one-off project.

Goldman Sachs, for example, integrated generative AI into its internal development platform and fine-tuned it on the bank's internal codebase and project documentation. This gives engineers context-aware, real-time coding solutions far beyond basic autocompletion—extending to automated code generation and even code testing—thereby significantly accelerating development cycles and boosting programmer productivity.

These leaders also make sure that generative AI's benefits translate into business value. They measure how much time AI saves and redirect that capacity to high-value work, ensuring that efficiency gains become business gains. They also modernize their environments—adopting cloud development environments, automated continuous integration and delivery pipelines, and modular architectures—to remove friction that could limit AI's impact. They also recognize that there's no one-size-fits-all approach and tailor targeted tools, playbooks, and trainings to each team's unique needs, ensuring smooth, fast adoption across diverse scenarios.

Common roadblocks to scaling generative AI

Even with generative AI's promise, many firms are stuck in pilot mode because of some common obstacles.

- **Lack of executive direction:** If senior leadership doesn't clearly prioritize generative AI, pilot efforts tend to fizzle.
- **Adoption resistance:** Under pressure, developers often fall back on old habits. Some engineers distrust AI or worry that it will undermine their role. Three of four companies say that the hardest part is getting people to change how they work. Overcoming this resistance requires strong change management.

- **Skills gaps:** Generative AI requires new skills such as writing prompts and reviewing AI output. But many firms haven't provided adequate training, so even powerful tools go underused.
- **No ROI tracking:** It's tough to prove generative AI's value without clear key performance indicators or plans for using the time saved. If you don't measure results, even real productivity gains won't show up in business terms.
- **Process or tooling mismatch:** Slow, manual processes in build, testing, or release will choke generative AI's benefits. Legacy toolchains that can't handle AI-generated code will also blunt any speed gains.

These issues explain why so many AI efforts never get out of the sandbox. The good news is that none of these barriers are insurmountable; each can be overcome with the right approach. Often, the toughest obstacles are people related, so overcoming them requires significant investment in training, communication, and cultural change.

Reimagine the software life cycle with AI at its core

To break out of pilot mode and get real returns from generative AI, tech leaders must go beyond incremental tool adoption and frame their roadmap as an AI-native reinvention of the software development life cycle. Starting with a vision of a future in which AI is seamlessly integrated into every phase of development allows teams to then plan backward to make that vision a reality. Leaders follow a roadmap to move from experimentation to scaled impact.

- **Set an AI-native vision anchored in business outcomes, not just tech metrics.** Define a bold, future-back ambition for how software will be built with AI at the core. Tie that directly to concrete business outcomes such as faster release cycles, lower defect rates, or higher customer satisfaction. And show where AI is generating real value (see *Figure 2*).
- **Turn saved time into business results.** Don't let productivity gains sit idle. Decide early how to use freed-up capacity—for instance, whether to ship more features, or reduce spending, or accelerate innovation—and tie those outcomes to financial impact. Scale successful practices across teams to maximize ROI.
- **Start with high-impact, easy wins aligned with the future vision.** Apply generative AI where it can succeed quickly—for example, generating new feature code or automating tests—and help pave the way toward an AI-native end state. Avoid fragile legacy systems at first; instead, focus on domains that are ready for AI. Early wins build momentum for broader use.
- **Cultivate AI-native talent and culture.** Provide hands-on training (such as prompt engineering or AI orchestration workshops), and actively manage the culture shift. Make upskilling a continuous effort, reassure engineers that AI is an assistant (not a replacement), and celebrate early wins to build buy-in.

Technology Report 2025

Figure 2: Key performance indicators and performance targets measure progress

Outcomes	Proposed key performance indicator	Expected target	Why it matters
Speed	Deployment frequency	From weekly to multiple deployments per day	Faster releases mean quicker delivery to customers and a competitive edge
	Lead time for changes	20% faster in 12 months	Shorter lead times accelerate time to market and responsiveness
Quality	Change failure rate (percentage of releases that cause problems)	Less than 15% in 12 months	Fewer failures reflect higher product quality and stability, reducing downtime and rework costs
	Mean time to recovery	Less than one hour to recover service	Limits downtime and business disruption
Value	Feature throughput	A 10% increase in output	Higher output means more business value delivered with the same resources
	Revenue per employee	Equal or higher	Indicates that AI lets employees focus on higher-value tasks
Experience	Developer satisfaction	Positive sentiment higher than 90%	Satisfied developers are more innovative and productive—and less likely to leave.
	Customer satisfaction	Satisfaction levels greater than 80% to 85%	Happier customers are more loyal and create repeat business

Note: Example is illustrative—actual metrics will vary across industries
Source: Bain & Company

- **Modernize processes and architecture for AI at scale.** A true AI-native approach often demands overhauling your development environment end to end. Eliminate process bottlenecks that could reduce AI's speed advantage. Adjust workflows so that faster coding leads to faster releases and isn't stuck in slow pipelines. Update development tools to handle AI outputs smoothly.
- **Prepare for autonomous workflows.** Finally, as generative AI evolves from copilot to autonomous agent, start experimenting with AI handling end-to-end development tasks. Developers' roles may shift to guiding these agents as "intent engineers" or "AI orchestrators." Assign an agent to build a simple app in a sandbox, with humans stepping in only if needed. These trials will show how far AI can go, where oversight is essential, and what skills or workflows need to evolve—thus signaling your intent to lead in the next wave of development.

Closing the gap: From experimentation to execution

Generative AI's promise is real, but capturing it requires moving beyond one-off pilots. It takes bold leadership to drive adoption, revamped processes to embed AI at every step, and a focus on measurable outcomes to analyze results and make adjustments. The winners won't be those dabbling in flashy demos but rather those redesigning their workflows to fully integrate AI and deliver tangible improvements. Already, some companies report 25% to 30% productivity boosts by pairing generative AI with end-to-end process transformation—far above the 10% gains from basic code assistants.

Experiments pay off only when backed by a well-defined approach that converts innovation into measurable results.

An even bigger leap is on the horizon as AI evolves from assistant to autonomous agent—a shift that could redefine software development and widen the gap between firms that treat AI as a novelty and those that embrace it as transformative. Generative AI's capabilities are steadily broadening, and the gains seen today are expected to continue growing over the next 12 to 24 months as models improve their performance and reliability. Tech executives must excel at implementing generative AI today while also preparing their teams for a more AI-driven development model tomorrow.

Experiments pay off only when backed by a well-defined approach that converts innovation into measurable results. Now is the time to act. Organizations that move decisively with a clear vision and bold execution will capture real returns and redefine how software is built; those that hesitate risk being left behind.



OPERATIONAL TRANSFORMATIONS

Building the Foundation for Agentic AI

Capturing the full potential of agents will require modernizing enterprise architecture.

By Pascal Gautheron, Chris Bell, and Stephen Hardy

At a Glance

- ▶ Agentic AI is a structural shift in enterprise tech, reshaping companies with agents that can reason, coordinate, and execute complex workflows.
- ▶ Most companies aren't ready: Capturing full value requires rethinking systems, data, and governance to support scalable, safe agent deployment.
- ▶ Tech leaders should continue to modernize core platforms while prioritizing interoperability, security, and accountability.
- ▶ Early movers are focusing their investments on the most valuable areas, building foundational capabilities, and using agents in the transformation.

Agentic AI isn't just another wave of automation; it's a structural shift in enterprise technology, one with the potential to completely redefine how work gets done. Previous waves of automation tackled parts of processes, leaving exceptions where humans had to step in. AI agents can reason, collaborate, and coordinate actions, allowing them to accomplish complex, multistep, nondeterministic processes that have so far depended on humans.

It's easy to see the transformative potential of this, from improved operational efficiency and customer experience to sharper decision making and beyond. Forward-looking leaders aren't asking if agentic AI will reshape their business but how to prepare their organizations to deploy it safely and effectively.

As executives reflect on how agentic AI might shape their business, from competitive positioning to their talent model, they will have to consider how it challenges the fundamentals of their IT architecture. Agentic AI architecture builds on the rise of composable microservices architecture and the use of enterprise cloud services that many companies have already been investing in. But to fully capture the value while navigating the risks, they will need to rethink how AI is embedded across the architecture—the systems, processes, and governance. Enterprises need to ensure agents have the context they need in real time, the ability to observe and explain behavior, and the guardrails to execute safely, securely, and cost-effectively. Current architectures simply cannot handle this balance when AI agents are used in the thousands across the enterprise—yet.

Architecture adapts to support agentic AI

Agentic AI should complement rather than take over the existing architecture. Tech teams will need to deploy it thoughtfully, with clear scope and controls in place. Agents are best for complex, nondeterministic problems that span multiple business domains and systems, rely on unstructured data and contextual reasoning, depend on real-time inputs, and until now have required human intervention.

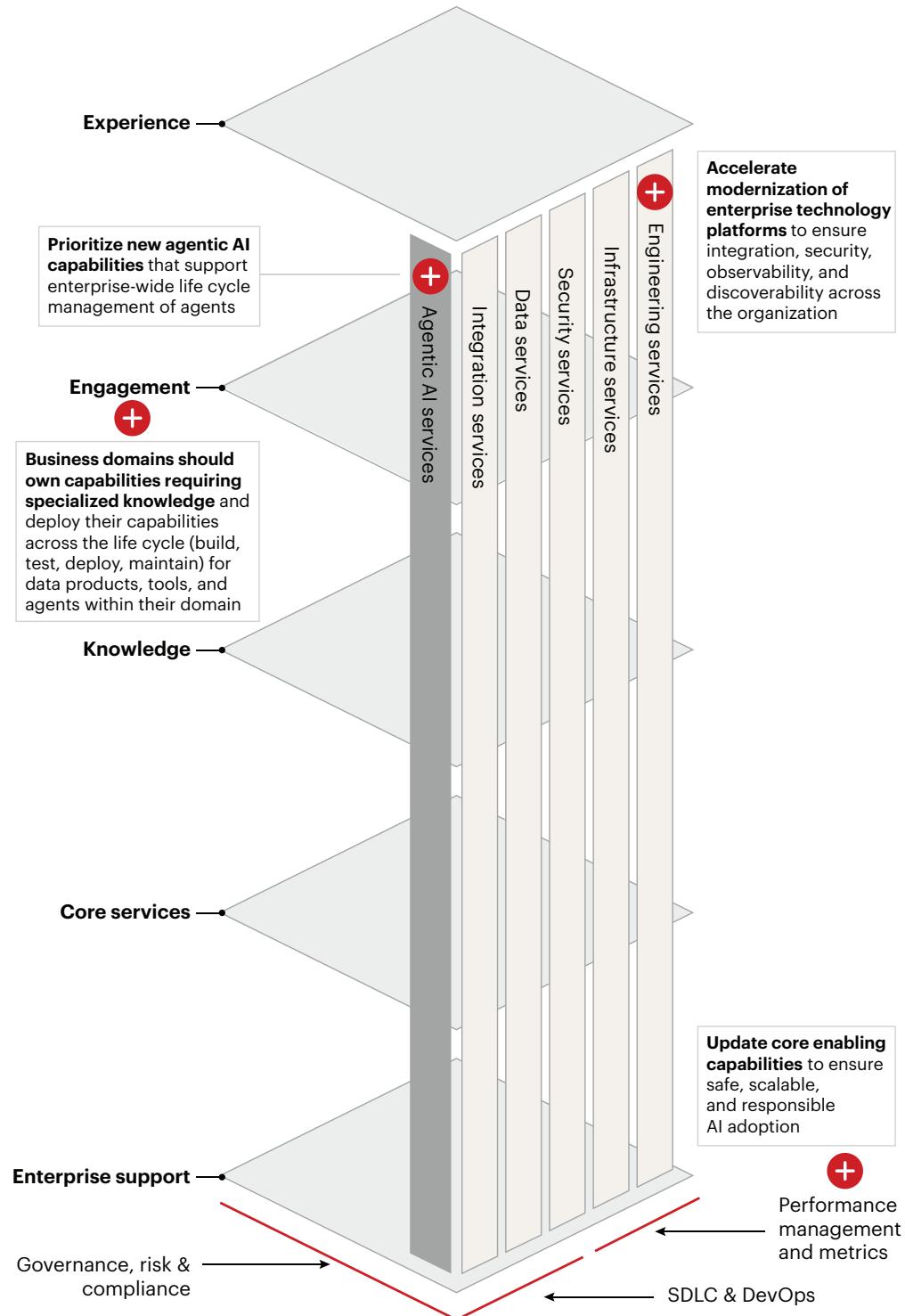
Agentic AI should complement rather than take over the existing architecture. Tech teams will need to deploy it thoughtfully, with clear scope and controls in place.

Higher-level orchestrator agents are like project managers that oversee a whole process, breaking it down into subtasks and tracking progress. Task agents execute individual tasks and send back the results to the orchestrator. The orchestrator then compiles results and adjusts workflows as needed.

Teams updating the IT architecture should consider several key principles as they evolve the design to support an agentic framework (*see Figure 1*).

Modernize the core platform. To fully realize the potential of agentic AI, many organizations will need to modernize their technology foundations. This means making core business capabilities easy for agents to find and use in real time. Achieving this may require reworking older, batch-based systems to be more flexible, accessible by APIs, and able to respond to real-time events. Adopting modular, industry-standard frameworks, like the Banking Industry Architecture Network, will help accelerate this shift. However,

Figure 1: Updating the IT architecture for agentic AI requires special attention to several facets



Source: Bain & Company

these modern systems will need to work alongside existing infrastructure for the foreseeable future, which could add some architectural complexity in the near term.

Ensure interoperability of agentic services. As agents roll out across the tech stack, consistent interoperability standards, such as the model context protocol (MCP), and frictionless integrations will be critical for breaking down silos and capturing the full value of agentic AI. Most organizations will need to support a mix of frameworks. These will include custom agents built on engineering tools, prebuilt agents embedded in vendor platforms, and dynamically generated agents in data platforms. Frameworks are themselves becoming more agentic. For example, a software development life cycle (SDLC) agent could coordinate a team of specialized agents (design, analyst, engineer, quality assurance) that collaborate to deliver a complete solution, from concept to deployment.

Distribute accountability. While central platform teams will control the core agentic platforms, accountability for assembling, training, testing, deploying, and monitoring agents needs to be distributed to business domains. Success will hinge on making domain expertise and knowledge assets—such as product documentation, business logic, feature stores, models, and data products—readily discoverable and accessible to agents.

Agentic AI isn't just another wave of automation; it's a structural shift in enterprise technology, one with the potential to completely redefine how work gets done.

Scale data access. Scalable access to structured and unstructured data is essential. Most organizations still lack the required ingestion pipelines for unstructured sources such as documents, emails, voice recordings, images, videos, and call transcripts. These data sources are critical for agentic reasoning, especially in manual or exception-driven processes where necessary knowledge often resides outside core systems of record or even outside the organization. For example, one European bank built foundational infrastructure to consistently use both structured and unstructured data to create a holistic view of each customer, enabling it to automate and personalize its engagement marketing—driving smarter, more targeted interactions at scale.

Update governance and controls. As agents take on more decision making, governance and controls must evolve. Real-time explainability, behavioral observability, and adaptive security are essential to mitigate risk, maintain customer trust, and avoid regulatory or reputational fallout. At the same time, organizations must manage the volatility of compute costs through dynamic resource allocation, edge deployment strategies, and AI-native financial operations practices.

Shift the engineering paradigm. Software engineering and DevOps processes, both tooling and workflows, need to evolve to manage the full life cycle of AI agents, including how they are tested, monitored, and safely deployed as they learn and adapt over time. AI agents are also poised to transform how engineering teams operate, taking on more of the day-to-day development, testing, deployment, and system operations. This shift will free up engineers to focus on higher-value work like architecture, strategy, and innovation.

Reimagine agent experience and access. In this new framework, agents become first-class channels and citizens. As channels, they are emerging as primary interfaces for interacting with customers and employees—on par with websites, mobile apps, and call centers. As citizens, they are fully embedded participants in business operations, empowered to act, make decisions, and collaborate across systems. This demands a reimagining of experience design: Conversational interfaces will dominate human engagement, while agent-to-agent coordination will drive autonomous action across workflows, systems, and even organizational boundaries. To make this work at scale—and safely—enterprises must establish robust frameworks for identity, consent, and fine-grained access control. For example, a South American bank that uses agents to facilitate real-time PIX payments through WhatsApp allows customers to simply send a photo or text describing the payment they want to make. The AI agent interprets the request, identifies the appropriate payment, confirms it with the customer, and then authorizes and sends the payment—all within a conversational experience.

The implementation imperative

Over the next three to five years, 5% to 10% of technology spending could be directed toward building foundational capabilities, including agent platforms, communication protocols, real-time data access and discoverability for agents, and modern security and observability frameworks.

Over time, investment in agentic AI will grow. Up to half of technology spending could be on agents running across the enterprise to support business domains.

Over time, investment in agentic AI will grow. Up to half of technology spending could be on agents running across the enterprise to support business domains.

Still, the long-term economics are favorable, as efficiency and process improvements will outweigh the costs. Investments will need to be tightly focused, with an emphasis on delivering value quickly to ensure buy-in from the business. Frameworks for successful transformations typically follow a pattern of four motions:

- Focus on a few business domains to generate early value, rather than just building capabilities. Reimagining processes from end to end accelerates returns, lowers cost per agent, and lays the groundwork for scalable, enterprise-wide adoption.
- Evaluate current architecture for agentic readiness, identifying the capabilities required to scale. This includes laying the groundwork for agent development toolchains, enabling seamless system interoperability, and fast-tracking the modernization of vector databases, event architectures, and core infrastructure.
- Define and embed observability, security, governance, and controls, providing traceability, accountability, anomaly detection, and cost discipline. For agentic AI to scale safely across the enterprise, these guardrails must be built in from the start, not bolted on later.
- Use agentic AI in the transformation to reduce effort, control costs, and ensure outcomes. Delivering value early helps fund the rest of an agentic transformation.

Companies that don't want to fall behind should be preparing and investing now: hiring and upskilling teams, embedding new capabilities, and understanding the necessary architectural changes. Those that wait will struggle to catch up. Agentic AI is already reshaping the enterprise, and only those that move decisively—redesigning their architecture, teams, and ways of working—will unlock its full value.

Bold ideas. Bold teams. Extraordinary results.

Bain & Company is a global consultancy that helps the world's most ambitious change makers define the future.

Across the globe, we work alongside our clients as one team with a shared ambition to achieve extraordinary results, outperform the competition, and redefine industries. We complement our tailored, integrated expertise with a vibrant ecosystem of digital innovators to deliver better, faster, and more enduring outcomes. Our 10-year commitment to invest more than \$1 billion in pro bono services brings our talent, expertise, and insight to organizations tackling today's urgent challenges in education, racial equity, social justice, economic development, and the environment.

