Tech Trends 2025

In Deloitte's 16th annual Tech Trends report, AI is the common thread of nearly every trend. Moving forward, it will be part of the substructure of everything we do.



Deloitte.
Insights

Table of contents

02 . . . Executive summary

INTRODUCTION
05 . . . AI everywhere: Like magic, but with algorithms

INTERACTION
09 . . . Spatial computing takes center stage

INFORMATION
17 . . . What's next for AI?

COMPUTATION
27 . . . Hardware is eating the world

BUSINESS OF TECHNOLOGY
37 . . . IT, amplified: AI elevates the reach (and remit) of the tech function

CYBER AND TRUST
45 . . . The new math: Solving cryptography in an age of quantum

CORE MODERNIZATION
53 . . . The intelligent core: AI changes everything for core modernization

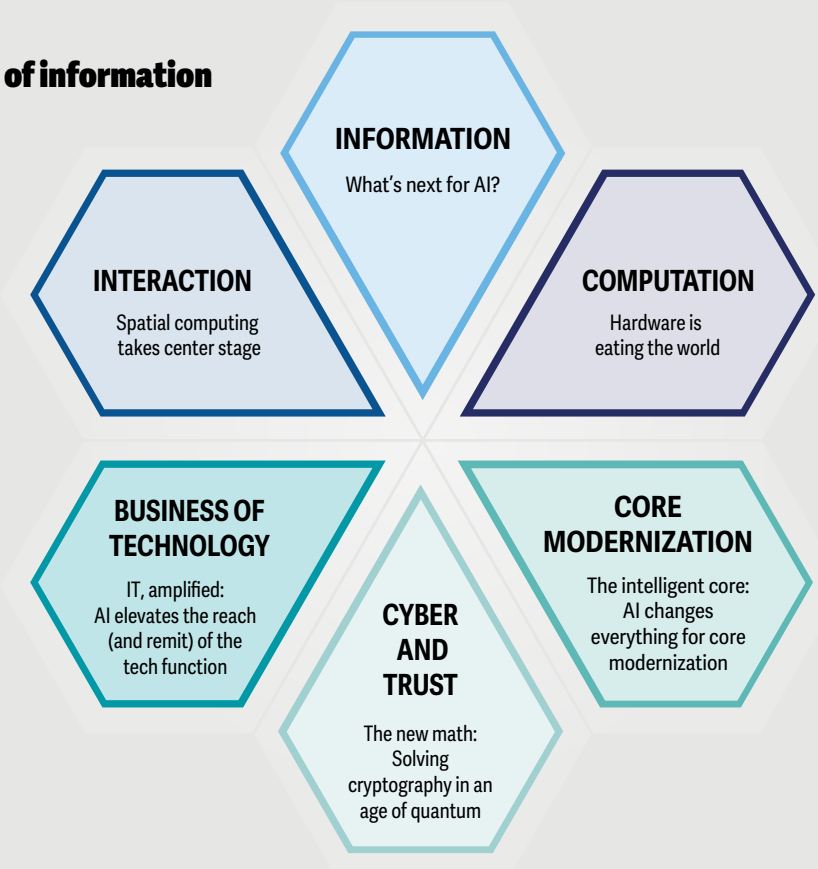
CONCLUSION
60 . . . Breadth is the new depth: The power of intentional intersections

Executive summary

Tech Trends, Deloitte's flagship technology report, explores the emergence of trends in three elevating forces (interaction, information, and computation) and three grounding forces (business of technology, cyber and trust, and core modernization)—all part of our macro technology forces framework (figure 1). Tech Trends 2025, our 16th trip around the sun, previews a future in which artificial intelligence will be as foundational as electricity to daily business and personal lives. As our team in Deloitte's Office of the CTO put finishing touches on Tech Trends 2025, we realized that AI is a common thread in nearly every trend. We expect that going forward, AI will be so ubiquitous that it will be a part of the unseen substructure of everything we do, and we eventually won't even know it's there.

Figure 1

Six macro forces of information technology



Introduction

AI everywhere: Like magic, but with algorithms

Generative AI continues to be the buzzword of the year, but Tech Trends 2025—and in fact, the future of technology—is about much more than AI. This year's report reveals the extent to which AI is being woven into the fabric of our lives. We'll eventually take it for granted and think of it in the same way that we think of HTTP or electricity: We'll just expect it to work. AI will perform quietly in the background, optimizing traffic in our cities, personalizing our health care, or creating adaptative and accessible learning paths in education. We won't proactively use it; we'll simply experience a world in which it makes everything work smarter, faster, and more intuitively—like magic, but grounded in algorithms. The six chapters of Tech Trends 2025 reflect this emerging reality.

Interaction

Spatial computing takes center stage

Spatial computing continues to spark enterprise interest because of its ability to break down information silos and create more natural ways for workers and customers to interact with information. We're already seeing enterprises find success with use cases like advanced simulations that allow organizations to test different scenarios to see how various conditions will impact their operations. With a stronger focus on effectively managing spatial data, organizations can drive more cutting-edge applications. In the coming years, advancements in AI could lead to seamless spatial computing experiences and improved interoperability, ultimately enabling AI agents to anticipate and proactively meet users' needs.

Information

What's next for AI?

To take advantage of the burgeoning excitement around generative AI, many organizations have already adopted large language models (LLMs), the best option for many use cases. But some are already looking ahead. Despite their general applicability, LLMs may not be the most efficient choice for all organizational needs. Enterprises are now considering small language models and open-source options for the ability to train LLMs on smaller, more accurate data sets. Together with multimodal models and AI-based simulations, these new types of AI are building a future where enterprises can find the right type of AI for each task. That includes AI that not only answers questions but also completes tasks. In the coming years, a focus on execution may usher in a new era of agentic AI, arming consumers and organizations with co-pilots capable of transforming how we work and live.

Computation

Hardware is eating the world

After years of software dominance, hardware is reclaiming the spotlight. As AI demands specialized computing resources, companies are turning to advanced chips to power AI workloads. In addition, personal computers embedded with AI chips are poised to supercharge knowledge workers by providing access to offline AI models while "future-proofing" technology infrastructure, reducing cloud computing costs, and enhancing data privacy. Although AI's increased energy demands pose sustainability challenges, advancements in energy sources and efficiency are making AI hardware more accessible. Looking forward, AI's continued integration into devices could revolutionize the Internet of Things and robotics, transforming industries like health care through smarter, more autonomous devices.

Business of technology

IT, amplified: AI elevates the reach (and remit) of tech talent

After years of progressing toward lean IT and everything-as-a-service offerings, AI is sparking a shift away from virtualization and austere budgets. Long viewed as the lighthouse of digital transformation throughout the enterprise, the IT function is now taking on AI transformation. Because of generative AI's applicability to writing code, testing software, and augmenting tech talent in general, forward-thinking technology leaders are using the current moment as a once-in-a-blue-moon

opportunity to transform IT across five pillars: infrastructure, engineering, finance operations, talent, and innovation. As both traditional and generative AI capabilities grow, every phase of tech delivery could see a shift from human in charge to human in the loop. Such a move could eventually return IT to a new form of lean IT, leveraging citizen developers and AI-driven automation.

Cyber and trust

The new math: Solving cryptography in an age of quantum

In their response to Y2K, organizations saw a looming risk and addressed it promptly. Today, IT faces a new challenge, and it will have to respond in a similarly proactive manner. Experts predict that quantum computers, which could mature within five to 20 years, will have significant implications for cybersecurity because of their ability to break existing encryption methods and digital signatures. This poses a risk to the integrity and authenticity of data and communications. Despite the uncertainty of the quantum computer timeline, inaction on post-quantum encryption is not an option. Emerging encryption standards offer a path to mitigation. Updating encryption practices is fairly straightforward—but it's a lengthy process, so organizations should act now to stay ahead of potential threats. And while they're at it, they can consider tackling broader issues surrounding cyber hygiene and cryptographic agility.

Core modernization

The intelligent core: AI changes everything for core modernization

Core systems providers have invested heavily in AI, rebuilding their offerings and capabilities around an AI-fueled or AI-first model. The integration of AI into core enterprise systems represents a significant shift in how organizations operate and leverage technology for competitive advantage. This transformation is about automating routine tasks and fundamentally rethinking and redesigning processes to be more intelligent, efficient, and predictive. It requires careful planning due to integration complexity, strategic investment in technology and skills, and a robust governance framework to ensure smooth operations. But beware of the automation paradox: The more complexity is added to a system, the more vital human workers become. Adding AI to core systems may simplify the user experience, but it will make them more complex at an architectural level. Deep technical skills are still critical for managing AI in core systems.

Conclusion

Breadth is the new depth: The power of intentional intersections

Organizations have long relied on innovation-driven new revenue streams, synergies created through mergers and acquisitions, and strategic partnerships. But increasingly, segmentation and specialization have given way to intentional intersections of technologies and industries. For example, when two technologies intersect, they are often complementary, but they can also augment each other so that both technologies ultimately accelerate their growth potential. Similarly, new opportunities can emerge when companies aim to extend their market share by purposefully partnering across seemingly disparate industries.

INTRODUCTION

AI everywhere: Like magic, but with algorithms

Tech Trends 2025 reveals how much artificial intelligence is being woven into the fabric of our lives—making everything work smarter, faster, and more intuitively

Kelly Raskovich

Two years after generative artificial intelligence staked its claim as the free space on everyone's buzzword-bingo cards, you'd be forgiven for imagining that the future of technology is simply … more AI. That's only part of the story, though. We propose that the future of technology isn't so much about more AI as it is about ubiquitous AI. We expect that, going forward, AI will become so fundamentally woven into the fabric of our lives that it's everywhere, and so foundational that we stop noticing it.

Take electricity, for example. When was the last time you actually thought about electrons? We no longer marvel that the lights turn on—we simply expect them to work. The same goes for HTTP, the unseen thread that holds the internet together. We use it every day, but I'd bet most of us haven't thought about (let alone uttered) the word "hypertext" in quite some time.

AI will eventually follow a similar path, becoming so ubiquitous that it will be a part of the unseen substructure of everything we do, and we eventually won't even know it's there. It will quietly hum along in the background, optimizing traffic in our cities, personalizing our health care, and creating adaptative and accessible learning paths in education. We won't "use" AI. We'll just experience a world where things work smarter, faster, and more intuitively—like magic, but grounded in algorithms. We expect that it will provide a foundation for business and personal growth while also adapting and sustaining itself over time.

Nowhere is this AI-infused future more evident than in this year's Tech Trends report, which each year explores emerging trends across the six macro forces of information technology (figure 1 in the executive summary). Half of the trends that we've chronicled are elevating forces—interaction, information, and computation—that underpin innovation and growth. The other half—the grounding forces of the business of technology, cyber and trust, and core modernization—help enterprises seamlessly operate while they grow.

As our team put the finishing touches on this year's report, we realized that this sublimation and diffusion of AI is already afoot. Not the "only trend" nor "every trend," AI is the scaffolding and common thread buttressing nearly every trend. (For those keeping a close eye at home, "The new math: Solving cryptography in an age of quantum"—about the cybersecurity implications of another game-changing technology, quantum computing—is the only one in which AI does not have a foundational role. Yet behind the scenes, AI advancements are accelerating advances in quantum.)

- **Spatial computing takes center stage:** Future AI advancements will enhance spatial-computing simulations, eventually leading to seamless spatial-computing experiences integrated with AI agents.

- **What's next for AI?:** As AI evolves, the enterprise focus on large language models is giving way to small language models, multimodal models, AI-based simulations, and agents that can execute discrete tasks.

- **Hardware is eating the world:** After years of software dominance, hardware is reclaiming the spotlight, largely due to AI's impact on computing chips and its integration into end-user devices, the Internet of Things, and robotics.

- **IT, amplified: AI elevates the reach (and remit) of tech talent:** AI's applicability to writing code, testing software, and augmenting tech talent is transforming IT and sparking a shift away from virtualization and austere budgets.

- **The intelligent core: AI changes everything for core modernization:** Core systems providers have invested heavily in AI, which may simplify the user experience and data-sharing across applications but will make these systems more complex at an architectural level.

Because we expect AI to become part of tomorrow's foundational core—like electricity, HTTP, and so many other technologies—it's exciting to think about how AI might evolve in the next few years as it marches toward ubiquity, and how we as humans may benefit. We here at Tech Trends will be chronicling every step of the journey.

Until next time,



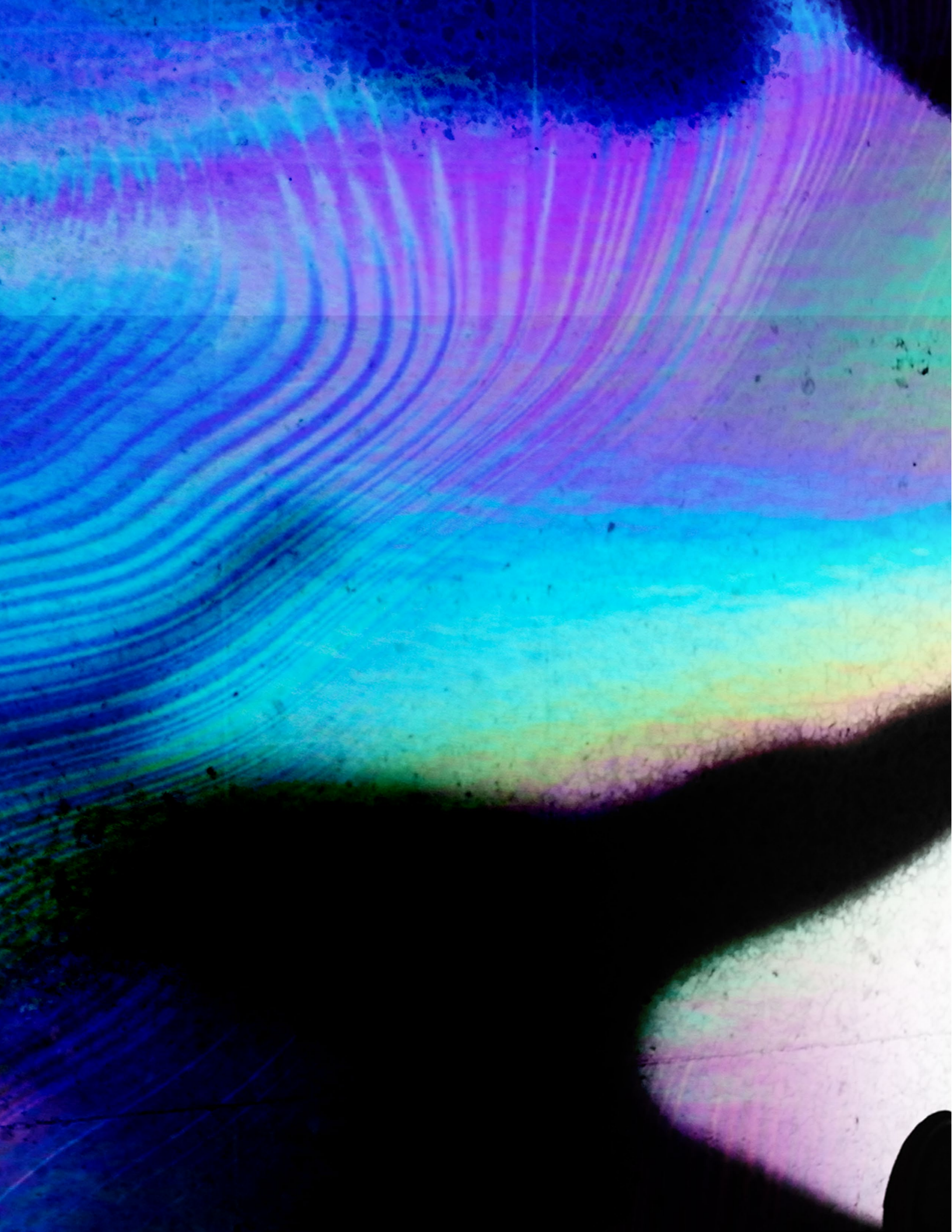
Kelly Raskovich
Office of the CTO
Executive editor, Tech Trends

Trending the trends

	INTERACTION		INFORMATION		COMPUTATION		BUSINESS OF TECHNOLOGY		CYBER AND TRUST	CORE MODERNIZATION
2025	Spatial computing takes center stage		What's next for AI?		Hardware is eating the world		IT, amplified		The new math	The intelligent core
2024	Interfaces in new places		Genie out of the bottle		Smarter, not harder		From DevOps to DevEx		Defending reality	Core workout
2023	Through the glass		Opening up to AI		Above the clouds		Flexibility, the best ability		In us we trust	Connect and extend
2022			Data sharing made easy		Blockchain: Ready for business	Cloud goes vertical	DEI tech: Tools for equity	The tech stack goes physical	Cyber AI	IT, disrupt thyself
2021	Rebooting the digital workplace	Bespoke for billions	Machine data revolution	ML Ops: Industrialized AI			Strategy, engineered	Supply unchained	Zero trust	Core revival
2020	Human experience platforms		Digital twins				Finance and the future of IT	Architecture awakens	Ethical technology and trust	
2019	Intelligent interfaces	Beyond marketing	AI-fueled organizations		NoOps in a serverless world		Connectivity of tomorrow		DevSecOps and the cyber imperative	
2018	Digital reality		Enterprise data sovereignty		API imperative	Blockchain to blockchains	No-collar workforce	Reengineering technology		The new core
2017	Mixed reality		Dark analytics	Machine intelligence	Everything as-a-service	Trust economy	IT unbounded	Inevitable architecture		
2016	Internet of Things	AR and VR go to work	Industrialized analytics		Democratized trust		Right speed IT	Autonomic platforms		Reimagining core systems

Note: To learn more about past Tech Trends, go to www.deloitte.com/us/TechTrends

Source: Deloitte analysis.



INTERACTION

Spatial computing takes center stage

What is the future of spatial computing? With real-time simulations as just the start, new, exciting use cases can reshape industries ranging from health care to entertainment.

Kelly Raskovich, Bill Briggs, Mike Bechtel, and Ed Burns

Today's ways of working demand deep expertise in narrow skill sets. Being informed about projects often requires significant specialized training and understanding of context, which can burden workers and keep information siloed. This has historically been true especially for any workflow involving a physical component. Specialized tasks demanded narrow training in a variety of unique systems, which made it hard to work across disciplines.

One example is computer-aided design (CAD) software. An experienced designer or engineer can view a CAD file and glean much information about the project. But those outside of the design and engineering realm—whether they're in marketing, finance, supply chain, project management, or any other role that needs to be up to speed on the details of the work—will likely struggle to understand the file, which keeps essential technical details buried.

Spatial computing is one approach that can aid this type of collaboration. As discussed in Tech Trends 2024, spatial computing offers new ways to contextualize business data, engage customers and workers, and interact with digital systems. It more seamlessly blends the physical and digital, creating an immersive technology ecosystem for humans to more naturally interact with the world.[1] For example, a visual interaction layer that pulls together contextual data from business software can allow supply chain workers to identify parts that need to be ordered and enable marketers to grasp a product's overall aesthetics to help them build campaigns. Employees across the organization can make meaning of and, in turn, make decisions with detailed information about a project in ways anyone can understand.

If eye-catching virtual reality (VR) headsets are the first thing that come to mind when you think about spatial computing, you're not alone. But spatial computing is about more than providing a visual experience via a pair of goggles. It also involves blending standard business sensor data with the Internet of Things, drone, light detection and ranging (LIDAR), image, video, and other three-dimensional data types to create digital representations of business operations that mirror the real world. These models can be rendered across a range of interaction media, whether a traditional two-dimensional screen, lightweight augmented reality glasses, or full-on immersive VR environments.

Spatial computing senses real-world, physical components; uses bridging technology to connect physical and digital inputs; and overlays digital outputs onto a blended interface (figure 1).[2]

Spatial computing's current applications are as diverse as they are transformative. Real-time simulations have emerged as the technology's primary use case. Looking ahead, advancements will continue to drive new and exciting use cases, reshaping industries such as health care, manufacturing, logistics, and entertainment—which is why the market is projected to grow at a rate of 18.2% between 2022 and 2033.[3] The journey from the present to the future of human-computer interaction promises to fundamentally alter how we perceive and interact with the digital and physical worlds.

Figure 1

The possibilities of spatial operations

Physical	Bridging	Digital
Wearables (for example, headset, smart eyewear, and pins)	Sensors (for example, LIDAR) and sensor fusion	Augmented reality objects
Next-gen displays	Computer vision	Interactive digital objects
Internet of Things devices (for example, biometric devices)	GPS/spatial mapping software	Holographic projections
Sensory tech (for example, haptic suits)	3D design and rendering tools	Audio outputs
Spatial audio devices	Comprehensive next-gen network infrastructure	Avatars
Cameras	Data lakes	Generative AI
Next-gen batteries		

Source: Abhijith Ravinutala et al., "Dichotomies spatial computing: Navigating towards a better future," Deloitte, April 22, 2024.

Now: Filled to the rim with sims

At its heart, spatial computing brings the digital world closer to lived reality. Many business processes have a physical component, particularly in asset-heavy industries, but, too often, information about those processes is abstracted, and the essence (and insight) is lost. Businesses can learn much about their operations from well-organized, structured business data, but adding physical data can help them understand those operations more deeply. That's where spatial computing comes in.

"This idea of being served the right information at the right time with the right view is the promise of spatial computing," says David Randle, global head of go-to-market for spatial computing at Amazon Web Services (AWS). "We believe spatial computing enables more natural understanding and awareness of physical and virtual worlds."[4]

One of the primary applications unlocked by spatial computing is advanced simulations. Think digital twins, but rather than virtual representations that monitor physical assets, these simulations allow organizations to test different scenarios to see how various conditions will impact their operations.

Imagine a manufacturing company where designers, engineers, and supply chain teams can seamlessly work from a single 3D model to craft, build, and procure all the parts they need; doctors who can view true-to-life simulations of their patients' bodies through augmented reality displays; or an oil and gas company that can layer detailed engineering models on top of 2D maps. The possibilities are as vast as our physical world is varied.

The Portuguese soccer club Benfica's sports data science team uses cameras and computer vision to track players

throughout matches and develop full-scale 3D models of every move its players make. The cameras collect 2,000 data points from each player, and AI helps identify specific players, the direction they were facing, and critical factors that fed into their decision-making. The data essentially creates a digital twin of each player, allowing the team to run simulations of how plays would have worked if a player was in a different position. X's and O's on a chalkboard are now three-dimensional models that coaches can experiment with.[5]

"There's been a huge evolution in AI pushing these models forward, and now we can use them in decision-making," says Joao Copeto, chief information and technology officer at Sport Lisboa e Benfica.[6]

This isn't only about wins and losses—it's also about dollars and cents. Benfica has turned player development into a profitable business by leveraging data and AI. Over the past 10 years, the team has generated some of the highest player-transfer deals in Europe. Similar approaches could also pay dividends in warehouse operations, supply chain and logistics, or any other resource planning process.

Advanced simulations are also showing up in medical settings. For instance, virtual patient scenarios can be simulated as a training supplement for nurses or doctors in a more dynamic, self-paced environment than textbooks would allow. This may come with several challenges, such as patient data concerns, integration of AI into existing learning materials, and the question of realism. But AI-based simulations are poised to impact the way we learn.[7]

Simulations are also starting to impact health care delivery. Fraser Health Authority in Canada has been a pioneer in leveraging simulation models to improve care.[8] By creating a first-of-its-kind system-wide digital twin, the public health authority in British Columbia generated powerful visualizations of patient movement through different care settings and simulations to determine the impact of deploying different care models on patient access. Although the work is ongoing, Fraser expects improvement in appropriate, need-based access to care through increased patient awareness of available services.

New: Data is the differentiator

Enterprise IT teams will likely need to overcome significant hurdles to develop altogether-new spatial computing applications. They likely haven't faced these hurdles when implementing more conventional software-based projects. While these projects have compelling business value, organizations will have to navigate some uncharted waters to achieve them.

For one thing, data isn't always interoperable between systems, which limits the ability to blend data from different sources. Furthermore, the spaghetti diagrams mapping out the path that data travels in most organizations are circuitous at best, and building the data pipelines to get the correct spatial data into visual systems is a thorny engineering challenge. Ensuring that data is of high quality and faithfully mirrors real-world conditions may be one of the most significant barriers to using spatial computing effectively.[9]

Randle of AWS says spatial data has not historically been well managed at most organizations, even though it represents some of a business's most valuable information.

"This information, because it's quite new and diverse, has few standards around it and much of it sits in silos, some of it's in the cloud, most of it's not," says Randle. "This data landscape encompassing physical and digital assets is extremely scattered and not well managed. Our customers' first problem is managing their spatial data."[10]

Taking a more systematic approach to ingesting, organizing, and storing this data, in turn, makes it more available to modern AI tools, and that's where the real learnings begin.

Data pipelines deliver the fuel that drives business

We've often heard that data is the new oil, but for an American oil and gas company, the metaphor is becoming reality thanks to significant effort in replumbing some of its data pipelines.

The energy company uses drones to conduct 3D scans of equipment in the field and its facilities, and then applies

computer vision to the data to ensure its assets operate within predefined tolerances. It's also creating high-fidelity digital twins of assets based on data pulled from engineering, operational, and enterprise resource planning systems.

The critical piece in each example? Data integration. The energy giant built a spatial storage layer, using application program interfaces to connect to disparate data sources and file types, including machine, drone, business, and image and video data.[11]

Few organizations today have invested in this type of systematic approach to ingesting and storing spatial data. Still, it's a key factor driving spatial computing capabilities and an essential first step for delivering impactful use cases.

Multimodal AI creates the context

In the past, businesses couldn't merge spatial and business data into one visualization, but that too is changing. As discussed in **"What's next for AI?"** multimodal AI—AI tools that can process virtually any data type as a prompt and return outputs in multiple formats—is already adept at processing virtually any input, whether text, image, audio, spatial, or structured data types.[12] This capability will allow AI to serve as a bridge between different data sources, and interpret and add context between spatial and business data. AI can reach into disparate data systems and extract relevant insights.

This isn't to say multimodal AI eliminates all barriers. Organizations still need to manage and govern their data effectively. The old saying "garbage in, garbage out" has never been more prescient. Training AI tools on disorganized and unrepresentative data is a recipe for disaster, as AI has the power to scale errors far beyond what we've seen with other types of software. Enterprises should focus on implementing open data standards and working with vendors to standardize data types.

But once they've addressed these concerns, IT teams can open new doors to exciting applications. "You can shape this technology in new and creative ways," says Johan Eerenstein, executive vice president of workforce enablement at Paramount.[13]

Next: AI is the new UI

Many of the aforementioned challenges in spatial computing are related to integration. Enterprises struggle to pull disparate data sources into a visualization platform and render that data in a way that provides value to the user in their day-to-day work. But soon, AI stands to lower those hurdles.

As mentioned above, multimodal AI can take a variety of inputs and make sense of them in one platform, but that could be only the beginning. As AI is integrated into more applications and interaction layers, it allows services to act in concert. As mentioned in **"What's next for AI?"** this is already giving way to agentic systems that are context-aware and capable of executing functions proactively based on user preferences.

These autonomous agents could soon support the roles of supply chain manager, software developer, financial analyst, and more. What will separate tomorrow's agents from today's bots will be their ability to plan ahead and anticipate what the user needs without even having to ask. Based on user preferences and historical actions, they will know how to serve the right content or take the right action at the right time.

When AI agents and spatial computing converge, users won't have to think about whether their data comes from a spatial system, such as LIDAR or cameras (with the important caveat that AI systems are trained on high-quality, well-managed, interoperable data in the first place), or account for the capabilities of specific applications. With intelligent agents, AI becomes the interface, and all that's necessary is to express a preference rather than explicitly program or prompt an application. Imagine a bot that automatically alerts financial analysts to changing market conditions, or one that crafts daily reports for the C-suite about changes in the business environment or team morale.

All the many devices we interact with today, be they phone, tablet, computer, or smart speaker, will feel downright cumbersome in a future where all we have to do is gesture toward a preference and let context-aware, AI-powered systems execute our command. Eventually, once these systems have learned our preferences, we may not even need to gesture at all.

The full impact of agentic AI systems on spatial computing may be many years out, but businesses can still work toward reaping the benefits of spatial computing. Building the data pipelines may be one of the heaviest lifts, but once built, they open up myriad use cases. Autonomous asset inspection, smoother supply chains, true-to-life simulations, and immersive virtual environments are just a few ways leading enterprises are making their operations more spatially aware. As AI continues to intersect with spatial systems, we'll see the emergence of revolutionary new digital frontiers, the contours of which we're only beginning to map out.

Endnotes

1. Abhijith Ravinutala et al., "Dichotomies Spatial Computing: Navigating Towards a Better Future," Deloitte, April 22, 2024.
2. Ibid.
3. Future Market Insights, *Spatial Computing Market Outlook (2022 to 2032)*, October 2022.
4. David Randle (global head of go-to-market, AWS), interview with the author, Sept. 16, 2024.
5. Joao Copeto, chief information and technology officer, Sport Lisboa e Benfica, interview with the author, August 27, 2024.
6. Ibid.
7. Isabelle Bousquette, "Companies finally find a use for virtual reality at work," *The Wall Street Journal*, Sept. 6, 2024.
8. Fraser Health, "Fraser Health Authority: System wide digital twin," October 2023.
9. Gokul Yenduri et al., "Spatial computing: Concept, applications, challenges and future directions," preprint, 10.48550/arXiv.2402.07912 (2024).
10. Randle interview.
11. Deloitte internal information.
12. George Lawton, "Multimodal AI," TechTarget, accessed Oct. 29, 2024.
13. Johan Eerenstein (senior vice president of workforce enablement, Paramount), interview with the author, July 16, 2024.

Continue the conversation

Industry leadership

Frances Yu
Unlimited Reality™ GM/Business lead | Principal | Deloitte Consulting LLP
+1 312 486 2563 | francesyu@deloitte.com

Frances Yu is a partner at Deloitte Consulting LLP, where she has served in a range of global practice leadership roles. She has helped Fortune 500 clients as well as Deloitte launch several new ventures, evolved growth strategies, and transformed their demand value chain. Currently, she is the US and global business lead and general manager for Deloitte's Unlimited Reality™, a multinetwork innovation business for the industrial metaverse era, focusing on spatial computing, digital twin, and multimodal AI and data.

Nishanth Raj
Unlimited Reality™ Spatial/Multimodal AI and data lead | Managing director | Deloitte Consulting LLP
+1 832 970 7560 | nisraj@deloitte.com

Nishanth Raj is a managing director and AI and data / Unlimited Reality™ leader at Deloitte Consulting, specializing in the Energy & Chemicals sector. With over two decades of consulting experience, he helps clients leverage technology, AI, and data to drive business value, and transform them into insights-driven organizations.

Stefan Kircher
Unlimited Reality™ CTO | Managing director | Deloitte Consulting LLP
+1 404 631 2541 | skircher@deloitte.com

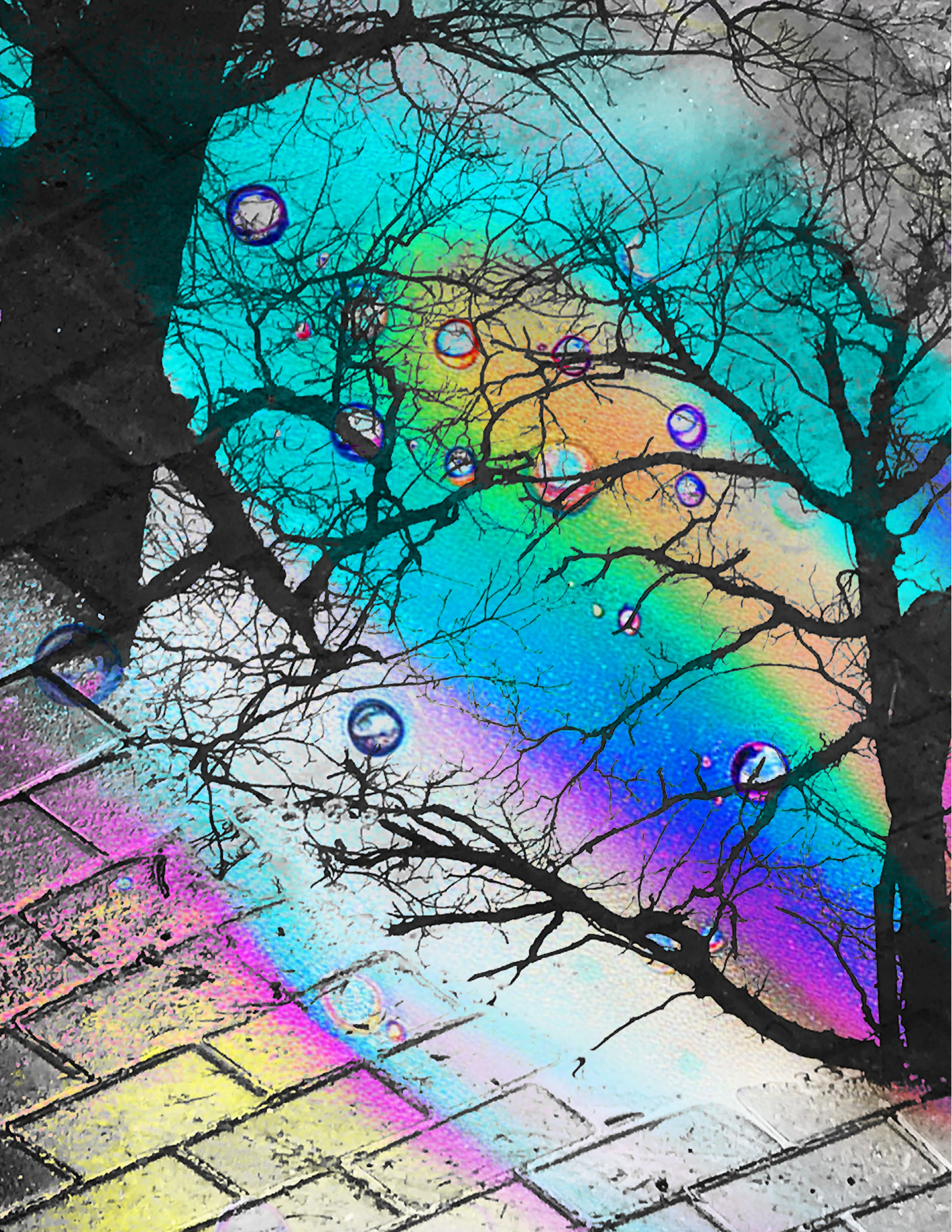
Stefan Kircher is a managing director in the Products & Solutions practice of Deloitte Consulting LLP and CTO for Deloitte's Unlimited Reality™ Business. He has over 25 years expertise in the industry, technology strategy, and solution-building across various industries, R&D, innovation, and partnerships with strategic tech partners like AWS.

Robert Tross
Unlimited Reality™ GPS market offering leader | Principal | Deloitte Consulting LLP
+1 703 251 1250 | rtross@deloitte.com

Robert Tross is a principal in Deloitte Consulting LLP's GPS Government Technology practice, leading the Unlimited Reality™ federal market offering. With over 25 years of experience, he specializes in omni-channel experiences across various platforms, including web, immersive/spatial, social media, mobile, wearables, and tablets, including others.

Acknowledgments

Much gratitude goes to the many subject matter leaders across Deloitte that contributed to our research for the Interaction chapter: **Lars Cromley**, **Stefan Kircher**, **Kaitlyn Kuczer, Lena La**, **Tim Murphy**, **Ali Newman**, **Bob Tross**, and **Frances Yu.**



INFORMATION

What's next for AI?

While large language models continue to advance, new models and agents are proving to be more effective at discrete tasks. AI needs different horses for different courses.

Kelly Raskovich, Bill Briggs, Mike Bechtel, and Abhijith Ravinutala

Blink and you'll miss it: The speed of artificial intelligence's advancement is outpacing expectations. Last year, as organizations scrambled to understand how to adopt generative AI, we cautioned Tech Trends 2024 readers to lead with need as they differentiate themselves from competitors and adopt a strategic approach to scaling their use of large language models (LLMs). Today, LLMs have taken root, with up to 70% of organizations, by some estimates, actively exploring or implementing LLM use cases.[1]

But leading organizations are already considering AI's next chapter. Instead of relying on foundation models built by large players in AI, which may be more powerful and built on more data than needed, enterprises are now thinking about implementing multiple, smaller models that can be more efficient for business requirements.[2] LLMs will continue to advance and be the best option for certain use cases, like general-purpose chatbots or simulations for scientific research, but the chatbot that peruses your financial data to think through missed revenue opportunities doesn't need to be the same model that replies to customer inquiries. Put simply, we're likely to see a proliferation of different horses for different courses.

A series of smaller models working in concert may end up serving different use cases than current LLM approaches. New open-source options and multimodal outputs (as opposed to just text) are enabling organizations to unlock entirely new offerings.[3]

In the years to come, the progress toward a growing number of smaller, more specialized models could once again move the goalposts of AI in the enterprise. Organizations may witness a fundamental shift in AI from augmenting knowledge to augmenting execution. Investments being made today in *agentic AI*, as this next era is termed, could upend the way we work and live by arming consumers and businesses with armies of silicon-based assistants. Imagine AI agents that can carry out discrete tasks, like delivering a financial report in a board meeting or applying for a grant. "There's an app for that" could well become "There's an agent for that."

Now: Getting the fundamentals right

LLMs are undoubtedly exciting but require a great deal of groundwork. Instead of building models themselves, many enterprises are partnering with companies like Anthropic or OpenAI or accessing AI models through hyperscalers.[4] According to Gartner®, AI servers will account for close to 60% of hyperscalers' total server spending.[5] Some enterprises have found immediate business value in using LLMs, while others have remained wary about the accuracy and applicability of LLMs trained on external data.[6] On an enterprise time scale, AI advancements are still in a nascent phase (crawling or walking, as we noted last year). According to recent surveys by Deloitte and Fivetran and Vanson Bourne, in most organizations, fewer than a third of generative AI experiments have moved into production, often because organizations struggle to access or cleanse all the data needed to run AI programs.[7] To achieve scale, organizations will likely need to further think through data and technology, as well as strategy, process, and talent, as outlined in a recent Deloitte AI Institute report.

According to Deloitte's 2024 State of Generative AI in the Enterprise Q3 report, 75% of surveyed organizations have increased their investments in data-life-cycle management due to generative AI.[8] Data is foundational to LLMs, because bad inputs lead to worse outputs (in other words, garbage in, garbage squared). That's why data-labeling costs can be a big driver of AI investment.[9] While some AI companies scrape the internet to build the largest models possible, savvy enterprises create the *smartest* models possible, which requires better domain-specific "education" for their LLMs. For instance, LIFT Impact Partners, a Vancouver-based organization that provides resources to nonprofits, is fine-tuning its AI-enabled virtual assistants on appropriate data to help new Canadian immigrants process paperwork. "When you train it on your organization's unique persona, data, and culture, it becomes significantly more relevant and effective," says Bruce Dewar, president and CEO of LIFT Impact Partners. "It brings authenticity and becomes a true extension of your organization."[10]

Data enablement issues are dynamic. Organizations surveyed by Deloitte said new issues could be exposed by the scale-up of AI pilots, unclear regulations around sensitive data, and questions around usage of external data (for example, licensed third-party data). That's why 55% of organizations surveyed avoided certain AI use cases due to data-related issues, and an equal proportion are working to enhance their data security.[11] Organizations could work around these issues by using out-of-the-box models offered by vendors, but differentiated AI impact will likely require differentiated enterprise data.

Thankfully, once the groundwork is laid, the benefits are clear: Two-thirds of organizations surveyed say they're increasing investments in generative AI because they've seen strong value to date.[12] Initial examples of real-world value are also appearing across industries, from insurance claims review to telecom troubleshooting and consumer segmentation tools.[13] LLMs are also making waves in more specialized use cases, such as space repairs, nuclear modeling, and material design.[14]

As underlying data inputs improve and become more sustainable, LLMs and other advanced models (like simulations) may become easier to spin up and scale. But size isn't everything. Over time, as methods for AI training and implementation proliferate, organizations are likely to pilot smaller models. Many may have data that can be more valuable than previously imagined, and putting it into action through smaller, task-oriented models can reduce time, effort, and hassle. We're poised to move from large-scale AI projects to AI everywhere, as discussed in this year's introduction.

New: Different horses for different courses

While LLMs have a vast array of use cases, the library is not infinite (yet). LLMs require massive resources, deal primarily with text, and are meant to augment human intelligence rather than take on and execute discrete tasks. As a result, says Vivek Mohindra, senior vice president of corporate strategy at Dell Technologies, "there is no one-size-fits-all approach to AI. There are going to be models of all sizes and purpose-built options—that's one of our key beliefs in AI strategy."[15]

Over the next 18 to 24 months, key AI vendors and enterprise users are likely to have a toolkit of models comprising increasingly sophisticated, robust LLMs along with other models more applicable to day-to-day use cases. Indeed, where LLMs are not the optimal choice, three pillars of AI are opening new avenues of value: small language models, multimodal models, and agentic AI (figure 1).

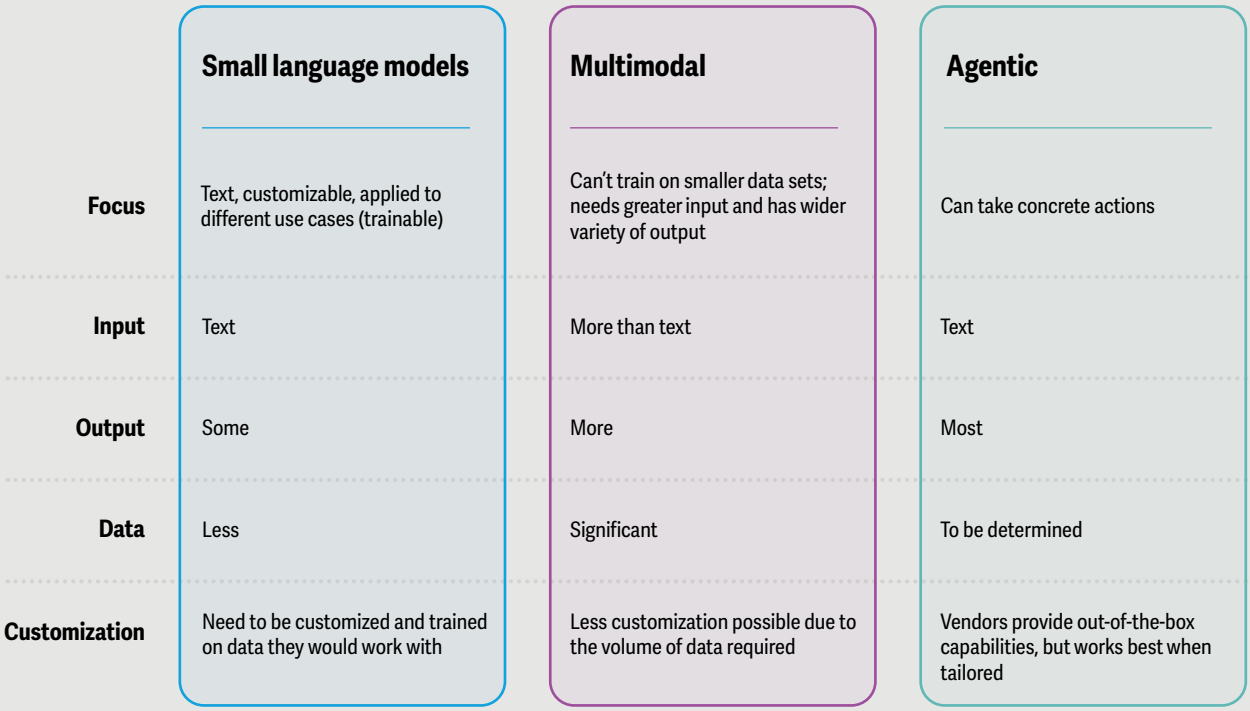
Small language models

LLM providers are racing to make AI models as efficient as possible. Instead of enabling new use cases, these efforts aim to rightsize or optimize models for existing use cases. For instance, massive models are not necessary for mundane tasks like summarizing an inspection report—a smaller model trained on similar documents would suffice and be more cost-efficient.

Small language models (SLMs) can be trained by enterprises on smaller, highly curated data sets to solve more specific problems, rather than general queries. For example, a company could train an SLM on its inventory information, enabling employees to quickly retrieve insights instead of manually parsing large data sets, a process that can sometimes take weeks. Insights from such an SLM could then be coupled with a user interface application for easy access.

Figure 1

Different AI for different needs



Source: Deloitte research.

Naveen Rao, vice president of AI at Databricks, believes more organizations will take this systems approach with AI: "A magic computer that understands everything is a sci-fi fantasy. Rather, in the same way we organize humans in the workplace, we should break apart our problems. Domain-specific and customized models can then address specific tasks, tools can run deterministic calculations, and databases can pull in relevant data. These AI systems deliver the solution better than any one component could do alone."[16]

An added benefit of smaller models is that they can be run on-device and trained by enterprises on smaller, highly curated data sets to solve more specific problems, rather than general queries, as discussed in "Hardware is eating the world." Companies like Microsoft and Mistral are currently working to distill such SLMs, built on fewer parameters, from their larger AI offerings, and Meta offers multiple options across smaller models and frontier models.[17]

Finally, much of the progress happening in SLMs is through open-source models offered by companies like Hugging Face or Arcee.AI.[18] Such models are ripe for enterprise use since they can be customized for any number of needs, as long as IT teams have the internal AI talent to fine-tune them. In fact, a recent Databricks report indicates that over 75% of organizations are choosing smaller open-source models and customizing them for specific use cases.[19] Since open-source models are constantly improving thanks to the contributions of a diverse programming community, the size and efficiency of these models are likely to improve at a rapid clip.

Multimodal models

Humans interact through a variety of mediums: text, body language, voice, videos, among others. Machines are now hoping to catch up.[20] Given that business needs are not contained to text, it's no surprise that companies are looking forward to AI that can take in and produce multiple mediums. In some ways, we're already accustomed to multimodal AI, such as when we speak to digital assistants and receive text or images in return, or when we ride in cars that use a mix of computer vision and audio cues to provide driver assistance.[21]

Multimodal generative AI, on the other hand, is in its early stages. The first major models, Google's Project Astra and OpenAI's GPT-4 Omni, were showcased in May 2024, and Amazon Web Services' Titan offering has similar capabilities.[22] Progress in multimodal generative AI may be slow because it requires significantly higher amounts of data, resources, and hardware.[23] In addition, the existing issues of hallucination and bias that plague text-based models may be exacerbated by multimodal generation.

Still, the enterprise use cases are promising. The notion of "train once, run anywhere (or any way)" promises a model that could be trained on text, but deliver answers in pictures, video, or sound, depending on the use case and the user's preference, which improves digital inclusion. Companies like AMD aim to use the fledgling technology to quickly translate marketing materials from English to other languages or to generate content.[24] For supply chain optimization, multimodal generative AI can be trained on sensor data, maintenance logs, and warehouse images to recommend ideal stock quantities.[25] This also leads to new opportunities with spatial computing, which we write about in **"Spatial computing takes center stage."** As the technology progresses and model architecture becomes more efficient, we can expect to see even more use cases in the next 18 to 24 months.

Agentic AI

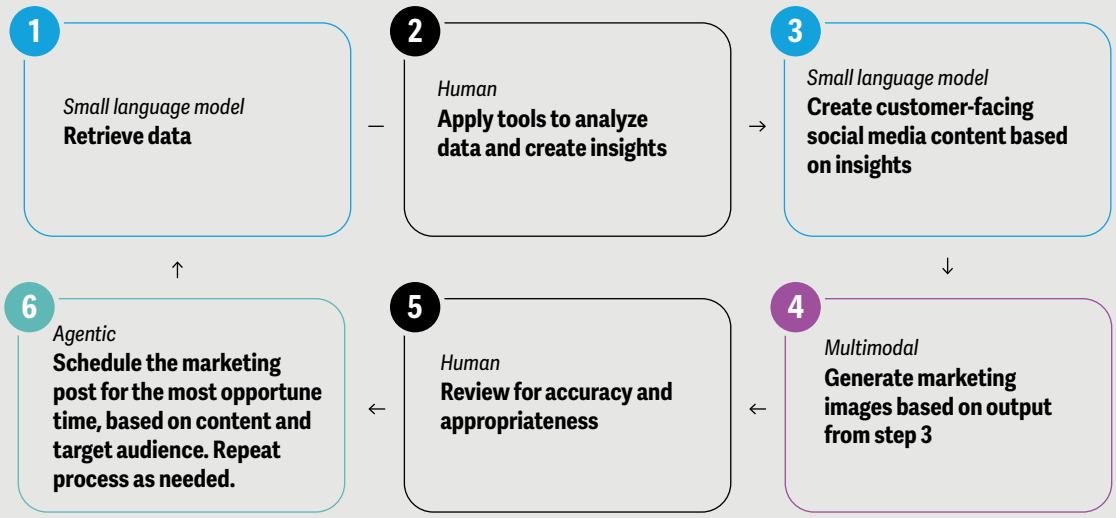
The third new pillar of AI may pave the way for changes to our ways of working over the next decade. Large (or small) action models go beyond the question-and-answer capabilities of LLMs and complete discrete tasks in the real world. Examples range from booking a flight based on your travel preferences to providing automated customer support that can access databases and execute needed tasks—likely without the need for highly specialized prompts.[26] The proliferation of such action models, working as autonomous digital agents, heralds the beginnings of agentic AI, and enterprise software vendors like Salesforce and ServiceNoware already touting these possibilities.[27]

Chris Bedi, chief customer officer at ServiceNow, believes that domain- or industry-specific agentic AI can change the game for humans and machine interaction in enterprises.[28] For instance, in the company's Xanadu platform, one AI agent can scan incoming customer issues against a history of incidents to come up with a recommendation for next steps. It then communicates to another autonomous agent that's able to execute on those recommendations, and a human in the loop reviews those agent-to-agent communications to approve the hypotheses. In the same vein, one agent might be adept at managing workloads in the cloud, while another provisions orders for customers. As Bedi says, "Agentic AI cannot completely take the place of a human, but what it can do is work alongside your teams, handling repetitive tasks, seeking out information and resources, doing work in the background 24/7, 365 days a year."[29]

Finally, aside from the different categories of AI models noted above, advancements in AI design and execution can also impact enterprise adoption—namely, the advent of liquid neural networks. "Liquid" refers to the flexibility in this new form of training AI through a neural network, a machine learning algorithm that mimics the human brain's structure. Similar to how **quantum computers are freed from the binary nature** of classical computing, liquid neural networks can do more with less: A couple dozen nodes in the network might suffice, versus 100,000 nodes in a more traditional network. The cutting-edge technology aims to run on less computing power, with more transparency, opening up possibilities for embedding AI into edge devices, robotics, and safety-critical systems.[30] In other words, it's not just the applications of AI but also its underlying mechanisms that are ripe for improvement and disruption in the coming years.

Figure 2

Compound AI journey



Source: Deloitte research.

Next: There's an agent for that

In the next decade, AI could be wholly focused on execution instead of human augmentation. A future employee could make a plain-language request to an AI agent, for example, "close the books for Q2 and generate a report on EBITDA." Like in an enterprise hierarchy, the primary agent would then delegate the needed tasks to agents with discrete roles that cascade across different productivity suites to take action. As with humans, teamwork could be the missing ingredient that enables the machines to improve their capabilities.[31] This leads to a few key considerations for the years to come (figure 2):

- **AI-to-AI communication.** Agents will likely have a more efficient way of communicating with each other than human language, as we don't need human-imitating chatbots talking to each other.[32] Better AI-to-AI communication can enhance outcomes, as fewer people will need to become experts to benefit from AI. Rather, AI can adapt to each person's communication style.[33]
- **Job displacement and creation.** Some claim that roles such as prompt engineer could become obsolete.[34] However, the AI expertise of those employees will remain pertinent as they focus on managing, training, and collaborating with AI agents as they do with LLMs today. For example, a lean IT team with AI experts might build the agents it needs in a sort of "AI factory" for the enterprise. The significant shift in the remaining workforce's skills and education may ultimately reward more human skills like creativity and design, as mentioned in previous Tech Trends.
- **Privacy and security.** The proliferation of agents with system access is likely to raise broad concerns about cybersecurity, which will only become more important as time progresses and more of our data is accessed by AI systems. New paradigms for risk and trust will be required to make the most out of applying AI agents.

- **Energy and resources.** AI's energy consumption is a growing concern.[35] To mitigate environmental impacts, future AI development will need to balance performance with sustainability. It will need to take advantage of improvements in liquid neural networks or other efficient forms of training AI, not to mention the hardware needed to make all of this work, as we discuss in **"Hardware is eating the world."**

- **Leadership for the future.** AI has transformative potential, as everyone has heard plenty over the last year, but only insofar as leadership allows. Applying AI as a faster way of doing things the way they've always been done will result in, at best, missed potential, and, at worst, amplified biases.[36] Imaginative, courageous leaders should dare to take AI from calcified best practices to the creation of "next practices," where we find new ways of organizing ourselves and our data toward an AI-enabled world.

When it comes to AI, enterprises will likely have the same considerations in the future that they do today: data, data, and data. Until AI systems can reach **artificial general intelligence** or learn as efficiently as the human brain,[37] they will be hungry for more data and inputs to help them be more powerful and accurate. Steps taken today to organize, streamline, and protect enterprise data could pay dividends for years to come, as data debt could one day become the biggest portion of technical debt. Such groundwork should also help enterprises prepare for the litany of regulatory challenges and ethical uncertainties (such as data collection and use limitations, fairness concerns, lack of transparency) that come with shepherding this new, powerful technology into the future.[38] The stakes of garbage in, garbage out are only going to grow: It would be much better to opt for genius in, genius squared.[39]

Endnotes

1. Carl Franzen, "More than 70% of companies are experimenting with generative AI, but few are willing to commit more spending," *VentureBeat*, July 25, 2023.
2. Tom Dotan and Deepa Seetharaman, "For AI giants, smaller is sometimes better," *The Wall Street Journal*, July 6, 2024.
3. Google Cloud, "Multimodal AI," accessed October 2024.
4. Silvia Pellegrino, "Which companies have partnered with OpenAI?," *Tech Monitor*, May 15, 2023; Maxwell Zeff, "Anthropic launches Claude Enterprise plan to compete with OpenAI," *TechCrunch*, September 4, 2024; Jean Atelsek and William Fellows, "Hyperscalers stress AI credentials, optimization and developer empowerment," *S&P Global Market Intelligence*, accessed October 2024.
5. Gartner, "Gartner forecasts worldwide IT spending to grow 8% in 2024," press release, April 17, 2024. GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.
6. Patricia Licatta, "Between sustainability and risk: Why CIOs are considering small language models," *CIO*, August 1, 2024.
7. Jim Rowan et al., "Now decides next: Moving from potential to performance," Deloitte's State of Generative AI in the Enterprise Q3 report, August 2024; Mark Van de Wiel, "New AI survey: Poor data quality leads to $406 million in losses," Fivetran, March 20, 2024.
8. Rowan et al., "Now decides next: Moving from potential to performance."
9. Sharon Goldman, "The hidden reason AI costs are soaring—and it's not because Nvidia chips are more expensive," *Fortune*, August 23, 2024.
10. *Deloitte Insights*, "Lifting up the nonprofit sector through generative AI," September 23, 2024.
11. Jim Rowan et al., "Now decides next: Moving from potential to performance."
12. Ibid.
13. Ibid.
14. Sandra Erwin, "Booz Allen deploys advanced language model in space," *SpaceNews*, August 1, 2024; Argonne National Laboratory, "Smart diagnostics: How Argonne could use Generative AI to empower nuclear plant operators," press release, July 26, 2024; Kevin Maik Jablonka et al., "14 examples of how LLMs can transform materials science and chemistry: A reflection on a large language model hackathon," *Digital Discovery* 5 (2023).
15. Phone interview with Vivek Mohindra, senior vice president of corporate strategy, Dell Technologies, October 11, 2024.
16. Phone interview with Naveen Rao, vice president of AI at Databricks, October 2, 2024.
17. YouTube, "Introducing the next evolution of generative AI: Small language models," Microsoft Dynamics 365, video, May 9, 2024; Llama team, "The Llama 3 herd of models," Meta, July 23, 2024.
18. Rachel Metz, "In AI, smaller, cheaper models are getting big attention," Bloomberg, August 8, 2024.
19. Databricks, "AI is in production," accessed October 2024.
20. *MIT Technology Review Insights*, "Multimodal: AI's new frontier," May 8, 2024.
21. Akesh Takyar, "Multimodal models: Architecture, workflow, use cases and development," LeewayHertz, accessed October 2024.
22. NeuronsLab, "Multimodal AI use cases: The next opportunity in enterprise AI," May 30, 2024.
23. Ellen Glover, "Multimodal AI: What it is and how it works," Built In, July 1, 2024.
24. Mary E. Morrison, "At AMD, opportunities, challenges of using AI in marketing," Deloitte's *CIO Journal* for *The Wall Street Journal*, July 2, 2024.
25. NeuronsLab, "Multimodal AI use cases: The next opportunity in enterprise AI."
26. Oguz A. Acar, "AI prompt engineering isn't the future," *Harvard Business Review*, June 6, 2023.
27. Salesforce, "Agentforce," accessed October 2024; ServiceNow, "Our biggest AI release is here," accessed October 2024.
28. Phone interview with Chris Bedi, chief customer officer at ServiceNow, September 30, 2024.
29. Ibid.
30. Brian Heater, "What is a liquid neural network, really?," *TechCrunch*, August 17, 2023.
31. Edd Gent, "How teams of AI agents working together could unlock the tech's true power," *Singularity Hub*, June 28, 2024.
32. Will Knight, "The chatbots are now talking to each other," *WIRED*, October 12, 2023.
33. David Ellis, "The power of AI in modeling healthy communications," *Forbes*, August 17, 2023.
34. Acar, "AI prompt engineering isn't the future."
35. James Vincent, "How much electricity does AI consume?," *The Verge*, February 16, 2024.
36. IBM, "Shedding light on AI bias with real world examples," October 16, 2023.
37. University of Oxford, "Study shows that the way the brain learns is different from the way that artificial intelligence systems learn," January 3, 2024.
38. Nestor Maslej et al., *The AI Index 2024 annual repor*t, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024.
39. Deloitte, Work Re-Architected video series, accessed October 2024.

Continue the conversation

Industry leadership

Jim Rowan

Head of AI | Principal | Deloitte Consulting LLP
Jimrowan@deloitte.com | +1 617 437 3470

Jim Rowan is a principal at Deloitte and is currently the Head of AI for Deloitte. He helps clients transform their businesses using data powered analytical and AI solutions that enable better decision making. Over the course of his career, Rowan has served clients across the life sciences, health care, and telecommunications industries. He also has deep knowledge of the finance function in these organizations, having led analytics, planning and forecasting, and close projects that enable the finance function to embrace digital transformations. Rowan formerly led AI & Data Operations within Deloitte Consulting's Strategy & Analytics practice.

Nitin Mittal

Global AI leader | Principal | Deloitte Consulting LLP

Nitin Mittal is a principal with Deloitte Consulting LLP. He currently serves as the US Artificial Intelligence (AI) Strategic Growth Offering Consulting leader and the Global Strategy, Analytics and M&A leader. He is the 2019 recipient of the AI Innovator of the Year award at the AI Summit New York. He specializes in advising clients to achieve competitive advantage through data and cognitive powered transformations that promote amplified intelligence and enable our clients to make strategic choices and transform ahead of disruption.

Throughout his career, Mittal has served as a trusted advisor to global clients and has worked across a number of industry sectors. His primary focus has been working with life sciences and health care clients, implementing large scale data programs that promote organizational intelligence, and the use of advanced analytics and AI to drive insights and business strategy.

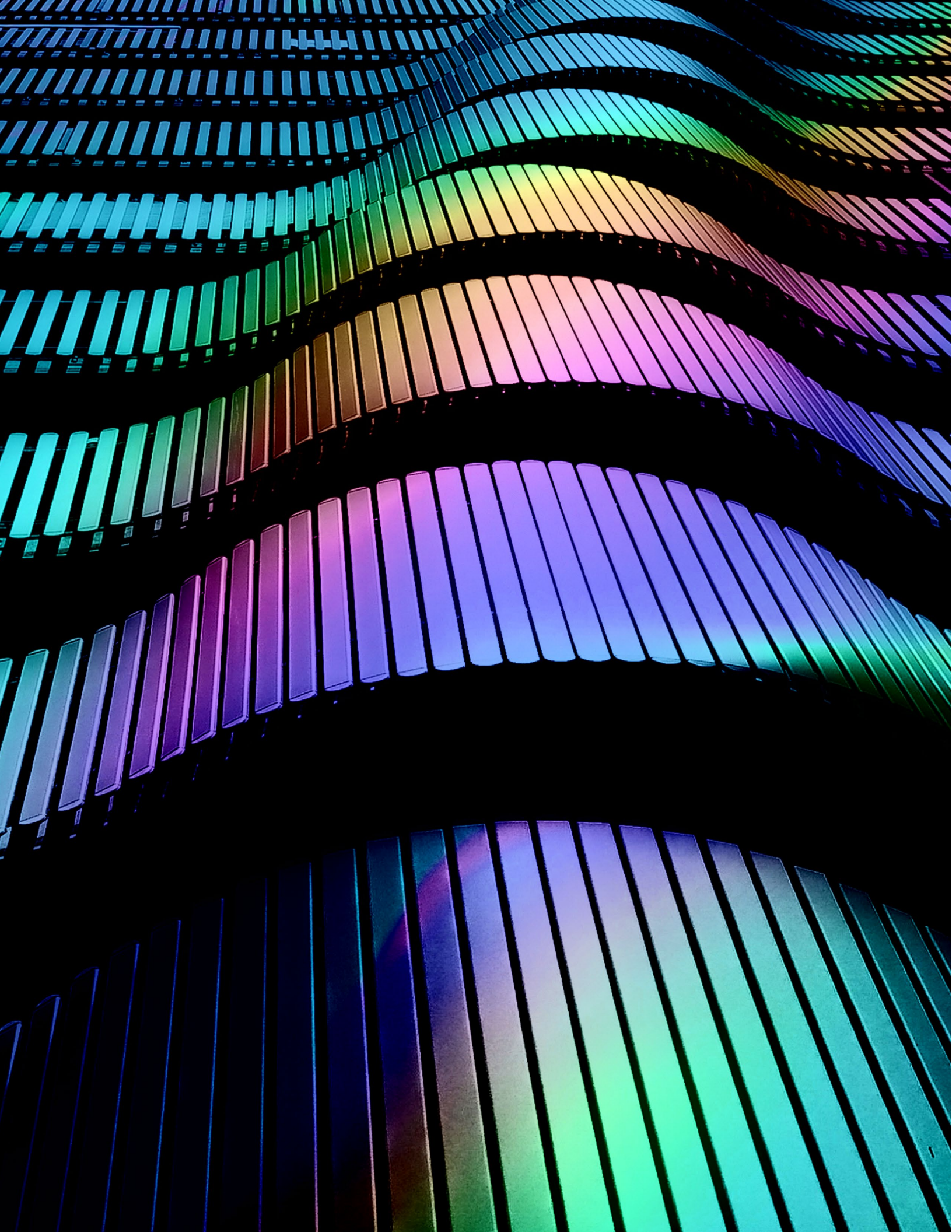
Lou DiLorenzo Jr

Principal | AI & Data Strategy Practice leader | US CIO & CDAO Programs, national leader | Deloitte Consulting LLP
+1 612 397 4000 | ldilorenzojr@deloitte.com

Lou DiLorenzo serves as the national leader of Deloitte Consulting's AI & Data Strategy practice and the Deloitte US CIO and CDAO Executive Accelerator programs. He is a member of Deloitte's Generative AI practice leadership team and heads the Generative AI Incubator. With over 20 years of cross-sector operating, entrepreneurial, and consulting experience, he has a successful record of bringing key stakeholders together to help lead change, develop new capabilities, and deliver positive financial results. Previously, DiLorenzo served as COO of a consumer health insurance startup and as Global CIO for the Food Ingredients & Bio Industrial division at Cargill. He is a frequent technology contributor to leading publications and hosts the podcast, *Techfluential*.

Acknowledgments

Much gratitude goes to the many subject matter leaders across Deloitte that contributed to our research for the information chapter: **Lou DiLorenzo**, **Lena La**, **Nitin Mittal**, **Sanghamitra Pati**, **Jim Rowan**, and **Baris Sarer**.



COMPUTATION

Hardware is eating the world

The AI revolution will demand heavy energy and hardware resources—making enterprise infrastructure a strategic differentiator once again

Kelly Raskovich, Bill Briggs, Mike Bechtel, and Abhijith Ravinutala

After years of "software eating the world," it's hardware's turn to feast. We previewed in the computation chapter of Tech Trends 2024 that as Moore's Law comes to its supposed end, the promise of the AI revolution increasingly depends on access to the appropriate hardware. Case in point: NVIDIA is now one of the world's most valuable (and watched) companies, as specialized chips become an invaluable resource for AI computation workloads.[1] According to Deloitte research based on a World Semiconductor Trade Statistics forecast, the market for chips used only for generative AI is projected to reach over US$50 billion this year.[2]

A critical hardware use case for enterprises may lie in AI-embedded end-user and edge devices. Take personal computers (PCs), for instance. For years, enterprise laptops have been commodified. But now, we may be on the cusp of a significant shift in computing, thanks to AI-embedded PCs. Companies like AMD, Dell, and HP are already touting the potential for AI PCs to "future-proof" technology infrastructure, reduce cloud computing costs, and enhance data privacy.[3] With access to offline AI models for image generation, text analysis, and speedy data retrieval, knowledge workers could be supercharged by faster, more accurate AI. That being said, enterprises should be strategic about refreshing end-user computation on a large scale—there's no use wasting AI resources that are limited in supply.

Of course, all of these advancements come at a cost. Data centers are a new focus of sustainability as the energy demands of large AI models continue to grow.[4] The International Energy Agency has suggested that the demands of AI will significantly increase electricity in data centers by 2026, equivalent to Sweden's or Germany's annual energy demands.[5] A recent Deloitte study on powering AI estimates that global data center electricity consumption may triple in the coming decade, largely due to AI demand.[6] Innovations in energy sources and efficiency are needed to make AI hardware more accessible and sustainable, even as it proliferates and finds its way into everyday consumer and enterprise devices. Consider that Unit 1 of the nuclear plant Three Mile Island, which was shut down five years ago due to economic reasons, will reopen by 2028 to power data centers with carbon-free electricity.[7]

Looking forward, AI hardware is poised to step beyond IT and into the Internet of Things. An increasing number of smart devices could become even more intelligent as AI enables them to analyze their usage and take on new tasks (as agentic AI, mentioned in "What's next for AI?" advances). Today's benign use cases (like AI in toothbrushes) are not indicative of tomorrow's robust potential (like AI in lifesaving medical devices).[8] The true power of hardware could be unlocked when smarter devices bring about a step change in our relationship with robotics.

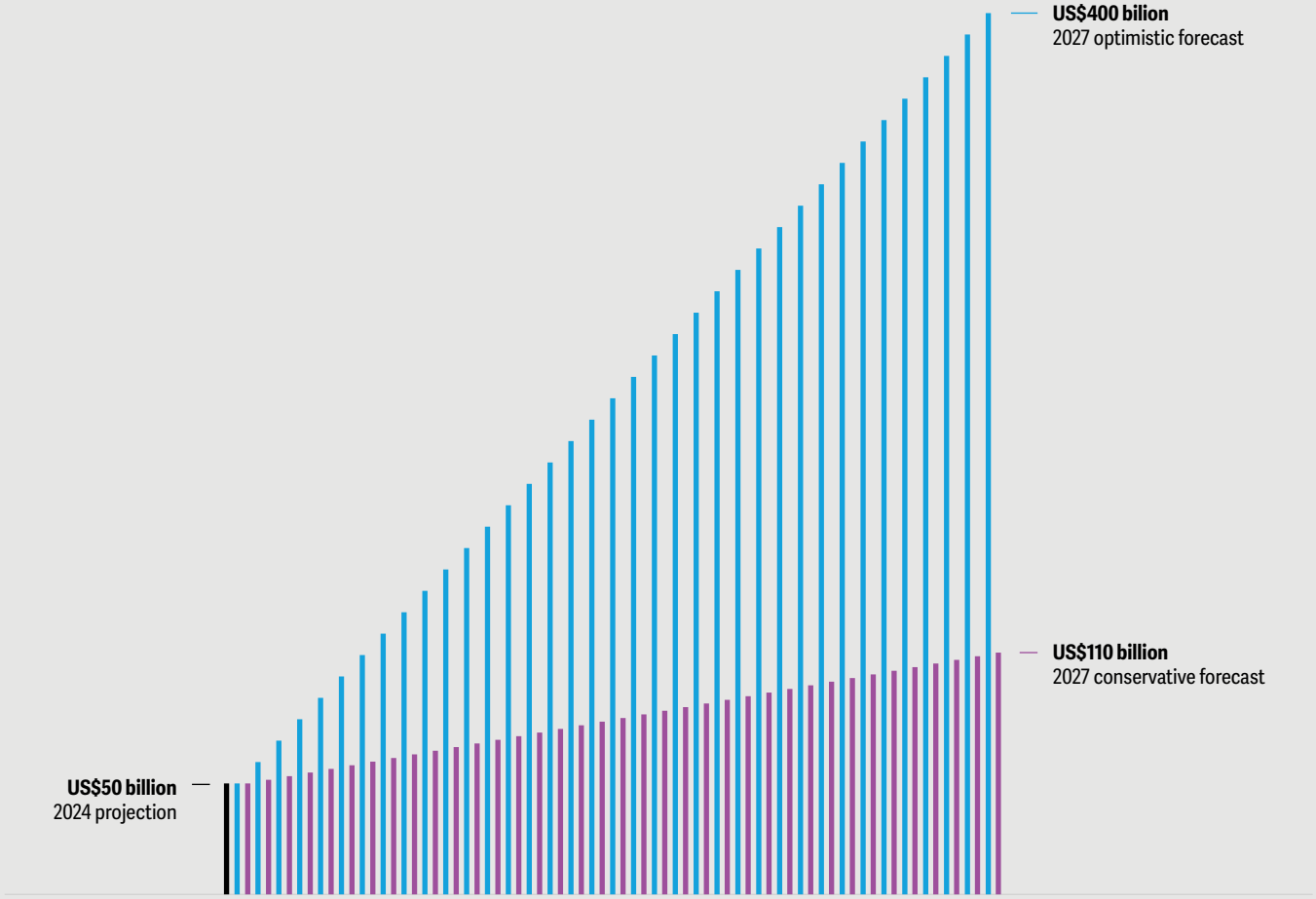
Now: Chips ahoy!

A generation of technologists has been taught to believe software is the key to return on investment, given its scalability, ease of updates, and intellectual property protections.[9] But now, hardware investment is surging as computers evolve from calculators to cogitators.[10] We wrote last year that specialized chips like graphics-processing units (GPUs) were becoming the go-to resources for training AI models. In its 2024 TMT Predictions report, Deloitte estimated that total AI chip sales in 2024 would be 11% of the predicted global chip market of

Figure 1

The surge in AI hardware investment

AI chip market forecasts



Source: Duncan Stewart et al., "Gen AI chip demand fans a semi tailwind ... for now," *Deloitte Insights*, November 29, 2023.

US$576 billion.[11] Growing from roughly $US50 billion today, the AI chip market is forecasted to reach up to US$400 billion by 2027, though a more conservative estimate is US$110 billion (figure 1).[12]

Large tech companies are driving a portion of this demand, as they may build their own AI models and deploy specialized chips on-premises.[13] However, enterprises across industries are seeking compute power to meet their IT goals. For instance, according to a Databricks report, the financial services industry has had the highest growth in GPU usage, at 88% over the past six months, in running large language models (LLMs) that tackle fraud detection and wealth management.[14]

All of this demand for GPUs has outpaced capacity. In today's iteration of the Gold Rush, the companies providing "picks and shovels," or the tools for today's tech transformation, are winning big.[15] NVIDIA's CEO Jensen Huang has noted that cloud GPU capacity is mostly

filled, but the company is also rolling out new chips that are significantly more energy-efficient than previous iterations.[16] Hyperscalers are buying up GPUs as they roll off the production line, spending almost $US1 trillion on data center infrastructure to accommodate the demand from clients who rent GPU usage.[17] All the while, the energy consumption of existing data centers is pushing aging power grids to the brink globally.[18]

Understandably, enterprises are looking for new solutions. While GPUs are crucial for handling the high workloads of LLMs or content generation, and central processing units are still table stakes, neural processing units (NPUs) are now in vogue. NPUs, which mimic the brain's neural network, can accelerate smaller AI workloads with greater efficiency and lower power demands,[19] enabling enterprises to shift AI applications away from the cloud and apply AI locally to sensitive data that can't be hosted externally.[20] This new breed of chip is a crucial part of the future of embedded AI.

Vivek Mohindra, senior vice president of corporate strategy at Dell Technologies, says, "Of the 1.5 billion PCs in use today, 30% are four years old or more. None of these older PCs have NPUs to take advantage of the latest AI PC advancements."[21] A great refresh of enterprise hardware may be on the horizon. As NPUs enable end-user devices to run AI offline and allow models to become smaller to target specific use cases, hardware may once again be a differentiator for enterprise performance. In a recent Deloitte study, 72% of respondents believe generative AI's impact on their industry will be "high to transformative."[22] Once AI is at our fingertips thanks to mainstream hardware advancements, that number may edge closer to 100%.

New: Infrastructure is strategic again

The heady cloud-computing highs of assumed unlimited access are giving way to a resource-constrained era. After being relegated to a utility for years, enterprise infrastructure (for example, PCs) is once again strategic. Specifically, specialized hardware will likely be crucial to three significant areas of AI growth: AI-embedded devices and the Internet of Things, data centers, and advanced physical robotics. While the impact on robotics may occur over the next few years, as we discuss in the next section, we anticipate that enterprises will be grappling with decisions about the first two areas over the next 18 to 24 months. While AI scarcity and demand persist, the following areas may differentiate leaders from laggards.

Edge footprint

By 2025, more than 50% of data could be generated by edge devices.[23] As NPUs proliferate, more and more devices could be equipped to run AI models without relying on the cloud. This is especially true as generative AI model providers opt for creating smaller, more efficient models for specific tasks, as discussed in **"What's next for AI?"** With quicker response times, decreased costs, and greater privacy controls, hybrid computing (that is, a mix of cloud and on-device AI workloads) could be a must-have for many enterprises, and hardware manufacturers are betting on it.[24]

According to Dell Technologies' Mohindra, processing AI at the edge is one of the best ways to handle the vast amounts of data required. "When you consider latency, network resources, and just sheer volume, moving data to a centralized compute location is inefficient, ineffective, and not secure," he says. "It's better to bring AI to the data, rather than bring the data to AI."[25]

One major bank predicts that AI PCs will account for more than 40% of PC shipments in 2026.[26] Similarly, nearly 15% of 2024 smartphone shipments are predicted to be capable of running LLMs or image-generation models.[27] Alex Thatcher, senior director of AI PC experiences and cloud clients at HP, believes that the refresh in devices will be akin to the major transition from command-line inputs to graphical user interfaces that changed PCs in the 1990s. "The software has fundamentally changed, replete with different tools and ways of collaborating," he says. "You need hardware that can accelerate that change and make it easier for enterprises to create and deliver AI solutions."[28] Finally, Apple and Microsoft have also fueled the impending hardware refresh by embedding AI into their devices this year.[29]

As choices proliferate, good governance will be crucial, and enterprises have to ask the question: How many of our people need to be armed with next-generation devices? Chip manufacturers are in a race to improve AI horsepower,[30] but enterprise customers can't afford to refresh their entire edge footprint with each new

advancement. Instead, they should develop a strategy for tiered adoption where these devices can have the most impact.

Build versus buy

For buying or renting specialized hardware, organizations may typically consider their cost model over time, the expected time frame of use, and the necessity for progress. However, AI is applying another level of competitive pressure to this decision. With hardware like GPUs still scarce and the market clamoring for AI updates from all organizations, many companies have been tempted to rent as much computing power as possible.

Organizations may struggle to take advantage of AI if they don't have their data enablement in order. Rather than scrambling for GPUs, it may be more efficient to understand where the organization is ready for AI. Some areas may concern private or sensitive data; investing in NPUs can keep those workloads offline, while others may be fine for the cloud. Thanks to the lessons of cloud in the past decade, enterprises know that the cost of runaway models operating on runaway hardware can quickly balloon.[31] Pushing these costs to operating expenditure may not be the best answer.

Some estimates even say that GPUs are underutilized.[32] Thatcher believes enterprise GPU utilization is only 15% to 20%, a problem that HP is addressing through new, efficient methods: "We've enabled every HP workstation to share its AI resources across our enterprise. Imagine the ability to search for idle GPUs and use them to run your workloads. We're seeing up to a sevenfold improvement in on-demand computing acceleration, and this could soon be industry standard."[33]

In addition, the market for AI resources on the cloud is ever-changing. For instance, concerns around AI sovereignty are increasing globally.[34] While companies around the world approved running their e-commerce platforms or websites on American cloud servers, the applicability of AI to national intelligence and data management makes some hesitant to place AI workloads overseas. This opens up a market for new national AI cloud providers or private cloud players.[35] GPU-as-a-service computing startups are an alternative to hyperscalers.[36] This means that the market for renting compute power may soon be more fragmented, which could give enterprise customers more options.

Finally, AI may be top of mind for the next two years, but today's build versus buy decisions could have impacts beyond AI considerations. Enterprises may soon consider using quantum computing for the next generation of cryptography (especially as AI ingests and transmits more sensitive data), optimization, and simulation, as we discuss in **"The new math: Solving cryptography in an age of quantum."**

Data center sustainability

Much has been said about the energy use of data centers running large AI models. Major bank reports have questioned whether we have the infrastructure to meet AI demand.[37] The daily power usage of major chatbots has been equated to the daily consumption of nearly 180,000 US households.[38] In short, AI requires unprecedented resources from data centers, and aging power grids are likely not up to the task. While many companies may be worried about getting their hands on AI chips like GPUs to run workloads, sustainability may well be a bigger issue.

Currently, multiple advancements that aim to make AI more sustainable are underway. Enterprises should take note of advancements in these areas over the next two years when considering data centers for AI (figure 2):

- **Renewable sources:** Pressure is mounting on the providers of data centers and AI-over-the-cloud to find sustainable energy sources—and the **rapidly growing focus on AI** may help transition the overall economy to renewables.[39] Major tech companies are already exploring partnerships with nuclear energy providers.[40] Online translation service DeepL hosts a data center in Iceland that's cooled by the naturally frigid air and is fully powered by geothermal and hydroelectric power.[41] And in El Salvador, companies are even exploring how they could power data centers with volcanos.[42]

- **Sustainability applications:** While *building* AI consumes a lot of energy, *applying* AI can, in many cases, offset some of these carbon costs. AI is already

being used to map and track deforestation, melting icebergs, and severe weather patterns. It can also help companies track their emissions and be more efficient in using data centers.[43]

- **Hardware improvements:** New GPUs and NPUs have already saved energy and cost for enterprises. Innovation is not stalling. Intel and Global Foundries recently unveiled new chips that can use light, rather than electricity, to transmit data.[44] This could revolutionize data centers, enabling reduced latency, more distributed construction, and improved reliability. While this fiber optic approach is expensive now, costs may come down over the next couple of years, enabling this type of chip to become mainstream.

Finally, an infrastructure resurgence wouldn't be complete without a nod to connectivity. As edge devices proliferate and companies rely on renting GPU usage from data centers, the complexities of interconnectivity could multiply. High-performance interconnect technologies like NVIDIA's NVLink are already primed for communications between advanced GPUs and other chips.[45] Advancements in 6G can integrate global terrestrial and non-terrestrial networks (like satellites) for ubiquitous connectivity, such that a company in Cape Town relying on a data center in Reykjavik has minimal lag.[46]

As *The Wall Street Journal* has noted, the AI transformation for enterprises is akin to the transition to electric that many car manufacturers are experiencing.[47] Technology infrastructure needs to be rethought on a component-by-component basis, and the decisions made today around edge footprint, investment in specialized hardware, and sustainability can have lasting impacts.

Next: We were promised robots

If today's hardware requires a strategic refresh, enterprises may have much more on their plates in the next decade when robotics become mainstream and smart devices become worthy of their label. Consider the example of the latest smart factories, which use a cascade of computer vision, ubiquitous sensors, and data to build machines that can learn and improve as they manufacture products.[48] Instead of simply providing readings or adjusting on one parameter, like a thermostat, mesh networks of multiple AI-embedded devices can create collaborative compute environments and orchestrate diverse resources.[49]

Another form of smart factory is being developed by Mytra, a San Francisco–based company that simplifies the manual process of moving and storing warehouse materials. The company has developed a fully modular storage system composed of steel "cubes," which can

Figure 2

Advancements in areas related to AI requirements

	Renewable sources	Energy-saving applications	Hardware improvements
Consider	Tracking the energy costs of AI on cloud	Applying AI to discover potential energy savings	Monitoring technological advancements in AI
Implement	Seek out innovative sustainability solutions	Optimize emissions tracking and data usage	Invest in new energy-efficient chips

Source: Deloitte research.

be assembled together in any shape that supports 3D movement and storage of material within, manipulated by robots and optimized through software.[50] Chris Walti, chief executive officer of Mytra, believes this modular approach unlocks automation for any number of unpredictable future applications: "It's one of the first general-purpose computers for moving matter around in 3D space."[51]

Walti believes there is immense potential to apply robotics to relatively constrained problems, such as moving material in a grid or driving a vehicle in straight lines.[52] Until now, in many cases, a good robot has been hard to find. Sustainability, security, and geopolitics are all salient concerns for such a technology. And that's after we even muster the infrastructure noted earlier, including data, network architecture, and chip availability, to make such a leap forward possible. As the saying goes, "hardware is hard."[53] Over the next decade, advancements in robotics applied to more and more complex situations could revolutionize the nature of manufacturing and other physical labor. The potential leads directly to humanoid robotics—bots that are dynamic, constantly learning, and capable of doing what we do.

Economists and businesses alike have argued that aging populations and labor shortages necessitate greater investment in robotics and automation.[54] In many cases, this entails large industrial robots completing relatively simple tasks, as noted above, but more complex tasks require "smarter" mechanical muscle that can move around as humans do. Take the example of Figure AI's humanoid robots tested at the BMW plant in Spartanburg, South Carolina.[55] The autonomous robot, through a combination of computer vision, neural networks, and trial and error, successfully assembled parts of a car chassis.[56]

As the furthest star of progress in this realm, we might anticipate humanoid robots performing a broad variety of tasks, from cleaning sewers to ferrying materials between hospital rooms or even performing surgeries.[57] Just as AI is currently transforming knowledge work, the increased presence of robots could greatly affect physical work and processes in **manufacturing** and beyond. In both cases, companies should be sure to find ways for humans and machines to work together more efficiently than either could do alone. Labor shortages addressed by robotics should then free up human time for more of the uniquely creative and complex tasks where we thrive. As the author Joanna Maciejewska has said astutely, "I want AI to do my laundry and dishes so that I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes."[58]

Endnotes

1. Jon Quast, "Artificial intelligence (AI) juggernaut Nvidia is one of the world's most valuable companies. Here's what investors should know," *The Motley Fool*, June 22, 2024.
2. Duncan Stewart et al., "Gen AI chip demand fans a semi tailwind … for now," *Deloitte Insights*, November 29, 2023; World Semiconductor Trade Statistics (WSTS), "Semiconductor market forecast spring 2023," June 6, 2023.
3. Rob Enderle, "AMD enters AI PC race, closes Microsoft Copilot+ launch gaps," *TechNewsWorld*, July 15, 2024; Saba Prasla, "Meet the future of computing with AI PCs," *Dell Blog*, May 31, 2024; HP, "HP unveils industry's largest portfolio of AI PCs," press release, March 7, 2024.
4. Taiba Jafari et al., "Projecting the electricity demand growth of generative AI large language models in the US," Center on Global Energy Policy, July 17, 2024.
5. International Energy Agency, *Electricity 2024: Analysis and forecast to 2026*, revised May 2024.
6. Deloitte, "Powering artificial intelligence," accessed November 18, 2024.
7. Constellation, "Constellation to launch Crane Clean Energy Center, restoring jobs and carbon-free power to the grid," press release, September 20, 2024.
8. Shira Ovide, "This $400 toothbrush is peak AI mania," *The Washington Post*, April 5, 2024; David Niewolny, "Boom in AI-enabled medical devices transforms healthcare," *NVIDIA Blog*, March 26, 2024.
9. Marc Andreessen, "Why software is eating the world," Andreessen Horowitz, August 20, 2011.
10. John Thornhill, "How hardware is (still) eating the world," *The Financial Times*, February 15, 2024.
11. Stewart et al., "Gen AI chip demand fans a semi tailwind … for now."
12. Ibid.
13. NVIDIA, "NVIDIA hopper GPUs expand reach as demand for AI grows," press release, March 21, 2023.
14. Databricks, *State of data + AI*, accessed October 2024.
15. John Thornhill, "The likely winners of the generative AI gold rush," *The Financial Times*, May 11, 2023.
16. Matt Ashare, "Nvidia sustains triple-digit revenue growth amid AI building boom," *CIO Dive*, August 29, 2024; NVIDIA, "Nvidia (NVDA) Q2 2025 earnings call transcript," *The Motley Fool*, August 28, 2024; Dean Takahashi, "Nvidia unveils next-gen Blackwell GPUs with 25X lower costs and energy consumption," *VentureBeat*, March 18, 2024.
17. Matt Ashare, "Big tech banks on AI boom as infrastructure spending heads for trillion-dollar mark," *CIO Dive*, August 5, 2024; Dell'Oro Group, "Worldwide data center capex to grow at a 24 percent CAGR by 2028," press release, August 1, 2024.
18. Evan Halper, "Amid explosive demand, America is running out of power," *The Washington Post*, March 7, 2024.
19. Chris Hoffman, "What the heck is an NPU, anyway? Here's an explainer on AI chips," *PCWorld*, September 18, 2024.
20. Anshel Sag, "At the heart of the AI PC battle lies the NPU," *Forbes*, April 29, 2024.
21. Phone interview with Vivek Mohindra, senior vice president of corporate strategy, Dell Technologies, October 11, 2024.
22. Christie Simons et al., *2024 global semiconductor industry outlook*, Deloitte, 2024.
23. Aditya Agrawal, "The convergence of edge computing and 5G," Control Engineering, August 7, 2023; Baris Sarer et al., "AI and the evolving consumer device ecosystem," Deloitte's *CIO Journal* for *The Wall Street Journal*, April 24, 2024.
24. Matthew S. Smith, "When AI unplugs, all bets are off," *IEEE Spectrum*, December 1, 2023.
25. Phone interview with Vivek Mohindra, senior vice president of corporate strategy, Dell Technologies, October 11, 2024.
26. Patrick Seitz, "AI PCs are here. Let the upgrades begin, computer makers say," *Investor's Business Daily*, July 5, 2024; Sam Reynolds, "AI-enabled PCs will drive PC sales growth in 2024, say research firms," *Computerworld*, January 11, 2024.
27. Phil Solis et al., "The future of next-gen AI smartphones," IDC, February 19, 2024.
28. Phone interview with Alex Thatcher, senior director of AI PC experiences and cloud clients at HP, October 4, 2024.
29. Rob Waugh, "Assessing Apple Intelligence: Is new 'on-device' AI smart enough for the enterprise?," *The Stack*, September 12, 2024; Matt O'Brien, "Microsoft's new AI-enabled laptops will have a 'photographic memory' of your virtual activity," *Fortune*, May 20, 2024. Tech Trends is an independent publication and has not been authorized, sponsored, or otherwise approved by Apple Inc.
30. Luke Larsen, "AMD just won the AI arms race," *Digital Trends*, June 3, 2024.
31. David Linthicum, "Learning cloud cost management the hard way," *InfoWorld*, July 16, 2024.
32. Tobias Mann, "Big Cloud deploys thousands of GPUs for AI – yet most appear under-utilized," *The Register*, January 15, 2024.
33. Phone interview with Alex Thatcher, senior director of AI PC experiences and cloud clients at HP, October 4, 2024.
34. Christine Mui, "Welcome to the global 'AI sovereignty' race," *Politico*, September 18, 2024.
35. Ibid.
36. Bobby Clay, "Graphics processing service providers step up to meet demand for cloud resources," *S&P Global Market Intelligence*, July 19, 2024.
37. Goldman Sachs, *Top of Mind* 129, June 25, 2024.
38. Cindy Gordon, "ChatGPT and generative AI innovations are creating sustainability havoc," *Forbes*, March 12, 2024.
39. Molly Flanagan, "AI and environmental challenges," Environmental Innovations Initiative, accessed October 2024; Deloitte, "Powering artificial intelligence."
40. Jennifer Hiller and Sebastian Herrera, "Tech industry wants to lock up nuclear power for AI," *The Wall Street Journal*, July 1, 2024.
41. Robert Scheier, "4 paths to sustainable AI," *CIO*, January 31, 2024.
42. Tom Dotan and Asa Fitch, "Why the AI industry's thirst for new data centers can't be satisfied," *The Wall Street Journal*, April 24, 2024.

43. Victoria Masterson, "9 ways AI is helping tackle climate change," World Economic Forum, February 12, 2024.
44. Kirk Ogunrinde, "Intel is using lasers to help meet AI demands on data centers," *Forbes*, June 26, 2024.
45. Rick Merritt, "What is NVLink?," NVIDIA, March 6, 2023.
46. Garry Kranz, "What is 6G? Overview of 6G networks & technology," *TechTarget*, last updated November 2023.
47. Steven Rosenbush, "AI will force a transformation of tech infrastructure," *The Wall Street Journal*, September 11, 2024.
48. Majeed Ahmad, "Sensor fusion with AI transforms the smart manufacturing era," *EE Times*, July 26, 2023.
49. Melissa Malec, "AI orchestration explained: The what, why & how for 2024," HatchWorks AI, last updated June 6, 2024.
50. Phone interview with Chris Walti, chief executive officer of Mytra, October 11, 2024.
51. Ibid.
52. Ibid.
53. Sara Holoubek and Jessica Hibbard, "Why hardware is hard," Luminary Labs, accessed October 2024.
54. Peter Dizikes, "Study: As a population gets older, automation accelerates," *MIT News*, September 15, 2021; Hans Peter Bronomo, "Inside Google's 7-year mission to give AI a robot body," *WIRED*, September 10, 2024.
55. BMW Group, "Successful test of humanoid robots at BMW Group Plant Spartanburg," press release, August 6, 2024.
56. Ibid.
57. Viktor Doychinov, "An army of sewer robots could keep our pipes clean, but they'll need to learn to communicate," *The Conversation*, January 26, 2021; Case Western Reserve University, "5 medical robots making a difference in healthcare," *Online Engineering Blog*, accessed October 2024; National Institute of Biomedical Imaging and Bioengineering (NIBIB), "Robot performs soft tissue surgery with minimal human help," press release, April 20, 2022.
58. Joanna Maciejewska's post on X, March 29, 2024.

Continue the conversation

Industry leadership

Nitin Mittal
Global AI leader | Principal | Deloitte Consulting LLP

Nitin Mittal is a principal with Deloitte Consulting LLP. He currently serves as the US Artificial Intelligence (AI) Strategic Growth Offering Consulting leader and the Global Strategy, Analytics and M&A leader. He is the 2019 recipient of the AI Innovator of the Year award at the AI Summit New York. He specializes in advising clients to achieve competitive advantage through data and cognitive powered transformations that promote amplified intelligence and enable our clients to make strategic choices and transform ahead of disruption.

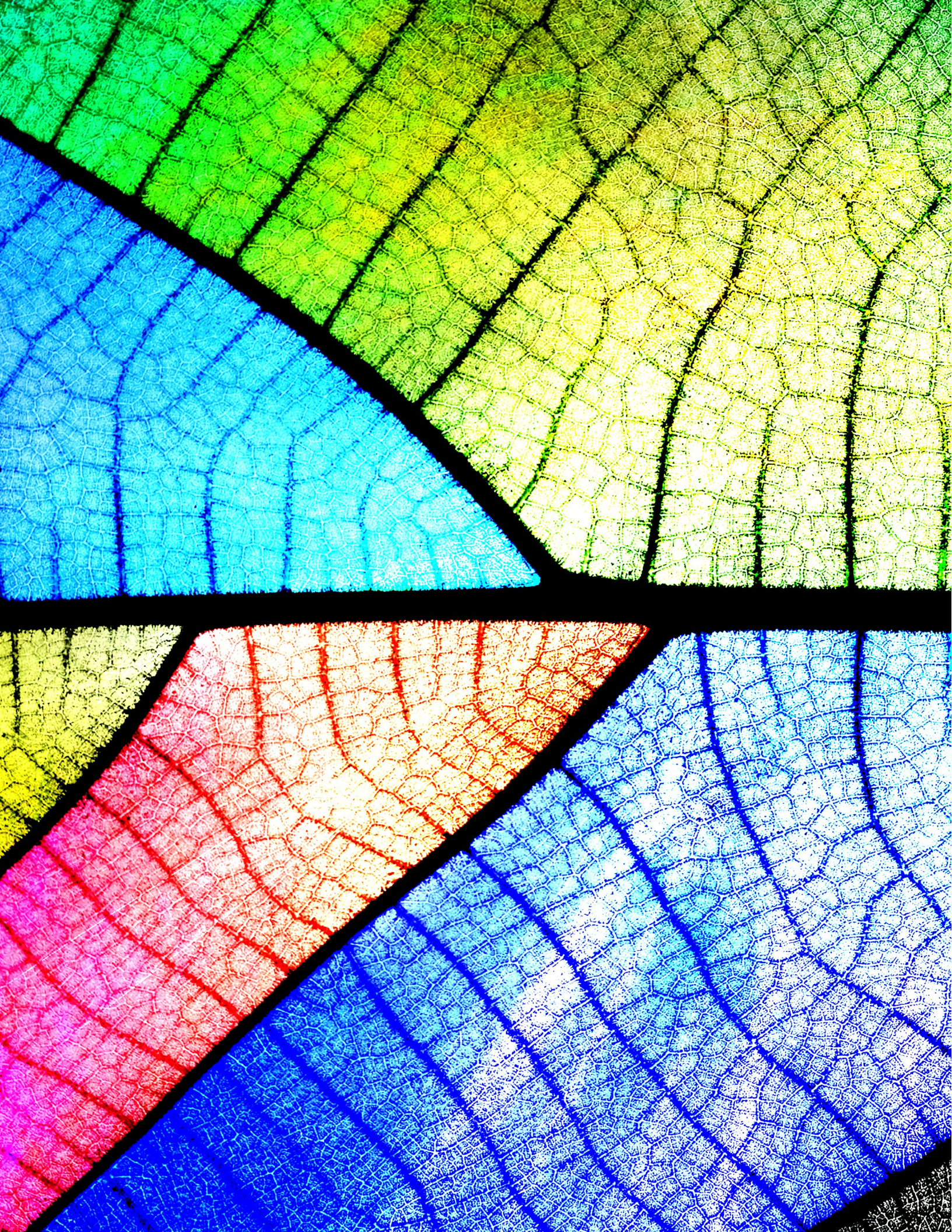
Throughout his career, Mittal has served as a trusted advisor to global clients and has worked across a number of industry sectors. His primary focus has been working with life sciences and health care clients, implementing large scale data programs that promote organizational intelligence, and the use of advanced analytics and AI to drive insights and business strategy.

Abdi Goodarzi
US Enterprise Performance Portfolio leader | Principal
+1 714 913 1091 | agoodarzi@deloitte.com

Abdi Goodarzi is a principal with Deloitte Consulting LLP, leading Deloitte's Enterprise Performance (EP) Offerings Portfolio. This portfolio of six offerings provide strategy, implement and operate services for variety of enterprise functions, from end-to-end business and IT transformation, to digital supply chain optimization, manufacturing and product strategies, and procurement as-a-service, to global finance, shared services, planning, ITSM, and full scale AMS and BPO. This portfolio offers competency in many ERP platforms such as SAP, Oracle, Workday Financials and Infor, in addition to ServiceNow, Anaplan, Ariba, and Coupa, as well as real estate solutions such as Nuvolo, as well as PLM, planning and fulfillment, and engineering solutions such as Siemens, PTC, O9, OMP and IBP.

Acknowledgments

Much gratitude goes to the many subject matter leaders across Deloitte that contributed to our research for the Computation chapter: **Lou DiLorenzo**, **Abdi Goodarzi**, **Lena La**, **Nitin Mittal**, **Manish Rajendran**, **Jim Rowan**, and **Baris Sarer**.



BUSINESS OF TECHNOLOGY

IT, amplified: AI elevates the reach (and remit) of the tech function

As the tech function shifts from leading digital transformation to leading AI transformation, forward-thinking leaders are using this as an opportunity to redefine the future of IT

Kelly Raskovich, Bill Briggs, Mike Bechtel, and Abhijith Ravinutala

Much has been said, including within the pages of Tech Trends, about the potential for artificial intelligence to revolutionize business use cases and outcomes. Nowhere is this more true than in the end-to-end life cycle of software engineering and the broader business of information technology, given generative AI's ability to write code, test software, and augment tech talent in general. Deloitte research has shown that tech companies at the forefront of this organizational change are ready to realize the benefits: They are twice as likely as their more conservative peers to say generative AI is transforming their organization now or will within the next year.[1]

We wrote in a Tech Trends 2024 article that enterprises need to reorganize their developer experiences to help IT teams achieve the best results. Now, the AI hype cycle has placed an even greater focus on the tech function's ways of working. IT has long been the lighthouse of digital transformation in the enterprise, but it must now take on *AI transformation*. Forward-thinking IT leaders are using the current moment as a once-in-a-generation opportunity to redefine roles and responsibilities, set investment priorities, and communicate value expectations. More importantly, by playing this pioneering role, chief information officers can help inspire other technology leaders to put AI transformation into practice.

After years of enterprises pursuing lean IT and everything-as-a-service offerings, AI is sparking a shift away from virtualization and austere budgets. Gartner predicts that "worldwide IT spending is expected to total $5.26 trillion in 2024, an increase of 7.5% from 2023."[2] As we discuss in "Hardware is eating the world," hardware and infrastructure are having a moment, and enterprise IT spending and operations may shift accordingly.

As both traditional AI and generative AI become more capable and ubiquitous, each of the phases of tech delivery may see a shift from human in charge to human in the loop. Organizations need a clear strategy in place before that occurs. Based on Deloitte analysis, over the next 18 to 24 months, IT leaders should plan for AI transformation across five key pillars: engineering, talent, cloud financial operations (FinOps), infrastructure, and cyber risk.

This trend may usher in a new type of lean IT over the next decade. If commercial functions see an increased number of citizen developers or digital agents that can spin up applications on a whim, the role of the IT function may shift from building and maintaining to orchestrating and innovating. In that case, AI may not only be undercover, as we indicate in the introduction to this year's report, but may also be overtly in the boardroom, overseeing tech operations in line with human needs.

Now: Spotlight—and higher spending—on IT

For years, IT has been under pressure to streamline sprawling cloud spend and curb costs. Since 2020, however, investments in tech have been on the rise thanks to pent-up demand for collaboration tools and the pandemic-era emphasis on digitalization.[3] According

to Deloitte research, from 2020 to 2022, the global average technology budget as a percentage of revenue jumped from 4.25% to 5.49%, an increase that approximately doubled the previous revenue change from 2018 to 2020.[4] And in 2024, US companies' average budget for digital transformation as a percentage of revenue is 7.5%, with 5.4% coming from the IT budget.[5]

As demand for AI sparks another increase in spending, the finding from Deloitte's 2023 Global Technology Leadership Study continues to ring true: Technology *is* the business, and tech spend is increasing as a result.

Today, enterprises are grappling with the new relevance of hardware, data management, and digitization in ramping up their usage of AI and realizing its value potential. In Deloitte's Q2 State of Generative AI in the Enterprise report, businesses that rated themselves as having "very high" levels of expertise in generative AI were increasing their investment in hardware and cloud consumption much more than the average enterprise.[6] Overall, 75% of organizations surveyed have increased their investments around data-life-cycle management due to generative AI.[7]

These figures point to a common theme: To realize the highest impact from gen AI, enterprises likely need to accelerate their cloud and data modernization efforts. AI has the potential to deliver efficiencies in cost, innovation, and a host of other areas, but the first step to accruing these benefits is for businesses to focus on making the right tech investments.[8] Because of these crucial investment strategies, the spotlight is on tech leaders who are paving the way.

According to Deloitte research, over 60% of US-based technology leaders now report directly to their chief executives, an increase of more than 10 percentage points since 2020.[9] This is a testament to the tech leader's increased importance in setting the AI strategy rather than simply enabling it. Far from a cost center, IT is increasingly being seen as a differentiator in the AI age, as CEOs, following market trends, are keen on staying abreast of AI's adoption in their enterprise.[10]

John Marcante, former global CIO of Vanguard and US CIO-in-residence at Deloitte, believes AI will fundamentally change the role of IT. He says, "The technology organization will be leaner, but have a wider purview. It will be more integrated with the business than ever. AI is moving fast, and centralization is a good way to ensure organizational speed and focus."[11]

As IT gears up for the opportunity presented by AI—perhaps the opportunity that many tech leaders and employees have waited for—changes are already underway in how the technology function organizes itself and executes work. The stakes are high, and IT is due for a makeover.

New: An AI boost for IT

Over the next 18 to 24 months, the nature of the IT function is likely to change as enterprises increasingly employ generative AI. Deloitte's foresight analysis suggests that, by 2027, even in the most conservative scenario, gen AI will be embedded into every company's digital product or software footprint (figure 1), as we discuss across five key pillars.[12]

Engineering

In the traditional software development life cycle, manual testing, inexperienced developers, and disparate tool environments can lead to inefficiencies, as we've discussed in prior Tech Trends. Fortunately, AI is already having an impact on these areas. AI-assisted code generation, automated testing, and rapid data analytics all save developers more time for innovation and feature development. The productivity gain from coding alone is estimated to be worth US$12 billion in the United States alone.[13]

At Google, AI tools are being rolled out internally to developers. In a recent earnings call, CEO Sundar Pichai said that around 25 percent of the new code at the technology giant is developed using AI. Shivani Govil, senior director of product management for developer products, believes that "AI can transform how engineering teams work, leading to more capacity to innovate, less toil, and higher developer satisfaction. Google's approach is to bring AI to our users and meet them where they are—by bringing the technology into products and tools that developers use every day to support them in their work. Over time, we can create even tighter alignment between the code and business requirements, allowing faster feedback loops, improved product market fit, and

Figure 1

How generative AI might transform IT ways of working

Over the next 18 to 24 months, enterprises may experience vast improvement in their technology teams as generative AI is increasingly embedded into ways of working. Deloitte's foresight analysis suggests that by 2027, even in the most conservative scenario, gen AI will be embedded into every company's digital product/software footprint. Manual and time-consuming processes like code reviews, infrastructure configuration, and budget management can be automated and improved, as we move from current to target state of AI in IT.

	The problem	Necessary changes	Recommended actions
Engineering	Manual, inefficient aspects of the traditional software development life cycle	Shift from writing code to defining the architecture, reviewing code, and orchestrating functionality	Tech leaders should expect human-in-the-loop code generation and review to become the standard
Talent	Executives struggle to hire workers with the right backgrounds and are forced to delay projects	AI can generate rich learning and development media as well as documentation to upskill talent	Tech leaders should implement regular AI-powered learning recommendations and personalization as a new way of working
Cloud financial operations	Runaway spend is common in the cloud, since resources can be provisioned with a click	AI-powered cost analysis, pattern detection, and resource allocation can optimize IT spend at new speeds	Leaders should consistently apply AI to help it earn its keep and optimize costs
Infrastructure	Nearly half of enterprises are handling tasks like security, compliance, and service management on a manual basis	Automated resource allocation, predictive maintenance, and anomaly detection could revolutionize IT systems	Leaders should work toward an IT infrastructure that can heal itself as needed through AI
Cyber	Generative AI and digital agents open up more attack surfaces than ever for bad actors	Automated data masking, incident response, and policy generation can optimize cybersecurity responses	Enterprises should take steps to further authenticate data and digital media through new tech or processes

Source: Deloitte research and analysis.

better alignment to the business outcomes."[14] In another example, a health care company used COBOL code assist to enable a junior developer with no experience in the programming language to generate an explanation file with 95% accuracy.[15]

As Deloitte recently stated in a piece on engineering in the age of gen AI, the developer role is likely to shift from writing code to defining the architecture, reviewing code, and orchestrating functionality through contextualized prompt engineering. Tech leaders should anticipate human-in-the-loop code generation and review to be the standard over the next few years of AI adoption.[16]

Talent

Technology executives surveyed by Deloitte last year noted that they struggle to hire workers with critical IT backgrounds in security, machine learning, and software architecture, and are forced to delay projects with financial backing due to a shortage of appropriately skilled talent.[17] As AI becomes the newest skill in demand, many companies may not even be able to find all the talent they need, leading to a hiring gap wherein nearly 50% of AI-related positions cannot be filled.[18]

As a result, tech leaders should focus on upskilling their own talent, another area where **AI can help**. Consider the potential benefits of AI-powered skills gap analyses and recommendations, personalized learning paths, and virtual tutors for on-demand learning. Bayer, the life sciences company, has used generative AI to summarize procedural documents and generate rich media such as animation for e-learning.[19] Along the same lines, AI could generate documentation to help a new developer understand a legacy technology, and then create an associated learning podcast and exam for that same developer.

At Google, developers thrive on hands-on experience and problem-solving, so leaders are keen to provide AI learning and tools (like coding assistants) that meet developers where they are on their learning journey. "We can use AI to enhance learning, in context with emerging technologies, in ways that anticipate and support the rapidly changing skills and knowledge required to adapt to them," says Sara Ortloff, senior director of developer experience at Google.[20]

As automation increases, tech talent would take an oversight role and enjoy more capacity to focus on innovation that can improve the bottom line (**as we wrote about last year**). This could help attract talent since, according to Deloitte research, the biggest incentive that attracts tech talent to new opportunities is the work they would do in the role.[21]

Cloud financial operations

Runaway spending became a common problem in the cloud era when resources could be provisioned with a click. Hyperscalers have offered data and tooling for finance teams and CIOs to keep better track of their team's cloud usage, but many of these FinOps tools still require manual budgeting and offer limited visibility across disparate systems.[22] The power of AI enables organizations to be more informed, proactive, and effective with their financial management. Real-time cost analysis, as well as robust pattern detection and resource allocation across systems, can optimize IT spending at a new speed.[23] AI can help enterprises identify more cost-saving opportunities through better predictions and tracking.[24] All of this is necessary because AI may significantly drive up cloud costs for large companies in the coming years. Applying AI to FinOps can help justify the investments in AI and optimize costs elsewhere while AI demand increases.[25]

Infrastructure

Across the very broad scope of IT infrastructure, from toolchain to service management, organizations haven't seen as much automation as they want.Just a few years ago, studies estimated that nearly half of large enterprises were handling key tasks like security, compliance, and service management on a completely manual basis.[26] The missing ingredient? Automation that can learn, improve, and react to the changing demands of a business. Now, that's possible.

Automated resource allocation, predictive maintenance, and anomaly detection could all be possible in a system that's set up to natively understand its own real-time status and then act.[27] This emerging view of IT is known as *autonomic*, in reference to the human body's autonomic nervous system that regulates its heart rate and breath, and adjusts dynamically to internal and external stimuli.[28] As mentioned above, such a system would enable the change from human in charge to human in the loop, as infrastructure takes care of itself and surfaces only the issues that require human intervention. That's why companies like eBay are already leveraging generative AI to scale their infrastructure and sort through troves of customer data, potentially leading to impactful changes to their platform.[29]

Cyber

Although AI may make many aspects of IT simpler or more efficient, it certainly introduces more complexity to cyber risk. **As we wrote about last year**, generative AI and synthetic media open up more attack surfaces than ever for phishing, deepfakes, prompt injection, and

others.[30] As AI proliferates and digital agents become the newest business-to-business representatives, these issues may become more severe. Enterprises should take steps to work on data authentication, as in the example of SWEAR, a security company that has pioneered a way to verify digital media through the blockchain.[31] Data masking, incident response, and automated policy generation are all also areas where generative AI can be applied to optimize cybersecurity responses and defend against attacks.[32]

Finally, as technology teams grow accustomed to the changes and challenges mentioned above, many will shift their focus to the innovation, agility, and growth that can be enabled by AI. Teams can streamline their IT workflows and reduce the need for manual intervention or offshoring, allowing IT to focus on higher-value activities.[33] Indeed, an entire reallocation of IT resources is likely to take place. As Ian Cairns, CEO of Freeplay, has noted, "As with any major platform shift, the businesses that succeed will be the ones that can rethink and adapt how they work and build software for a new era."[34]

Next: IT itself as a service

The current moment is like an all-hands-on-deck siren sounding for many IT teams, where product managers, domain experts, and business unit leaders are diving into the details of AI to stand up working proofs of concepts. If the bet pays off and companies are able to improve their margins with this new technology, IT may complete its transition from a cost center and enabler to a true competitive differentiator. By then, the role of the CIO and their management of the tech estate could be dramatically altered.

Imagine a scenario over the next decade where IT transitions from a centrally controlled function to an innovation leader, providing reusable code blocks and platforms that business units can use to develop their own solutions. While IT-as-a-service may not be new, the previous understanding was that several aspects of a company's IT infrastructure would be handed off to a new vendor.[35] Looking forward, that vendor could be replaced by each organization's internally trained and secure AI agents.

In this sense, IT itself could become a service run through online portals, where a combination of low-code or no-code technologies and advanced AI allows nontechnical users to create and run applications.[36] For example, the role of the chief architect could look very different with many legacy tasks performed by a digital agent. Just as cloud computing blocks can today be opened with a click, entire applications may be available at a click in the next five to 10 years. Continuous tech learning and fluency would become essential across the enterprise, not just in IT, as employees and citizen developers would be encouraged to adapt to the latest technologies. Trust and security responsibilities would also broaden, with technology teams retaining humans in the loop to review data privacy, cybersecurity, and ethical AI practices.

Though the advancement of AI may call into question the future role of IT, it actually elevates the technology function in the enterprise once it's embedded everywhere. Savvy tech leaders will need to develop a bevy of skills as tech and AI become even more important in the enterprise. These skills include journey and process knowledge, program and product management, business development, trust and compliance expertise, and ecosystem management (including AI tools and shareability). Leaders may also need to take on a new role as the enterprise's educator and evangelist of AI, in order to drive change management.

Marcante says, "AI capabilities may be democratized for the business and spur innovation, but tech leaders have to drive the agenda. There has to be a set of guiding principles and goals that people can point to globally to move their enterprise forward."[37]

Endnotes

1. Faruk Muratovic, Duncan Stewart, and Prashant Raman, "Tech companies lead the way on generative AI: Does code deserve the credit?" *Deloitte Insights*, August 2, 2024.
2. Gartner, "Gartner forecasts worldwide IT spending to grow 7.5% in 2024," press release, July 16, 2024.
3. Lou DiLorenzo Jr. et al., "From tech investment to impact: Strategies for allocating capital and articulating value," *Deloitte Insights*, September 13, 2023.
4. Ibid.
5. Tim Smith et al., "Focusing on the foundation: How digital transformation investments have changed in 2024," *Deloitte Insights* October 14, 2024.
6. Nitin Mittal et al., "Now decides next: Getting real about Generative AI," Deloitte's State of Generative AI in the Enterprise Q2 report, April 2024.
7. Ibid.
8. Elizabeth Sullivan (ed.), "Gen AI investments increasingly extend beyond the AI itself," *Deloitte Insights Magazine* 33, September 26, 2024.
9. Belle Lin, "AI puts CIOs in the spotlight, right next to the CEO," *The Wall Street Journal*, June 12, 2024.
10. Benjamin Finzi et al., "Three roles CEOs need to play to scale generative AI," Deloitte, 2024.
11. John Marcante, former global CIO of Vanguard and US CIO-in-residence at Deloitte, Deloitte interview, October 8, 2024.
12. Laura Shact et al., "Four futures of generative AI in the enterprise: Scenario planning for strategic resilience and adaptability," *Deloitte Insights*, October 25, 2024.
13. Muratovic et al., "Tech companies lead the way on generative AI."
14. Shivani Govil, senior director and project manager of developer tools, Google, Deloitte interview, September 4, 2024.
15. Faruk Muratovic et al., "How can organizations engineer quality software in the age of gen AI?," *Deloitte Insights*, October 28, 2024.
16. Ibid.
17. David Jarvis, "Tech talent is still hard to find, despite layoffs in the sector," *Deloitte Insights*, August 14, 2023.
18. Mark Dangelo, "Needed AI skills facing unknown regulations and advancements," *Thomson Reuters*, December 6, 2023.
19. Donald H. Taylor, *The global sentiment survey 2024*, February 2024.
20. Sara Ortloff, senior director of developer user experience, Google, Deloitte interview, September 4, 2024.
21. Linda Quaranto et al., "Winning the war for tech talent in FSI organizations," Deloitte, February 2022.
22. David Linthicum, "What's going on with cloud finops?," InfoWorld's *Cloud Computing Blog*, February 27, 2024.
23. PwC, "FinOps and AI: Balancing innovation and cost efficiency," *CIO*, September 24, 2024.
24. Fred Delombaerde, "Will AI and LLMs transform FinOps?," video, FinOps Foundation, May 20, 2024.
25. Linthicum, "What's going on with cloud finops?"
26. Nicholas Dimotakis, "IT's dirty little secret: Manual processes are still prevalent," *Forbes*, February 25, 2021.
27. Michael Nappi, "Go beyond with autonomic IT to drive the autonomous business," *ScienceLogic Blog*, May 15, 2024.
28. Science Direct, Autonomic Computing, accessed October 2024.
29. John Kell, "How eBay uses generative AI to make employees and online sellers more productive," *Fortune*, August 14, 2024.
30. Mike Bechtel and Bill Briggs, "Defending reality: Truth in an age of synthetic media," *Deloitte Insights*, December 4, 2023.
31. Jason Crawforth, "My take: Tackling the problem of deepfakes," *Deloitte Insights*, August 7, 2024.
32. Palo Alto Networks, "What is generative AI in cybersecurity?," *Cyberpedia*, accessed October 29, 2024.
33. Ilya Gandzeichuk, "How AI can transform the IT service industry in the next 5 years," *Forbes*, May 16, 2024.
34. Ian Cairns, "Generative AI forces rethink of software development process," *Deloitte Insights*, July 1, 2024.
35. Canon, "4 reasons why 'as-a-service' is the future for IT teams," accessed October 2024.
36. CloudBlue, "What is IT as a service?" November 28, 2022; Isaac Sacolick, "7 innovative ways to use low-code tools and platforms," *InfoWorld*, April 22, 2024.
37. Phone interview with John Marcante, former global CIO of Vanguard and US CIO-in-residence at Deloitte, October 8, 2024.

Continue the conversation

Industry leadership

Tim Smith

Tech Strategy & Business Transformation leader | Principal | US Monitor Deloitte | Deloitte Consulting LLP
+1 212 313 2979 | timsmith6@deloitte.com

Tim Smith is a principal with Deloitte Consulting LLP and serves as the US leader for Monitor Deloitte's Technology Strategy & Business Transformation practice. He has more than 20 years of cross-sector technology advisory and implementation experience in the United States and abroad. Tim works with clients to unlock the value within the technology estate via integrated choices across operating models, architectures, and ecosystems. Tim resides in New York City. He earned a BSc in systems engineering from the University of Virginia and an MBA from the London Business School.

Anjali Shaikh

US CIO Program Experience director | Managing director | Deloitte Consulting LLP
+1 714 436 7237 | anjalishaikh@deloitte.com

Anjali Shaikh is the experience director for Deloitte's technology executive programs, serving as an advisor to CIOs, CDAOs, and other tech leaders and providing strategic direction for program development. Shaikh leads a team of skilled practitioners responsible for creating customized experiences and developing valuable insights that help executives navigate complex challenges; shape the tech agenda; build and lead effective teams; and excel in their careers.

Lou DiLorenzo Jr

Principal | AI & Data Strategy Practice leader | US CIO & CDAO Programs, national leader | Deloitte Consulting LLP
+1 612 397 4000 | ldilorenzojr@deloitte.com

Lou DiLorenzo serves as the national leader of Deloitte Consulting's AI & Data Strategy practice and the Deloitte US CIO and CDAO Executive Accelerator programs. He is a member of Deloitte's Generative AI practice leadership team and heads the Generative AI Incubator. With over 20 years of cross-sector operating, entrepreneurial, and consulting experience, he has a successful record of bringing key stakeholders together to help lead change, develop new capabilities, and deliver positive financial results. Previously, DiLorenzo served as COO of a consumer health insurance startup and as Global CIO for the Food Ingredients & Bio Industrial division at Cargill. He is a frequent technology contributor to leading publications and hosts the podcast, *Techfluential*.

Acknowledgments

Much gratitude goes to the many subject matter leaders across Deloitte that contributed to our research for the Business of Technology chapter: **Kenny Brown, Lou DiLorenzo, Diana Kearns-Manolatos, Siva Muthu, Chris Purpura, Anjali Shaikh,** and **Tim Smith.**



The new math: Solving cryptography in an age of quantum

Quantum computers are likely to pose a severe threat to today's encryption practices. Updating encryption has never been more urgent.

Kelly Raskovich, Bill Briggs, Mike Bechtel, and Ed Burns

Cybersecurity professionals already have a lot on their minds. From run-of-the-mill social engineering hacks to emerging threats from AI-generated content, there's no shortage of immediate concerns. But while focusing on the urgent, they could be overlooking an important threat vector: the potential risk that a cryptographically relevant quantum computer (CRQC) will someday be able to break much of the current public-key cryptography that businesses rely upon. Once that cryptography is broken, it will undermine the processes that establish online sessions, verify transactions, and assure user identity.

Let's contrast this risk with the historical response to Y2K, where businesses saw a looming risk and addressed it over time, working backward from a specific time to avert a more significant impact.[1] The potential risk of a CRQC is essentially the inverse case: The effect is expected to be even more sweeping, but the date at which such a cryptographically relevant quantum computer will become available is unknown. Preparing for CRQCs is generally acknowledged to be highly important but is often low on the urgency scale because of the unknown timescale. This has created a tendency for organizations to defer the activities necessary to prepare their cybersecurity posture for the arrival of quantum computers.

"Unless it's here, people are saying, 'Yeah, we'll get to it, or the vendors will do it for me. I have too many things to do and too little budget,'" says Mike Redding, chief technology officer at cybersecurity company Quantropi.[2] "Quantum may be the most important thing ever, but it doesn't feel urgent to most people. They're just kicking the can down the road."

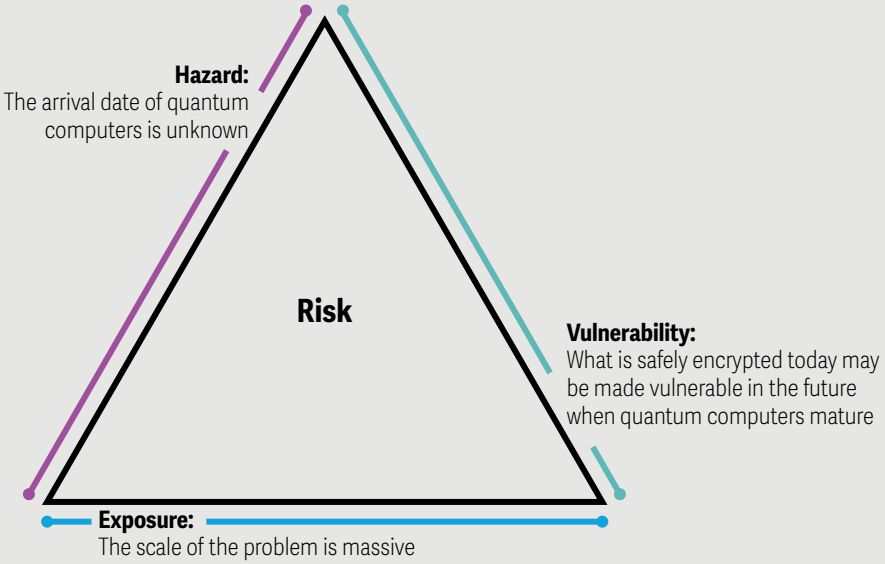
This complacent mindset could breed disaster because the question isn't *if* quantum computers are coming—it's *when*. Most experts consider the exact time horizon for the advent of a CRQC to be irrelevant when it comes to encryption. The consensus is that one will likely emerge in the next five to 10 years, but how long will it take organizations to update their infrastructures and third-party dependencies? Eight years? Ten years? Twelve? Given how long it took to complete prior cryptographic upgrades, such as migrating from cryptographic hashing algorithms SHA1 to SHA2, it is prudent to start now.

In a recent report, the US Office of Management and Budget said, "It is likely that a CRQC will be able to break some forms of cryptography that are now commonly used throughout government and the private sector. A CRQC is not yet known to exist; however, steady advancements in the quantum computing field may yield a CRQC in the coming decade. Accordingly … federal agencies must bolster the defense of their existing information systems by migrating to the use of quantum-resistant public-key cryptographic systems."[3]

The scale of the problem is potentially massive, but fortunately, tools and expertise exist today to help enterprises address it. Recently released postquantum cryptography (PQC) algorithm standards from the US National Institute of Standards and Technology (NIST) could help to neutralize the problem before it becomes costly,[4] and many other governments around the world are also working on this issue.[5] Furthermore, a reinvigorated cyber mindset could set enterprises on the road to better security.

Figure 1

The triangle of risk and the implications of the losses if each of the three is not considered



Source: Colin Soutar, Itan Barmes, and Casper Stap, "Don't let drivers for quantum cyber readiness take a back seat!" Deloitte, accessed November 2024.

Now: Cryptography everywhere

Two of the primary concerns for cybersecurity teams are technology integrity and operational disruption.[6] Undermining digital signatures and cryptographic key exchanges that enable data encryption are at the heart of those fears. Losing the type of cryptography that can guarantee digital signatures are authentic and unaltered would likely deal a major blow to the integrity of communications and transactions. Additionally, losing the ability to transmit information securely could potentially upend most organizational processes.

Enterprises are starting to become aware of the risks posed by quantum computing to their cybersecurity. According to Deloitte's Global Future of Cyber survey, 52% of organizations are currently assessing their exposure and developing quantum-related risk strategies. Another 30% say they are currently taking decisive action to implement solutions to these risks.

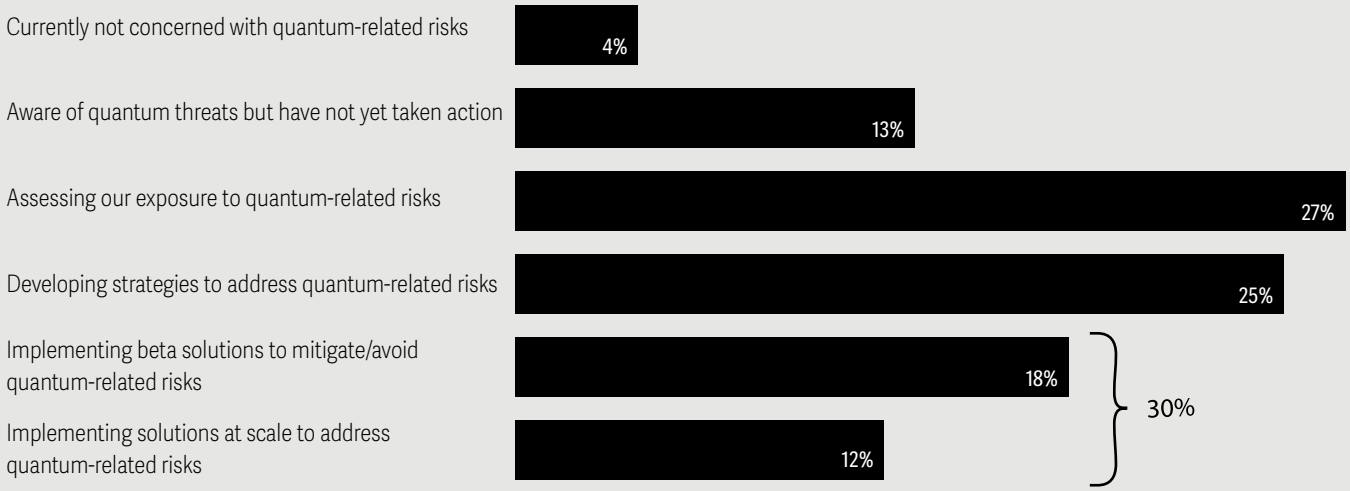
"The scale of this problem is sizeable, and its impact in the future is imminent. There may still be time when it hits us, but proactive measures now will help avoid a crisis later. That is the direction we need to take," says Gomeet Pant, group vice president of security technologies for the India-based division of a large industrial products firm.[7]

Cryptography is now so pervasive that many organizations may need help identifying all the places it appears. It's in applications they own and manage, and in their partner and vendor systems. Understanding the full scope of the organizational risk that a CRQC would pose to cryptography (figure 1) requires action across a wide range of infrastructures, supply chains, and applications. Cryptography used for data confidentiality and digital signatures to maintain the integrity of emails, macros, electronic documents, and user authentication would all be threatened, undermining the integrity and authenticity of digital communications.[8]

Figure 2

The quantum connection

How organizations are thinking about the approaching quantum era and the need for quantum cybersecurity readiness



Note: n = 1,196 C-suite executives and senior leaders.

Source: Deloitte, "The promise of cyber: Enhancing transformational value through cybersecurity resilience," accessed November 2024.

To make matters worse, enterprises' data may already be at risk, even though there is no CRQC yet. There's some indication that bad actors are engaging in what's known as "harvest now, decrypt later" attacks—stealing encrypted data with the notion of unlocking it whenever more mature quantum computers arrive. Organizations' data will likely continue to be under threat until they upgrade to quantum-resistant cryptographic systems.

"We identified the potential threat to customer data and the financial sector early on, which has driven our groundbreaking work toward quantum-readiness," said Yassir Nawaz, director of the emerging technology security organization at JP Morgan. "Our initiative began with a comprehensive cryptography inventory and extends to developing PQC solutions that modernize our security through crypto-agile processes."[9]

Given the scale of the issues, upgrading to quantum-safe cryptography could take years, maybe even a decade or more, and we're likely to see cryptographically relevant quantum computers sometime within that range.[10] The potential threat posed by quantum to cryptography may feel over the horizon, but the time to start addressing it is now (figure 2).

"It is important that organizations start preparing now for the potential threat that quantum computing presents," said Matt Scholl, computer security division chief at NIST. "The journey to transition to the new postquantum-encryption standards will be long and will require global collaboration along the way. NIST will continue to develop new post-quantum cryptography standards and work with industry and government to encourage their adoption."[11]

New: Upgrading to a quantum-safe future

There's good news, though. While upgrading cryptography to protect against the threat of quantum computers requires a comprehensive and widespread effort, given sufficient time, it should be a relatively straightforward operation.

Initial steps include establishing governance and policy, understanding current cryptographic exposure, assessing how best to prioritize remediation efforts across the infrastructure and supply chain, and building a comprehensive road map for internal updates and contractual mechanisms to ensure vendors meet the updated standards.

"The first step to reclaim control over decades of cryptographic sprawl across IT is to leverage modern cryptography management solutions, which empower organizations with critical observability and reporting capabilities," says Marc Manzano, general manager of cybersecurity group SandboxAQ.[12]

Once these initial steps are completed, organizations can begin updating encryption algorithms. In August 2024, NIST released new standards containing encryption algorithms that organizations can implement. The agency says these encryption methods should withstand attacks from quantum computers by changing how data is encrypted and decrypted.[13]

Current encryption practices encode data using complex math problems that outpace the computing power of even today's most powerful supercomputers. But quantum computers will likely be able to crack these problems quickly. The updated NIST standards move away from today's large-number-factoring math problems and leverage lattice and hash problems, which are sufficiently complex to bog down even quantum computers.[14]

Large tech companies are already beginning their transition. Following the release of NIST's updated standards, Apple updated its iMessage application to use quantum-secure encryption methods.[15] Google announced that it implemented the new standards in its cryptography library and will use them in its Chrome web browser.[16] IBM, which has invested heavily in developing quantum computing technology, has integrated postquantum cryptography into several of its platforms, and Microsoft has announced that it will add quantum-secure algorithms to its cryptographic library.[17]

In 2021, the National Cybersecurity Center of Excellence (NCCoE) at NIST started the Migration to PQC project. It has grown to over 40 collaborators, many of whom have cryptographic discovery and inventory tools with differing capabilities. The project demonstrates the use of these tools in a manner that will enable an organization to plan for their use. Other collaborators are focused on testing the PQC algorithms for use in protocols to understand their interoperability and performance as they prepare to implement PQC in their products.[18]

"An organization needs to understand where and how it uses cryptographic products, algorithms, and protocols to begin moving towards quantum-readiness," says Bill Newhouse, co-lead for the Migration to PQC project at the NCCoE. "Our project will demonstrate use of the tools and how the output of the tools supports risk analysis that will enable organizations to prioritize what it will migrate to PQC first."[19]

Next: Leveraging postquantum cryptography to prepare for future threats

While enterprises upgrade their encryption practices, they should consider what else they might do. This can be likened to cleaning out the basement: What can be done to clean out the back corners no one has looked at in a decade? They will map out highly technical, low-level capabilities in core systems that haven't been assessed in years. Perhaps they will uncover other potential issues that can be addressed while upgrading cryptography, such as enhancing governance, improving key management processes, implementing a zero trust strategy, upgrading cryptography while modernizing legacy systems, or simply sunsetting tools that haven't been used in a while.

Organizations that engage in proper cyber hygiene are likely to strengthen their broader cyber and privacy practices. They will likely be more cautious about collecting and sharing anything other than strictly necessary data, establish more robust and accountable governance mechanisms, and continually assess trust between digital

components. Beyond protecting against the far-off threat of quantum attacks, these practices harden an enterprise's defenses today by building secure habits into everyday activities.

Enterprises should consider how to create a reproducible set of activities to protect their cryptographic systems against various types of attacks and failures, a concept known as cryptographic resilience. Today, organizations need to prepare for the quantum threat vector, but tomorrow, the next new risk will require a different approach. Security teams shouldn't have to go through this entire exercise again when a new threat emerges—instead, they should develop the muscles necessary to add or swap out cryptographic capabilities quickly and seamlessly.[20]

As our digital and physical lives become more closely linked, our friendships, reputations, and assets are undergoing a digital transformation. These areas are mediated digitally and secured cryptographically. Going forward, the privacy and integrity of messages, transactions, and an increasing share of the human condition will be built upon a foundation of digital trust. Protecting cryptography isn't only about protecting enterprise data stores—it's about shielding increasingly sensitive areas of our lives.

"As our reliance on cryptography intensifies in the digital economy, organizations must act swiftly to prepare for a controlled transition to maintain the trust they've built with customers and partners," says Michele Mosca, founder and CEO of evolutionQ. "It's crucial for organizations to develop a quantum-safe road map and partner with vendors to kick-start this vital shift. Prioritizing the security of your most sensitive information isn't just prudent—it's essential."[21]

Quantum computers are likely to bring significant benefits to a range of areas, such as drug discovery, financial modeling, and other use cases, that improve people's lives. These potential benefits should not be overshadowed by the attendant security challenges. This is why enterprises should start hardening their defenses now so that they are prepared to reap the potential benefits of quantum computing without major disruption from its risks.

Endnotes

1. Damian Carrington, "Was Y2K bug a boost?," BBC News Online, January 4, 2000.
2. Mike Redding, chief technology officer, Quantropi, Deloitte interview, August 27, 2024.
3. Executive Office of the President of the United States, *Report on post-quantum cryptography*, July 2024.
4. National Institute of Standards and Technology (NIST), "NIST releases first 3 finalized post-quantum encryption standards," press release, August 13, 2024.
5. European Commission, "Commission publishes Recommendation on Post-Quantum Cryptography," press release, April 11, 2024.
6. Emily Mossburg et al., *The promise of cyber: Enhancing transformational value through cybersecurity resilience*, Deloitte, 2024.
7. Gomeet Pant, group vice president of security technologies for the India-based division of a large industrial products firm, Deloitte interview, October 25, 2024.
8. Katherine Noyes, "NIST's postquantum cryptography standards: 'This is the start of the race'," *CIO Journal* for *The Wall Street Journal*, June 12, 2024.
9. Yassir Nawaz, director of emerging technology security, JP Morgan, Deloitte interview, October 14, 2024.
10. Colin Soutar, Itan Barmes, and Casper Stap, "Don't let drivers for quantum cyber readiness take a back seat!" Deloitte, 2023.
11. Matt Scholl, computer security division chief, NIST, Deloitte interview, September 3, 2024.
12. Marc Manzano, general manager, SandboxAQ, Deloitte interview, October 15, 2024.
13. NIST, "NIST releases first 3 finalized post-quantum encryption standards."
14. NIST, "What is post-quantum cryptography?," August 13, 2024.
15. Apple Security Research, "iMessage with PQ3: The new state of the art in quantum-secure messaging at scale," February 21, 2024. iMessage is a trademark of Apple Inc., registered in the U.S. and other countries. Tech Trends is an independent (publication) and has not been authorized, sponsored, or otherwise approved by Apple Inc.
16. Chiara Castro, "Chrome to adopt NIST-approved post quantum encryption on desktop," *techradar*, September 17, 2024.
17. Dan Goodin, "As quantum computing threats loom, Microsoft updates its core crypto library," Ars Technica, Sep. 12, 2024; Paul Smith-Goodson, "IBM Prepares for a Quantum-Safe Future Using Crypto-Agility," *Forbes*, August 8, 2024.
18. NIST, "NCCoE announces technology collaborators for the Migration to Post-Quantum Cryptography project," July 15, 2022.
19. Bill Newhouse, lead, Migration to PQC project at the NCCoE, Deloitte interview, October 16, 2024.
20. Soutar et al., "Don't let drivers for quantum cyber readiness take a back seat!"
21. Michele Mosca, founder and CEO, evolutionQ, Deloitte interview, October 18, 2024.

Continue the conversation

Industry leadership

Colin Soutar
Managing director, Cyber | Deloitte & Touche LLP
+1 571 447 3817 | csoutar@deloitte.com

Dr. Colin Soutar is a managing director within Deloitte & Touche LLP, who leads Deloitte's US and Global quantum cyber readiness program. He's a member of Deloitte's US Government & Public Services (GPS) Cyber practice, where he leads Innovation, Assets, and Ecosystems and Alliances.

Prior to his current role, Dr. Soutar served almost 10 years as the chief technology officer for a Canadian-based biometric and identity management public company. He started his career with a two-year postdoctoral fellowship at NASA Johnson Space Center, developing pattern recognition techniques for autonomous rendezvous and capture operations. He thrives on driving new business opportunities for emerging technologies, within the complex landscape of risk and regulation. He was part of the team in 2013 that developed the National Institute of Standards and Technology (NIST) Cybersecurity Framework, and subsequently helped NIST to develop specific guidance for biometric technologies, identity, IoT, and privacy.

Sunny Aziz
Principal | Cyber & Strategic Risk Services | Deloitte & Touche LLP
+1 713 982 2877 | saziz@deloitte.com

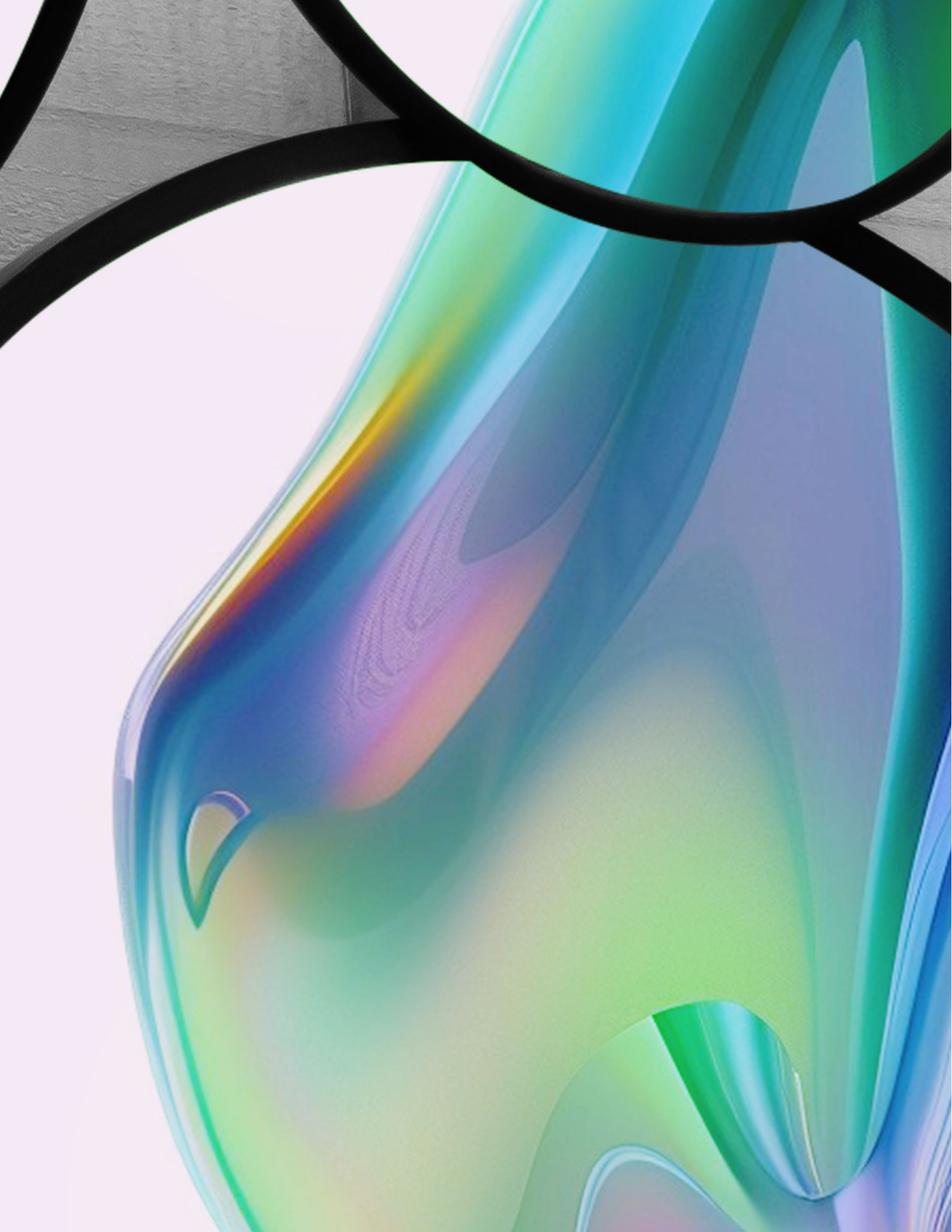
Sunny Aziz is a Principal in Deloitte's Cyber & Strategic Risk Services with over 25 years of experience in assisting clients manage, implement, and operate complex cyber programs. Aziz advises clients on cyber strategies and executing large cyber transformation initiatives. Aziz also serves as Deloitte's Financial Services Industry Insurance sector lead for Cyber, specializing in Managed Security Services, Cyber Strategy & Assessments, Identity & Access Management, Cloud and Infrastructure Security, IT Risk & Compliance Management, Incident Response, Threat & Vulnerability Management, Third Party Risk Management, and Privacy and Data Protection.

Itan Barnes, PhD
Global Quantum Cyber Readiness Capability lead | Deloitte NL
+31 (0)88 288 5589 | ibarmes@deloitte.nl

Itan Barnes leads the cryptography and quantum security capability at the cyber team of Deloitte NL, and also serves as the Global Quantum Cyber Readiness Capability lead. His team focuses on the various aspects of cryptography management such as PKI, certificate lifecycle management, cryptographic key management, and quantum risk.

Acknowledgments

Much gratitude goes to the many subject matter leaders across Deloitte that contributed to our research for the Cyber chapter: **Scott Buchholz**, **Colin Soutar**, and **Masayoshi Terabe**.



CORE MODERNIZATION

The intelligent core: AI changes everything for core modernization

For years, core and enterprise resource planning systems have been the single source of truth for enterprises' systems of records. AI is fundamentally challenging that model.

Kelly Raskovich, Bill Briggs, Mike Bechtel, and Ed Burns

Many core systems providers have gone all in on artificial intelligence and are rebuilding their offerings and capabilities around an AI-first model. The integration of AI into core enterprise systems represents a significant shift in how businesses operate and leverage technology for competitive advantage.

It's hard to overstate AI's transformative impact on core systems. For years, the core and the enterprise resource planning tools that sit on top of it were most businesses' systems of record—the single source of truth. If someone had a question about any aspect of operations, from suppliers to customers, the core had the answer.

AI is not simply augmenting this model; it's fundamentally challenging it. AI tools have the ability to reach into core systems and learn about an enterprise's operations, understand its process, replicate its business logic, and so much more. This means that users don't necessarily have to go directly to core systems for answers to their operational questions, but rather can use whatever AI-infused tool they're most familiar with. Thus, this transformation goes beyond automating routine tasks to fundamentally rethinking and redesigning processes to be more intelligent, efficient, and predictive. It has the potential to unleash new ways of doing business by arming workers with the power of AI along with information from across the enterprise.

No doubt, there will be integration and change management challenges along the way. IT teams will need to invest in the right technology and skills, and build robust data governance frameworks to protect sensitive data. The more AI is integrated with core systems, the more complicated architectures become, and this complexity will need to be managed. Furthermore, teams will need to address issues of trust to help ensure AI systems are handling critical core operations effectively and responsibly.

But tackling these challenges could lead to major gains. Eventually, we expect AI to progress beyond being the new system of record to become a series of agents that not only do analyses and make recommendations but also take action. The ultimate endpoint is autonomous decision-making, enabling enterprises to operate quickly compared with their current pace of operations.

Now: Businesses need more from systems of record

Core systems and, in particular, enterprise resource planning (ERP) platforms are increasingly seen as critical assets for the enterprise. There's a clear recognition of the value that comes from having one system hold all the information that describes how the business operates. For this reason, the global ERP market is projected to grow at a rate of 11% from 2023 through 2030. This growth is driven by a desire for both greater efficiency and more data-driven decision-making.[1]

The challenge is that relatively few organizations are realizing the benefits they expect from these tools. Despite

an acknowledgment that a centralized single source of truth is key to achieving greater operational efficiency, many ERP projects don't deliver. According to Gartner research, by 2027, more than 70% of recently implemented ERP initiatives will fail to fully meet their original business case goals.[2]

Part of the reason ERP projects may fail to align with business goals is that the systems tend to be one-size-fits-all. Businesses needed to mirror their operations to fit the ERP system's model. Applications across the organization were expected to integrate with the ERP. It was the system of record and held all business data and business logic, so the organization acquiesced to these demands, even if they were hard to meet. However, this produced a certain level of disconnect between the business and the ERP system.

AI is breaking this model. Some enterprises are looking to reduce their reliance on monolithic ERP implementations, and AI is likely to be the tool that allows them to by opening up data sets and enabling new ways of working.

New: AI augments the core

With some evolution, ERP systems will likely maintain their current position as systems of record. In most large enterprises, they still hold virtually all the business data, and organizations that have spent the last several years implementing ERP systems will likely be reluctant to move on from them.

Orchestrating the platform approach

In this model, today's core systems become a platform upon which AI innovations are built. However, this prospect raises multiple questions around AI orchestration that IT and business leaders will have to answer. Do they use the modules provided by vendors, use third-party tools, or, in the case of more tech-capable teams, develop their own models? Relying on vendors means waiting for functionality but may come with greater assurance of easy integration.

Another question is how much data to expose to AI. One of the benefits of generative AI is its ability to read and interpret data across different systems and file types. This is where opportunities for new learnings and automation come from, but it could also present privacy and security challenges. In the case of core systems, we're talking about highly sensitive HR, finance, supplier, and customer information. Feeding this data into AI models without attention to governance could create new risks.

There's also the question of who should own initiatives to bring AI to the core. This is a highly technical process that demands the skills of IT—but it also supports critical operational functions that the business should be able to put its fingerprints on.

The answer to these questions will likely look different from use case to use case and even enterprise to enterprise. But teams should think about them and develop clear answers before going all in on AI in the core. These answers form the foundation upon which rests the larger benefits of the technology.

"To get the most out of AI, companies should develop a clear strategy anchored in their business goals," says Eric van Rossum, chief marketing officer for cloud ERP and industries at SAP. "AI shouldn't be considered as a stand-alone functionality, but rather as an integral, embedded capability in all business processes to support a company's digital transformation."[3]

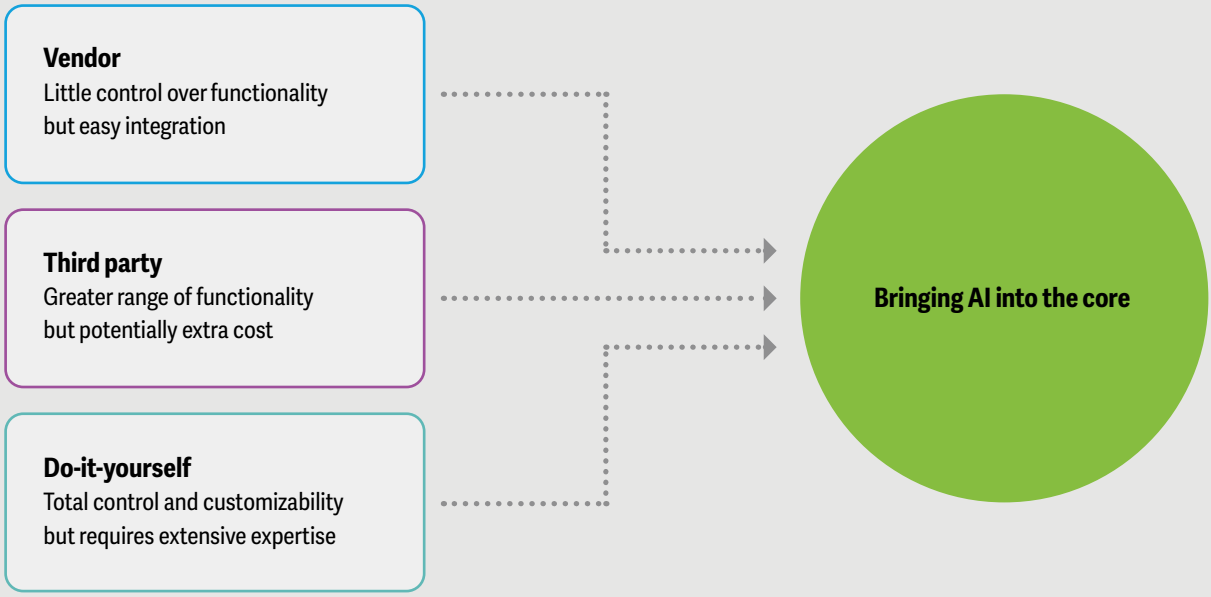
AI enables new ways of working

Forward-looking enterprises are already answering these orchestration questions. Graybar, a wholesale distributor of electrical, industrial, and data communications solutions, is in the middle of a multiyear process of modernizing a 20-year-old core system implementation, which started with upgrades to its HR management tools and is now shifting to ERP modernization. It's leaning on the best modules available from its core systems vendors when it makes sense, while also layering on third-party integrations and homegrown tools when there's an opportunity to differentiate its products and services.[4]

The growth of AI presented leaders at the company with an opportunity to not only upgrade its tech stack, but also to think about how to reshape processes to drive new efficiencies and revenue growth. Trust has been a key part of the modernization efforts. The company is rolling out AI in narrowly tailored use cases where tools

Figure 1

When adding AI functionality to core systems, enterprises have three choices, each with its own benefits and drawbacks



Source: Deloitte research.

only have access to specific databases based on what they need to accomplish the assigned task. And in each instance, humans are kept in the loop to help ensure the accuracy of information that comes from AI tools before it reaches customers.

Graybar is piloting AI in sales and customer service and plans to expand to inventory forecasting and planning. It's adding AI to ordering systems to help surface cross-sell and upsell ideas to sales agents. It's also developing an AI-based tool that will help agents build quotes for customers. The tool will allow workers to use natural language to query product catalogs, pull together options for customers, and compile the information into a communication for the customer.

"These tasks used to take hours or days to complete; now it takes minutes," says David Meyer, chief financial officer at Graybar. "Empowered with AI-based tools, employees can now focus their time on selling and business development versus spending half a day looking for info and typing up a response to a customer request."[5]

This change is about more than just freeing up some time for customer-facing staff. Graybar leadership is eyeing billions of dollars in new revenue growth from expanding its use of AI in core systems. AI in the core is all about driving growth by enabling new ways of working.

Software company ServiceNow is seeing this trend play out with many of its clients, says Michael Park, senior vice president and global head of AI go-to-market at ServiceNow. One especially impactful use case he's seeing is in new employee onboarding. Every new hire needs access to HR systems as well as tools and data specific to their role. In the past, the worker would have had to engage with a range of helpdesk workers, retrieve passwords, log into different systems, and assemble the

credentials they needed to start doing their job. Now, AI enables the HR systems to learn more quickly what new hires need and to automatically provision access by the start date.

This automated learning approach can be applied to all sorts of business processes, Park says. Automating these tasks through gen AI capabilities such as summarization, notes generation, conversational chat, AI search, and task automation may save two minutes or two days, depending on the use case. Once they offload simple workloads to bots, enterprises can redeploy workers to more valuable tasks, like improving service levels, driving margin growth, or developing new product offerings, a trend ServiceNow is seeing with its customers.

"AI in core systems is merely a new capability, a tool to be employed," Park says. "The bigger strategic imperative is using these new capabilities to redefine the status quo for exponential value creation versus just bringing over existing processes onto a new technology capability."[6]

AI in the core, and beyond

As more and more software tools across the enterprise become embedded with AI, workloads that were traditionally owned by core systems could eventually leave the core entirely. With AI, business logic doesn't need to reside in the core. AI can train on structured and unstructured data from across the enterprise. Organizations' business data will be instrumental in developing the most accurate and insightful outputs from AI models. Leveraging the core to help harmonize this data and subsequent AI models for insights will provide companies an opportunity to run their operations on truly insight-driven actions.

In this model, the core becomes just another repository of training data that AI can use to learn and improve business process management. This is where the real power of AI in the core comes in.

Every technology provider knows it needs to build AI into its offerings now, says Chris Bedi, chief customer officer at ServiceNow.[7] ERP systems will continue to be effective as the enterprise's system of record, providing transactional control and reliability as a source of truth. But increasingly, work is being done across domains, with AI as the connective tissue. This means a lot of the major efficiency gains will come from business process innovations happening outside the core.

"AI tech built into systems of record is going to be decent at incremental improvements to existing ways of working," Bedi says. "But for that step function change, it has to come from AI that works across domains, that takes advantage of data that's not just resident of one system of record, [that] can look at all of it, run the model on all of it, take actions across all of it. That's the real unlock here."[8]

Next: More automation creates opportunities—and potential risks

For many enterprises, core modernization has been a years-long, ongoing task. They may be tempted to view AI as just the latest look to something they're already familiar with. This may not be the right mindset.

This modernization will likely look very different from past rounds. The speed and scale of change will likely be faster and larger than previous efforts. In the past, modernization was primarily about implementing upgrades, a laborious and time-consuming task, but nevertheless one that was well understood. Software vendors typically provide an upgrade path to give their users a playbook to follow.

This time around, there is no prewritten playbook. The architecture will likely be different because a lot of it will involve AI modules in peripheral software interacting with core systems. Rather than the business aligning everything it does with the core, now the core has to be aligned with what the business is doing. This may become particularly challenging when enterprises take advantage of AI to create new business processes backed by core data. The job becomes more complex and demands more expertise and different skills. Similar to what we discuss in **"IT, amplified: AI elevates the reach (and remit) of the tech function,"** understanding business problems will become a crucial skill for IT teams adding AI to their core systems. This will likely be a major change for IT workers who, in the past, advanced their careers based on deep technical expertise.

Once core systems are modernized through AI, maintaining them becomes a very different exercise. As mentioned in **"What's next for AI?"** AI agents could soon execute many core functions. Imagine a customer service bot that can interact with customers, understand their issues, and diagnose problems. This bot may then be able to interact with another bot that can take actions like process returns or ship new items. Leading companies are already starting to do this. For example, luxury retailer Saks' customer service bots can interact with ordering and inventory systems to smooth delivery of items bought online, ease returns, and empower customer service representatives.[9] In the truly agentic future, we expect to see more of these kinds of bots that work autonomously and across various systems. Then, maintaining core systems becomes about overseeing a fleet of AI agents.

Done wisely, AI may help reduce technical debt for core systems and push for a cleaner core, which could make enterprise systems less complex to maintain and cater to business demand in a more agile manner.

The core is on the cusp of a major AI-driven revolution. Early adopters are riding the first crest of this wave to increased efficiency and new ways of generating revenue, but soon enterprises will likely turn over much larger core functions to autonomous agents. It remains to be seen what organizations will do with the improved efficiency and effectiveness that come with this change. But the opportunity exists to reshape not just how the core operates but, at a more fundamental level, how business gets done.

Endnotes

1. Grand View Research, *ERP software market size, share & growth report*, 2030, accessed November 5, 2024.
2. Gartner, "Enterprise resource planning to optimize operations," accessed November 5, 2024.
3. Eric van Rossum, chief marketing officer for cloud ERP and industries, SAP, interview with the author, October 1, 2024.
4. David Meyer, chief financial officer, Graybar, interview with the author, September 26, 2024.
5. Ibid.
6. Michael Park, senior vice president and global head of AI go-to-market, ServiceNow, interview with the author, September 19, 2024.
7. Chris Bedi, chief customer officer, ServiceNow, interview with the author, September 29, 2024.
8. Ibid.
9. Salesforce, "Saks elevates luxury shopping with unified data and AI service agents," accessed November 5, 2024.

Continue the conversation

Industry leadership

Thorsten Bernecker
Application Modernization & Migration practice leader | Principal | Deloitte Consulting LLP
+1 512 226 4418 | tbernecker@deloitte.com

Thorsten Bernecker is a principal with Deloitte Consulting LLP and leads the Application Modernization & Migration practice within the Application Modernization & Innovation offering. Having founded the software company innoWake, and growing it from a small business to the global leader for modernizing legacy technology, he has been able to unite a keen eye for disruptive technology with a business sense for successfully taking a small startup through an exponential growth stage. Deloitte acquired innoWake in 2017 and Bernecker now heads global strategy and leadership for this group.

Zillah Austin
Technology Strategy & Business Transformation | Principal | Deloitte Consulting LLP
+1 202 716 5974 | ziaustin@deloitte.com

Zillah Austin is a principal at Deloitte Consulting within the Technology Strategy and Business Transformation practice. She has more than 20 years of industry and consulting experience, leading and managing business transformations and delivering large-scale technology solutions for Global Energy, Resources and Industrials clients. Austin defines and implements IT strategies for energy and resources clients, helping major corporations operationalize business and IT strategies, while also improving the effectiveness of their technology processes. She has deep expertise in aligning IT strategies to architecture, governance, program management, operating models, and service management.

Abdi Goodarzi
US Enterprise Performance Portfolio leader | Principal | Deloitte Consulting LLP
+1 714 913 1091 | agoodarzi@deloitte.com

Abdi Goodarzi is a principal with Deloitte Consulting LLP, leading Deloitte's Enterprise Performance (EP) Offerings Portfolio. This portfolio of six offerings provides strategy, implement and operate services for a variety of enterprise functions, including end-to-end Business and IT transformation, digital supply chain optimization, manufacturing and product strategies, procurement as-a-service, global finance, shared services, planning, ITSM, and full scale AMS and BPO. This portfolio offers competency in many ERP platforms such as SAP, Oracle, Workday Financials and Infor, in addition to ServiceNow, Anaplan, Ariba and Coupa. Plus Real Estate solutions such as Nuvolo, as well as PLM, Planning and Fulfillment and engineering solutions like Siemens, PTC, O9, OMP and IBP.

Acknowledgments

Much gratitude goes to the many subject matter leaders across Deloitte that contributed to our research for the Core Modernization chapter: **Zillah Austin**, **Thorsten Bernecker**, **Lars Cromley**, **Tim Gaus**, **Abdi Goodarzi**, **Kelly Herod**, **Chip Kleinheksel**, **Kasey Lobaugh**, and **Jinlei Lui**.

CONCLUSION

Breadth is the new depth: The power of intentional intersections

In an increasingly convergent world, enterprises would do well to explore intentional industry and technology intersections that propel innovation across boundaries

Mike Bechtel and Raquel Buscaino

In consulting, we often rely on the MECE problem-solving framework, which proposes that a problem can be more readily solved if it can be broken down into distinct "mutually exclusive" (ME) tasks that, when taken in sum, provide a "collectively exhaustive" (CE) solution.

Yet, we are increasingly living in a convergent world where the MECE principle isn't always easily applied. That's evident in this report's six chapters. Although we've neatly packaged six trends into distinct chapters, they're far from separate and isolated. For that matter, neither are today's technologies, organizations, and industries—and most of the rest of the world. Increasingly, separation, segmentation, and specialization are being replaced by a complex web of intersections—a convergence of "unusual suspects" that can be found across both industries and technologies. Consider the confluence of blockchain and generative artificial intelligence for better detection of, and protection from, synthetic media; or that of space tech and biotech, for protecting astronauts from the effects of long-term space travel.

Companies have long relied on innovation-driven revenue streams, synergies created through mergers and acquisitions, and strategic business partnerships to drive new growth. More than ever, they should double down on such intentional, dedicated pursuits of breadth. The business case for breadth reveals that the most promising (and profitable) futures will likely emerge from industry and technology convergence. This convergence can help uncover two key perspectives:

1. Insight into adjacent industries whose current research and development efforts might hold the keys to an organization's future

2. Clarity on how different technologies might be combined so that the sum is greater than its respective parts: synergy, if you will—a concept that has itself gone through the hype cycle and emerged intact

Let's take a deeper dive into each of these.

Industry intersections: Exploring beyond industry boundaries

Cyberpunk science fiction writer William Gibson is often attributed with the well-known quote, "The future is already here; it's just not evenly distributed."[1]

Overused? Yes. Relevant now more than ever? Also yes. Gibson's statement can help leaders see that their organizations' next big breakthrough likely exists today in another industry, geography, or competitor.

Let's take a look at the space and life sciences industries. One could argue that there's minimal synergy between the two, but we'd counter with the following example:

The unique properties of microgravity in space allow for pharmaceutical product inputs to be developed with more uniformity and higher production quality.[2]

Although the idea of manufacturing in microgravity might seem fantastical, it's far from theoretical: Companies like Eli Lilly and Merck are already investing in this possibility.[3] Biopharma companies that overlook the space sector as a relevant partner could miss a potential discovery that could directly have an impact on their core business.

Many other examples of industry convergence reiterate the importance of searching beyond one's own industry for innovative solutions and answers. Auto giants Toyota and Mitsubishi Heavy Industries are partnering with space agencies to build lunar rovers,[4] while clothing retailer lululemon is partnering with biotech companies such as LanzaTech and Samsara Eco to develop more sustainable fabrics.[5] Meanwhile, food delivery now accounts for about a third of transportation company Uber's total revenues,[6] and e-commerce leader Amazon has made significant strides in the health care sector with Amazon Pharmacy.[7]

Tech intersections: Compounding growth and integration

Whereas industry intersections can serve as a wide-angle camera lens for searching adjacent industries for insight, technology intersections offer a slightly different perspective. They help us better understand how technologies and innovations can compound growth.

Technologies are tools, often applied to specific problems. But what separates a hammer from a jackhammer is that a jackhammer is the combination of several tools (a hammer, chisel, and an energy source) that together create a more efficient tool. Rather than viewing technologies in isolation, it's important to think of them as tightly integrated, with the ability to compound each other's growth.

For example, quantum machine learning applies quantum computing principles to machine learning programs to increase efficiency. Networking technologies like 5G networks and edge computing are so tightly coupled that they are often grouped into a singular shorthand name, 5G edge. And as we discussed in **"Hardware is eating the world,"** smart factories are combining computer vision, sensors, and data to build machines that can learn and improve, potentially leading to the development of humanoid robotics.[8]

And what about artificial intelligence, the tool of the moment? We discussed in our **introduction** the expectation that AI will eventually become as ubiquitous and foundational as electricity, which suggests that it will have endless convergence points with all manner of downstream technologies. As just one example, let's explore the intersection of AI and robotics. Although both technologies can be viewed distinctly, the real magic happens when they are combined—when mechanical minds meet mechanical muscles. AI enables robots to operate autonomously, allowing the robots to collect more data about the world and their movement through it, which is, in turn, fed into the AI algorithm's training data, improving the algorithm itself. When viewing technologies as intersectional by nature, we can start to see the **flywheel effect** bolstering growth and innovation.

What does this mean for business and technology leaders? While having "mutually exclusive" technology teams focused on à la carte technologies is functionally efficient, it's also imperative to build bridges between teams. Choosing slightly improved hammers over a jackhammer is forfeiting innovation for the tyranny of incrementalism.

Renaissance reimagined

The term "renaissance person" embodies an ideal, in a time of rapid change around science, art, and commerce, that people who build expertise across several areas of knowledge are poised to lead. In today's world, accelerating industry and technology intersections affirm that breadth is the new depth. Generalists are needed more than ever. As the amount of available information approaches infinity, so, too, does the demand for interdisciplinary dot connectors—the big-picture thinkers who can identify correlations and links between seemingly unrelated industries, technologies, and other ideas.

If, as we mentioned, AI becomes as ubiquitous as electricity, the second- and third-order effects could be profound. The advent of electricity influenced immense

changes in society, such as urban migration, industrialization, and radio communication.[9] We may be on the cusp of similar changes through AI that alter the way we work, live, and communicate. Expertise in historical methods may not be as important as the vision to imagine and execute new intersections of AI with the **macro technology forces** we've covered in this report, such as AI applied to spatial computing and core modernization.

For leaders, this serves as a nudge to see odd-combination dual degrees, bridges between disparate teams, and interest in adjacent industries as necessary *features,* not bugs. If organizations can see beyond the silos of specialization and embrace these intentional intersections, we might very well find ourselves on the cusp of a reimagined renaissance. What convergence will your organization discover next?

Endnotes

1. *The Economist*, "Broadband blues," June 21, 2001.
2. Axiom Space, "Protein crystallization," accessed October 2024.
3. Ibid.
4. Natsumi Kawasaki, "Toyota, Mitsubishi heavy to work together on lunar vehicles," *Nikkei Asia*, July 21, 2023.
5. *Bio.News*, "LanzaTech x Lululemon collab births a new sustainable fashion item," April 24, 2024; lululemon, "lululemon and Samsara Eco unveil world's first enzymatically recycled nylon 6,6 product," press release, February 20, 2024.
6. Arjun, "How Uber makes revenue: Key streams and strategies explained," Appscrip, September 19, 2024.
7. Bruce Japsen, "Amazon rolls out same-day prescription delivery with help from AI," *Forbes*, March 26, 2024.
8. Majeed Ahmad, "Sensor fusion with AI transforms the smart manufacturing era," *EE Times*, July 26, 2023.
9. Smithsonian Institution, "Lighting a Revolution", accessed October 2024.

About the authors

Kelly Raskovich
kraskovich@deloitte.com

Kelly Raskovich is a senior manager and lead within Deloitte's Office of the CTO (OCTO), and serves as the executive editor of Tech Trends, Deloitte's flagship report on emerging technologies. Her mission is to educate clients, shape the future of Deloitte's technology brand and offerings, cultivate talent, and enable businesses to achieve future growth. She is responsible for technology eminence, client engagement, and marketing/PR efforts. Prior to her leadership role, she led several data and analytics projects for global Fortune 500 organizations across the oil and gas industry.

Bill Briggs
wbriggs@deloitte.com

As chief technology officer, Bill Briggs helps clients anticipate the impact that emerging technologies may have on their business in the future and how to get there from the realities of today. He is responsible for research, eminence, and incubation of emerging technologies affecting clients' businesses and shaping the future of Deloitte Consulting LLP's technology-related services and offerings. Briggs also serves as executive sponsor of Deloitte's CIO Program, offering CIOs and other technology executives insights and experiences to navigate the complex challenges they face in business and technology.

Briggs earned his undergraduate degree in Computer Engineering from the University of Notre Dame, and his MBA from the Kellogg School of Management at Northwestern University. He proudly serves on the board of directors for the Kids In Need Foundation, partnering with teachers and students in under-resourced schools and providing the support needed for teachers to teach and learners to learn.

Mike Bechtel
mibechtel@deloitte.com

As chief futurist with Deloitte Consulting LLP, Mike Bechtel helps clients develop strategies to thrive in the face of discontinuity and disruption. His team researches the novel and exponential technologies most likely to impact the future of business, and builds relationships with the startups, incumbents, and academic institutions creating them.

Prior to joining Deloitte, Bechtel led Ringleader Ventures, an earlystage venture capital firm he cofounded in 2013. Before Ringleader, he served as CTO of Start Early, a national not-for-profit focused on early childhood education for at-risk youth. Bechtel began his career in technology research and development at a global professional services firm where his dozen US patents helped result in him being named that firm's global innovation director. He currently serves as professor of corporate innovation at the University of Notre Dame.

Ed Burns
edburns@deloitte.com

Ed Burns leads the client stories initiative within the Office of the CTO known as Trend Lines. This project serves as a key research input to Tech Trends and other eminence. Prior to his current role he led a tech news publication that covered all things AI, analytics, and data management.

Abhijith Ravinutala
aravinutala@deloitte.com

Abhijith Ravinutala is a professional storyteller with Deloitte's Office of the CTO. Through research, writing, and presentations, he helps Deloitte and its clients envision and better prepare for the future of technology. His background in strategy consulting has exposed him to a variety of industries, and as a writer he takes keen interest in the intersections of technology ethics, AI, and human impacts. In addition to writing Tech Trends, he has also led Deloitte publications on AI and CEOs, xTech Futures: BioTech, and the Dichotomies series, recently featured at SXSW 2024.

Raquel Buscaino
rbuscaino@deloitte.com

Raquel Buscaino leads Deloitte's Novel & Exponential Technologies (NExT) team where she and her team sense, and make sense of, emerging technologies that are likely to change the way we work and live. From brain-computer interfaces and synthetic biology to space exploration and quantum computing, Buscaino and her team aim to distill signal from noise, value from hype, and profitable actions from ambiguous concepts. The NExT team uses this research to create world-class thought leadership, such as Deloitte Tech Trends and xTech Futures publications.

Buscaino is also the host of the Deloitte *TECHTalks podcast* where she interviews industry leaders about what's new and next in tech. Prior to her leadership on Deloitte's NExT team, she worked in Deloitte's blockchain & digital assets practice where she specialized in blockchain consortiums and led global blockchain workshops for Deloitte and its clients.

Acknowledgments

Special thanks

Ed Burns, **Heidi Morrow**, and **Abhijith Ravinutala** for being the creative engine powering Tech Trends. Ed and Abhi, your exceptional dedication, leadership, and editorial chops have truly elevated our work, not to mention your ability to deftly weave research and interviews into compelling narratives and your flexibility in managing feedback from multiple stakeholders. And Heidi, thank you for being a standard bearer for the principles of excellent design while enthusiastically embracing new ways of leading the design portion of Tech Trends. The beautiful report imagery, figures, videos, and other graphics are a testament to your leadership. We're lucky and thankful that the three of you are part of the team.

Sarah Mortier for diving headfirst into your new role of managing Tech Trends production and making it yours. It has been a thrill to watch your confidence grow as you identified challenges, proposed improvements, and ultimately kept us on track to deliver the editorial earlier than ever before. We appreciate you for being an eager and enthusiastic learner and we can't wait to see how you will "wow" us in year two.

Caroline Brown, for leading the Tech Trends editorial and design production team with good cheer, humor, and grace under pressure. Your leadership and strategic vision have been instrumental in taking Tech Trends to the next level, and we're incredibly grateful for you.

Imelda Mendoza and **Bella Stash** for the breath of fresh air you brought to the Tech Trends process by pitching in with research, data, and PMO support. We appreciate your enthusiastic and cheerful willingness to tackle whatever came your way.

MacKenzie Hackathorn, **Haley Gove Lamb**, **Kiran Makhijani**, and **Angel Lacambra**, for the work you'll do in bringing Tech Trends to life for our clients and account teams. Thank you for taking our work and making it real.

Stefanie Heng for your continued willingness to step in and help us figure out Tech Trends and the publication process. We appreciate your commitment to our team, even as you transition to a new role, and we look forward to seeing the heights you'll fly to next. We'll miss you more than you can ever know!

Deanna Gorecki, **Ben Hebbe**, **Bri Henley**, **Tracey Parry**, **Abria Perry**, **Madelyn Scott**, and **Mikaeli Robinson** for your unwavering dedication and innovative strategies in promoting Tech Trends. Your tireless efforts in marketing, communications, and PR significantly amplify our reach and impact year after year. Thank you for recognizing and believing in the value and impact of Tech Trends.

Taylor Brockman, **Raquel Buscaino**, **Lucas Erb**, **Danny Greene**, **Mark Osis**, and **Hillary Umphrey** for being our brain trust as we identified trends and for conducting initial research and showing us which way the compass is pointing in the long term. Thank you for generously sharing your knowledge with us and helping us hone our research craft.

Hannah Bachman, **Aditi Rao**, and the entire Deloitte Insights team for evolving our partnership and growing with us as we continue to look for ways to improve Tech Trends. We appreciate your continued support, flexibility, and grace as the needs of our team and our practice change.

Sylvia Chang, **Manya Kuzemchenko**, **Melissa O'Brien**, **Molly Piersol**, **Natalie Pfaff**, **Harry Wedel**, **Jaime Austin**, **Govindh Raj**, **Megha Priya**, **Naveen Bhusare**, and all the Marketing Excellence creative team members who helped develop report images and figures. Your creativity and dedication have resulted in a beautiful report and hub page that exceeds all expectations. We are grateful not only for your artistic vision and the captivating visuals that bring our work to life, but also for your commitment to collaboration and exploration.

Additional thanks

The authors would like to thank the Office of the CTO Market-Making team, without whom this report would not be possible: **Caroline Brown**, **Ed Burns**, **MacKenzie Hackathorn**, **Stefanie Heng**, **Bri Henley**, **Dana Kublin**, **Angel Lacambra**, **Haley Gove Lamb**, **Kiran Makhijani**, **Sangeet Mohanty**, **Heidi Morrow**, **Sarah Mortier**, **Abria Perry**, **Abhijith Ravinutala**, and **Bella Stash**.

Continue the conversation

Our insights can help you take advantage of emerging trends. If you're looking for fresh ideas to address your challenges, let's talk.

The Office of the CTO

The Deloitte US Office of the CTO is a team centered on engineering technology futures. We identify, research, and incubate emerging technology solutions to shape demand for future markets, cultivate talent, and enable businesses for future growth.

If you'd like to connect and discuss more, please feel free to contact us at OCTO@deloitte.com.

Executive editor



Kelly Raskovich

Client & Marketing Lead, Office of the CTO
Deloitte Consulting LLP
kraskovich@deloitte.com

Kelly Raskovich is a senior manager and lead within Deloitte's Office of the CTO (OCTO), and serves as the executive editor of Tech Trends, Deloitte's flagship report on emerging technologies. Her mission is to educate clients, shape the future of Deloitte's technology brand and offerings, cultivate talent, and enable businesses to achieve future growth. She is responsible for technology eminence, client engagement, and marketing/PR efforts. Prior to her leadership role, she led several data and analytics projects for global Fortune 500 organizations across the oil and gas industry.

Executive sponsors



Bill Briggs
Global chief technology officer
Deloitte Consulting LLP
wbriggs@deloitte.com

As chief technology officer, Bill Briggs helps clients anticipate the impact that emerging technologies may have on their business in the future and how to get there from the realities of today. He is responsible for research, eminence, and incubation of emerging technologies affecting clients' businesses and shaping the future of Deloitte Consulting LLP's technology-related services and offerings. Briggs also serves as executive sponsor of Deloitte's CIO Program, offering CIOs and other technology executives insights and experiences to navigate the complex challenges they face in business and technology.

Bill earned his undergraduate degree in Computer Engineering from the University of Notre Dame, and his MBA from the Kellogg School of Management at Northwestern University. He proudly serves on the board of directors for the Kids In Need Foundation, partnering with teachers and students in under-resourced schools and providing the support needed for teachers to teach and learners to learn.



Mike Bechtel
Chief futurist
Deloitte Consulting LLP
mibechtel@deloitte.com

As chief futurist with Deloitte Consulting LLP, Mike Bechtel helps clients develop strategies to thrive in the face of discontinuity and disruption. His team researches the novel and exponential technologies most likely to impact the future of business, and builds relationships with the startups, incumbents, and academic institutions creating them.

Prior to joining Deloitte, Bechtel led Ringleader Ventures, an early-stage venture capital firm he cofounded in 2013. Before Ringleader, he served as CTO of Start Early, a national not-for-profit focused on early childhood education for at-risk youth. Bechtel began his career in technology research and development at a global professional services firm where his dozen US patents helped result in him being named that firm's global innovation director. He currently serves as professor of corporate innovation at the University of Notre Dame.

Deloitte.
Insights

Sign up for Deloitte Insights updates at **www.deloitte.com/insights**

Deloitte Insights contributors

Editorial: Aditi Rao, Hannah Bachman, Debashree Mandal, Pubali Dey, and Cintia Cheong
Creative: Manya Kuzemchenko, Sylvia Yoon Chang, Natalie Pfaff, Molly Piersol, Harry Wedel, and Govindh Raj
Deployment: Atira Anderson
Cover artwork: Manya Kuzemchenko and Sylvia Yoon Chang; Getty Images, Adobe Stock

About Deloitte Insights

Deloitte Insights publishes original articles, reports and periodicals that provide insights for businesses, the public sector and NGOs. Our goal is to draw upon research and experience from throughout our professional services organization, and that of coauthors in academia and business, to advance the conversation on a broad spectrum of topics of interest to executives and government leaders.

Deloitte Insights is an imprint of Deloitte Development LLC.

About this publication

This publication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or its and their affiliates are, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your finances or your business. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser. None of Deloitte Touche Tohmatsu Limited, its member firms, or its and their respective affiliates shall be responsible for any loss whatsoever sustained by any person who relies on this publication.

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms.

Copyright © 2024 Deloitte Development LLC. All rights reserved.
Member of Deloitte Touche Tohmatsu Limited