IDC FutureScape

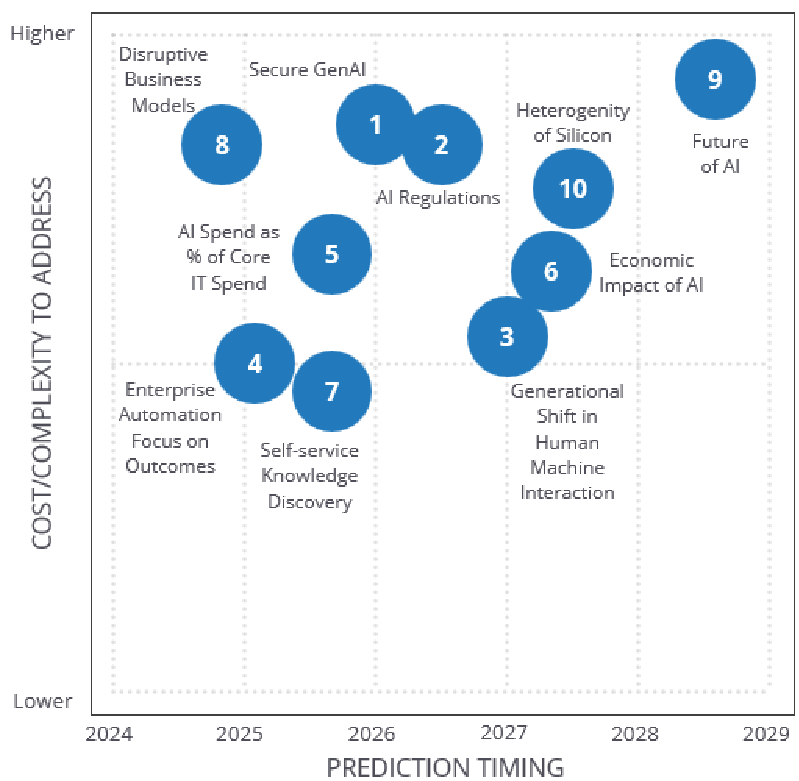
IDC FutureScape: Worldwide Artificial Intelligence and Automation 2024 Predictions

Ritu Jyoti, David Schubmehl, Hayley Sutherland, Maureen Fleming
Shari Lava, Neil Ward-Dutton, Jennifer Cooke, Laurie Buczek
Melih Murat, Jennifer Hamel, Deepika Giri, Shane Rau
Andrew Gens, Matt Arcaro, Kathy Lange, Mary Wardley
Nobuko Iisaka, Raghunandhan Kuppuswamy, Michele Rosen

IDC FUTURESCAPE FIGURE

FIGURE 1

IDC FutureScape: Worldwide Artificial Intelligence and Automation 2024 Top 10 Predictions



Note: Marker number refers only to the order the prediction appears in the document and does not indicate rank or importance, unless otherwise noted in the Executive Summary.

Source: IDC, 2023

October 2023, IDC #AP50341323

EXECUTIVE SUMMARY

According to IDC's Worldwide Semiannual Artificial Intelligence Systems Spending Guide, 2023 (V2) – which tracks artificial intelligence (AI) software, hardware, and services across industries and use cases – enterprises worldwide are expected to invest $166 billion on AI solutions in 2023. This spending is expected to grow to $423 billion at a compound annual growth rate (CAGR) of 26.9% for 2022–2027. This is more than four times greater than the five-year CAGR of 5.7% for worldwide IT spending over the same period. In addition, worldwide intelligent process automation (IPA) software will reach $65.3 billion in 2027, growing at a CAGR of 21.7% from 2021 to 2026. In 2022, the IPA market grew 16.7% to $24.5 billion compared with 2021's growth of 20.5% to $21.0 billion.

In this study, the global team of IDC analysts describe key drivers affecting IT and business decision makers responsible for this spending and the effective use of associated solutions. This study also presents the top 10 predictions affecting AI and automation initiatives through 2028.

Each prediction is assessed based on its impact (a mix of cost and complexity to address) and time frame to the expected stated adoption level. This study also offers IDC analysts' guidance to IT and business decision makers as they develop or revise their strategies and create resource allocation plans for investment in AI and automation.

The following 10 predictions represent the expected trends with the greatest potential impact on artificial intelligence and automation initiatives:

- **Prediction 1:** By 2026, 70% of cloud and software platform providers will bundle GenAI safety and governance packages with their primary services, reducing GenAI risk scenarios by three times.
- **Prediction 2:** By 2026, AI regulatory divergence across geographies will create major challenges for multinational organizations, increasing implementation time and effort for sensitive use cases by up to 10%.
- **Prediction 3:** By 2027, GenAI digital assistants will be the UI for 25% of interactions with enterprise software, including software development and higher for operating business processes like customer engagement.
- **Prediction 4:** After a temporary focus on GenAI, 60% of enterprises will refocus on mandating automation strategies based on outcomes rather than specific technologies by 2025.
- **Prediction 5:** By 2025, the G2000 will allocate over 40% of core IT spend to AI initiatives, leading to a double-digit increase in the rate of product and process innovations.
- **Prediction 6:** By 2027, the overall economic impact of AI will neutralize, as organizations overcome initial disruption and refocus resources on innovation and new business opportunities, fueling economic expansion.
- **Prediction 7:** By 2025, two-thirds of businesses will leverage a combination of generative AI and RAG to power domain-specific, self-service knowledge discovery, improving decision efficacy by 50%.
- **Prediction 8:** By 2024, 33% of G2000 companies will exploit innovative business models to double their monetization potential of generative AI.

- **Prediction 9:** By 2028, 30% of G2000 companies will experiment with AGI (currently speculative) systems that will have a transformative effect on society and create significant opportunities and threats.
- **Prediction 10:** By 2027, the spending on server accelerators (GPU, FPGA, and AI ASIC and ASSP) compared with server CPUs will invert, achieving a 55/45 rate.

This IDC study provides IDC's 2024 top 10 predictions for artificial intelligence and automation.

"ChatGPT's explosive global popularity has given us AI's first true inflection point in public adoption," says Ritu Jyoti, group vice president, Worldwide Artificial Intelligence and Automation Market Research and Advisory Services at IDC. "As AI and automation investments grow, focus on outcomes, governance, and risk management is paramount."

IDC FUTURESCAPE PREDICTIONS

Summary of External Drivers

- **AI everywhere —** Generative AI takes the spotlight
- **The drive to automate —** Maximizing efficiency and new opportunities
- **Shifting tech regulatory landscape —** Navigating risk and opportunity
- **The digital business imperative —** Competitiveness and outcomes
- **Dynamic work and skills requirements —** New work mode era
- **Everything as a service intensifies —** Transforming models to drive change
- **Geopolitical turbulence —** Sovereignty under emerging fragmentation

Predictions: Impact on Technology Buyers

Prediction 1: By 2026, 70% of Cloud and Software Platform Providers Will Bundle GenAI Safety and Governance Packages with Their Primary Services, Reducing GenAI Risk Scenarios by Three Times

The excitement that organizations are feeling today about the promise of GenAI and the impact on their business operations and interactions with customers will be tempered by the significant risks that GenAI can also introduce. Accelerated efficiency and catastrophic risk are the two sides to the shiny new GenAI coin. Deep fakes, toxicity, hallucinations, copyright infringement, prompt injections, scams and advanced phishing, and accelerated data privacy breaches are just a few of the ways GenAI could be used to disrupt businesses, steal identities, and influence outcomes.

The most effective approach to reducing this risk is to "fight fire with fire" — essentially turning to the power of GenAI to identify bad actors and issues. Cloud and software platform providers can leverage GenAI technologies to identify patterns and anomalies and identify potential risks and remediate problems before they become widespread. Although they will not be able to remove all the risks, they will be able to speed risk mitigation by three times, providing better security and resilience for their customers.

Cloud and software platform providers will bundle GenAI safety and governance packages with their primary services to add value and differentiate their offerings. For example, content moderation platforms can use AI models to detect offensive or inappropriate content in text and images quickly and efficiently to create better online experiences for everyone. Software platform providers can watermark AI-generated images and videos to tell them apart from real things. The ability to protect an organization will be just as important as the core services they provide. Organizations that invest in cloud and software platform providers as their partners on the GenAI journey will benefit from the deep investments and expertise these providers offer.

Associated Drivers

- **AI everywhere —** Generative AI takes the spotlight
- **The digital business imperative —** Competitiveness and outcomes
- **Shifting tech regulatory landscape —** Navigating risk and opportunity

IT Impact

- The scale and sophistication of GenAI-fueled digital safety issues will be more than most organizations can address with internal IT resources.
- IT organizations will only be as secure as their technology and service partners. IT leadership will focus a significant amount of time and resources on vetting out their solution providers to ensure that they are prepared to identify and remediate risks.

Business Impact

- GenAI-fueled digital safety and security breaches can cause businesses downtime, loss of customer trust and, potentially, loss of customers.
- With many software vendors across industries looking at the inclusion of GenAI capabilities in their offerings, the inclusion of GenAI safety detection or security will be a major potential differentiator going forward, especially when looking at GenAI use in regulated industries.

Guidance

- Lean on trusted partners that have deep expertise and GenAI resources that can help define, plan, and execute your secure GenAI strategy.
- Establish a robust provider evaluation and assessment framework.

Prediction 2: By 2026, AI Regulatory Divergence Across Geographies Will Create Major Challenges for Multinational Organizations, Increasing Implementation Time and Effort for Sensitive Use Cases by Up to 10%

As AI technologies and systems become progressively more integrated into people's lives, concerns regarding their ethical implications, potential risks, and misuse are now commonplace. The European Union (EU) embarked on the journey to regulate the deployment and development of AI systems through the introduction of the EU AI Act in 2019, and it is now progressing toward ratification. If current negotiations between EU authorities and individual EU states are successful, the Act will come into effect in June 2024, with a two-year transition period for AI developers and providers to adjust. The regulation identifies four risk categories for AI applications.

The approach being taken in the United States indicates a modest shift toward enforcing protections for individuals and a degree of control over AI implementation, although, so far, no new regulation is planned. In the Asia/Pacific region, meanwhile, governments are much more cautious about regulating AI — broadly speaking, governments in the region see regulation of AI as impeding the spirit of innovation in progressive digital economies and weakening the strategic roles that their countries may play in the development of AI technologies and solutions. Separately, Canada is pressing ahead with an Artificial Intelligence and Data Act (AIDA) that follows a use case risk categorization approach like the EU's AI Act; India has also hinted at a use case-based approach in its Digital Act under development.

AI is already used extensively across public and private sectors, and technology, solution, and services providers are very likely to push back strongly against heavy regulation of AI technology that makes no distinction between different AI use cases. The result will be that some AI use cases will likely be untouched, even where countries are pushing ahead strongly on AI regulation; whereas the implementation of other use cases, based on their potential to infringe on individuals' rights, will likely be placed under some degree of restriction in many countries. Beyond potentially requiring the use of additional legal/risk advisory services upfront, diverse regulatory requirements are likely to result in organizations taking more phased approaches to AI rollouts, which will also increase time to value.

Associated Drivers

- **Shifting tech regulatory landscape —** Navigating risk and opportunity
- **Geopolitical turbulence —** Sovereignty under emerging fragmentation
- **AI everywhere —** Generative AI takes the spotlight

IT Impact

- Model management, explainability, and provenance will become important for regulatory compliance.
- Some AI use cases will require regional/local sovereign rather than global platforms to support them to ensure organizations can comply with local regulations.
- More technologies and more vendors will complicate enterprise AI strategies and require challenging data integration.
- IT will be expected to be aware of changes in regulations and adapt AI systems appropriately.

Guidance

- Reframe your data, analytics, and AI strategies within national or regional boundaries. For example, you would need to cordon off sensitive data, models, and associated processes within regional boundaries to comply with local rules and regulations.
- Note that AI cannot be managed as a one-off project. Apply a systematic approach to model development and management as the model life cycle becomes more important than ever with complex regulations.
- Examine your vendor ecosystem across AI products and services. Ensure you have the right tools and expertise to manage a fragmented regulatory landscape.
- Establish an AI governance board.

Prediction 3: By 2027, GenAI Digital Assistants Will Be the UI for 25% of Interactions with Enterprise Software, Including Software Development and Higher for Operating Business Processes Like Customer Engagement

The world of artificial intelligence has shifted dramatically since the release of ChatGPT on November 30, 2022. While many AI research organizations have been experimenting with large language models (LLMs) and GenAI, OpenAI created viral interest in conversational interfaces based on the GPT-3 LLM. Since that time, consumers, developers, end users, and many organizations have experimented with conversational interfaces as an easy and quick way to interact with LLMs to answer questions, offer dialog, create code, and even perform automation.

In addition to this, the need for intelligent conversational assistance within organizations has grown rapidly over the past several years, fueled initially by the COVID-19 pandemic and now via the requirements of digital business. To date, there are almost 100 companies seeking to address these emerging interface opportunities, and new companies are continuing to emerge. As CIOs, CMOs, and other senior executives embark on the use of conversational AI to increase revenue, enhance customer satisfaction, improve internal productivity, or broaden their customer communication channels, vendors are rapidly adding and integrating GenAI into their conversational AI platforms.

The combination of these two trends will create a new trend: conversation as the standard user interface for both enterprise and consumer applications and solutions. IDC is already seeing that conversational AI interfaces are significantly affecting customer engagement, sales, marketing, and even the IT help desk. Many organizations are engaging with conversational AI vendors and contact center vendors to incorporate conversational AI into their overall solution mix.

As GenAI continues to evolve and improve, conversation will be the de facto standard for enterprises going forward. Already, numerous software vendors have announced or released conversational digital assistants based on GenAI techniques and models. Organizations need to prepare for this impending deluge of digital assistants and consider whether building their own GenAI-based digital assistant on top of a large language model fueled with their own proprietary data is necessary and worthwhile. Already, IDC has seen organizations announce such internal applications.

Associated Drivers

- **AI everywhere —** Generative AI takes the spotlight
- **Dynamic work and skills requirements —** New work mode era
- **The drive to automate —** Maximizing efficiency and new opportunities

IT Impact

- Due to regional AI regulations and laws, conversation as an interface will need to ensure regional and local compliance, which may require additional due diligence on the part of worldwide IT.
- IT organizations will need to keep abreast of the development and use of LLMs, both for natural language understanding/processing and for creative processes such as conversational interfaces, especially if the organization chooses to build its own model.
- Many AI-based applications and managed services will make use of these LLMs within the enterprise, potentially affecting everything from conversational AI applications and machine translation to creative endeavors such as product design and development. IT needs to be aware of this and work to guide and advise business management.

Guidance

- Identify and learn where LLMs may play a major role across your organization in the future. Consider both what is needed to build and operate AI technology solutions using LLMs and what resources within the organization will be needed to successfully use and work with LLMs.
- Seek guidance and support from your AI and cloud partners and enterprise vendors to identify and learn which applications may make use of LLMs and conversation as an interface. Understand what limitations or considerations need to be taken when using this type of software.

- Develop strategies and plans for the use of responsible AI where these LLMs and conversation as an interface will be used. Ensure that appropriate data use policies are adopted and followed relative to the use of this type of software.

Prediction 4: After a Temporary Focus on GenAI, 60% of Enterprises Will Refocus on Mandating Automation Strategies Based on Outcomes Rather Than Specific Technologies by 2025

As enterprises mature their understanding of automation, project sponsors have shifted from a technology focus to an outcomes mindset where they require tangible proof of value delivered for their investments measured by KPIs aligned with business and financial outcomes. IDC's March 2023 *Worldwide Automation and Process Improvement Survey* showed that 71% of business and technology decision makers view this as a top priority and concern, and 10% said the inability to link investments with outcomes caused them to pull back from making automation investments.

This goal remains illusive. In late 2022, automation centers of excellence (COEs) and their vendors struggled with determining how to align their projects with business KPIs and financial metrics. Then GenAI exploded into the market, bubbling up to become an intense area of focus. Vendors rapidly refocused to embrace GenAI to simplify development, create new automation use cases, and improve existing capabilities. And these efforts aligned with enterprises need to experiment and build competencies to learn how GenAI could enhance their organization, including their automation and process improvement efforts.

GenAI became the magic bullet that enabled both enterprises and their vendors to revert to paying attention to technology instead of outcomes. But the core issue of value delivery remains. By 2025, business sponsors will understand where GenAI fits in and will be back with the same outcomes-based questions that will also extend to GenAI use cases used to automate and improve business processes.

Associated Drivers

- **AI everywhere —** Generative AI takes the spotlight
- **The drive to automate —** Maximizing efficiency and new opportunities

IT Impact

- Centers of excellence that have created skills in value engineering — that include the ability to align metrics with outcomes — will be in a better position than their counterparts focused on technology-led projects, even for AI-powered automation.
- COEs with the aforementioned skills will have new roles in their team and become more collaborative and consultative with their business counterparts.

Business Impact

- Groups trying to improve their business processes may improve their competency in value engineering because they are not being supported by the COE.
- Business sponsors are likely to be more skeptical about the benefits of automation and become more reluctant to sponsor projects when there is limited proof of value, while the opposite is true when the COE upskills its value engineering skills and capabilities.
- Business sponsors look for improvements using substitute technologies — for example, finding packaged solutions and applications.

Guidance

- Ensure that enterprise COEs work with their vendors to identify efforts to rebalance the metrics provided in their technology platforms to become more useful to business.
- Collaborate with professional services partners building value engineering competencies.

Prediction 5: By 2025, the G2000 Will Allocate Over 40% of Core IT Spend to AI Initiatives, Leading to a Double-Digit Increase in the Rate of Product and Process Innovations

Artificial intelligence has, especially over the past two to three years, heralded game-changing innovation in both product development and operational processes, driving a fundamental shift in how enterprises function, meet customer needs, and bolster internal productivity. In the area of product innovation and advancement, AI empowers organizations to create and deliver exceptionally customized and relevant offerings. AI has the capacity to examine customer preferences and behaviors, empowering businesses to customize their products to cater to individual requirements. For instance, in the realm of ecommerce, AI-driven recommendation systems are employed by platforms to propose products to customers, resulting in an improved user experience and bolstered sales.

In addition, AI has effectively demonstrated value in driving process innovation within enterprises across industry sectors. By automating repetitive and rule-based tasks, AI-powered robotic process automation (RPA) optimizes workflows, reduces errors, and enhances operational efficiency. For instance, within the financial services sector, AI-driven chatbots can efficiently handle routine customer queries, allowing human agents to concentrate on more intricate responsibilities. Furthermore, AI facilitates predictive maintenance in manufacturing, enabling companies to identify and rectify equipment issues proactively, thereby preventing costly downtime.

It would not be an understatement to mention that the impact of AI on innovation extends beyond mere process improvement; it leads to the creation of entirely new business models and has the potential to disrupt several industry verticals. Both start-ups and established enterprises are leveraging AI to develop innovative solutions that were previously inconceivable. In summary, AI's influence on both product and process innovation is profound, promoting ingenuity, efficiency, and the emergence of entirely novel possibilities within the business landscape.

Associated Drivers

- **AI everywhere** — Generative AI takes the spotlight
- **The digital business imperative** — Competitiveness and outcomes
- **The drive to automate** — Maximizing efficiency and new opportunities

IT Impact

- IT teams would need to have enhanced focus on data management, especially activities around data integration, quality, and governance. AI model bias and explainability are key areas of concern for IT and data science professionals.
- Enterprises would require ample AI/ML-related skills development. An acute shortage of skilled AI/ML professionals could potentially prove to be a bottleneck for scaling of AI initiatives in the future.
- As AI systems become more sophisticated, so do cybersecurity threats. IT teams need to bolster security measures to protect AI models and data from potential breaches.

- IT departments are tasked with overseeing expenses related to AI, encompassing hardware, software, and cloud services, and the continuous upkeep of these resources. Skillful management of these costs is imperative to prevent exceeding budget limits.

Guidance

- Put in place strong data governance procedures to guarantee the quality, security, and adherence to regulations of your data. Enforce measures such as controlling data access, encrypting data, and tracking data lineage to safeguard sensitive information.
- Regularly assess and mitigate risks associated with AI projects, including technical, operational, and ethical risks. Implement contingency plans for potential issues.
- Establish ethical guidelines for AI usage within the organization to ensure responsible and unbiased AI development and deployment.
- Invest in training and upskilling IT team members in AI, machine learning, and data science. This will empower them to understand AI technologies, make informed decisions, and effectively manage AI-related projects.

Prediction 6: By 2027, the Overall Economic Impact of AI Will Neutralize, as Organizations Overcome Initial Disruption and Refocus Resources on Innovation and New Business Opportunities, Fueling Economic Expansion

AI disruption is upon us and will have impacts on the way most work is done today, much in the same way computers and then the internet did. In the shorter term, this has the potential for negative economic consequences, but more because of the current state of the global economy than AI itself. The capabilities of AI are becoming known during a period of rapid inflation and weakened demand, causing organizations to search for efficiency gains. But when economic conditions improve, companies will also use AI to take advantage of business opportunities that will eventually exceed any short-term losses.

AI will boost the reach and capabilities of platforms and services. It can lead to the creation of new business models, enhance user experience on ecommerce platforms, optimize digital advertising, personalize content delivery, and improve data analytics, among other benefits. Each of these enhancements can drive increased online activity and commerce, leading to increasingly positive economic impacts that will have neutralized by 2028 at the latest and creating trillions of dollars in economic growth beyond.

To better capitalize on the coming economic opportunities and innovation, most organizations will need to do some foundational work. Companies must make specific investments in skills to ensure employees have the requisite knowledge to harness AI capabilities in business operations and innovations. In addition, organizations must be able to harness the power of their data using connectivity-based automation such as integrations and APIs data technologies. Data is oxygen to AI services, and without accurate and up-to-date data, AI will not have situational awareness to serve customers and employees. Organizations need to have an AI transformation strategy to guide investments, measure results, and build momentum.

Associated Drivers

- **AI everywhere —** Generative AI takes the spotlight
- **The drive to automate —** Maximizing efficiency and new opportunities
- **The digital business imperative —** Competitiveness and outcomes

IT Impact

- There will be increasing pressure on IT to develop and deepen AI skills, including prompt engineering and fine-tuning, to ensure AI models can evolve with the business.
- As AI technologies are incorporated into the technology stack, connecting AI generative and predictive AI services such as chatbots will need to be connected to transactional systems to have accurate and up-to-date data when interacting with stakeholders.
- There will be pressure on IT to govern and secure AI, but those are only the hygiene needs. IT will need to steer as well as manage the overall AI strategy.

Guidance

- Develop broad AI skills across interpretative, predictive, and generative AI.
- Identify underlying data and supporting technologies to improve AI capabilities faster and more iteratively.
- Build an AI transformation strategy with the overall business strategy in mind that will encompass many of the existing uses of AI in the organization already.

Prediction 7: By 2025, Two-Thirds of Businesses Will Leverage a Combination of Generative AI and RAG to Power Domain-Specific, Self-Service Knowledge Discovery, Improving Decision Efficacy by 50%

While large language models such as BERT and GPT-3 have been part of many search and knowledge discovery systems for several years, in 2022 and 2023, the latest advances in generative AI have shown the art of the possible with public models like ChatGPT. This, in turn, has prompted a surge of demand for capabilities such as natural language question answering and conversational search to support self-service knowledge discovery for business users, customers, and partners. Thanks to their ability to work with unstructured data, advances in LLMs also represent a new paradigm to advance document AI capabilities by addressing some of the challenges of working with single or groups of documents and can allow organizations to unlock a broad range of use cases that leverage image and video data, including immersive search, building and enforcing document creation standards, and similarity scoring.

For external-facing use cases such as product search, harnessing generative AI-enhanced discovery technologies optimizes both CX and tangible business outcomes. In B2B, it fast-tracks deal closures by precisely meeting large client needs. For B2C, personalized journeys lead to higher cart values and quicker conversions. Examples include virtual shopping assistants, visual search enhancement, and customer service enhancement.

However, as organizations move to take advantage of the latest advances in LLMs and generative AI, they must contend with risks such as "hallucinations," in which LLMs generate plausible sounding but ultimately inaccurate statements. For use cases in self-service product discovery, knowledge discovery, and question answering, many vendors are bounding and grounding generative models in real-world/enterprise data by feeding generative LLMs with results from search systems via techniques such as Retrieval-Augmented Generation (RAG). Similarly, many conversational AI vendors are connecting with back-end search engines and vector databases to support RAG for question answering rather than relying directly on LLM's pretrained knowledge. This combination of search technologies with LLMs/generative AI is already helping provide accurate, relevant, and enterprise-ready capabilities.

Associated Drivers

- **AI everywhere —** Generative AI takes the spotlight
- **The digital business imperative —** Competitiveness and outcomes
- **Everything as a service intensifies —** Transforming models to drive change

IT Impact

- IT teams will be expected to evaluate the ability of search and knowledge discovery tools to provide GenAI functionality with appropriate security and governance guardrails.
- IT teams will need to support and educate business end users that are not skilled (or interested) in self-service knowledge discovery tools to encourage adoption and achieve business benefits such as improved decision making.
- For use cases such as enterprise search and competitive intelligence, these tools also need to be integrated with key data sources. For ecommerce use cases, IT will need to bridge the information divide between presale research and post-sale support. Domain-specific specializations, including integrations, should be considered.

Guidance

- Engage professional services partners for advice in selecting the right tools for your business use case, IT architecture, governance policies, and risk tolerance.
- Design search experiences with the business end user in mind, working backward from the point of decision-making or action-taking, to make tools easier to adopt and mitigate change management issues. Engage the business early in the solution design process.
- Ensure integration across systems to facilitate AI-driven tools to generate insights throughout the entire sales life cycle. Consider whether domain-specific accelerators offered by services firms, such as pretrained industry- or function-specific models, could provide your organization with a running start to implementing an AI-enabled search solution.

Prediction 8: By 2024, 33% of G2000 Companies Will Exploit Innovative Business Models to Double Their Monetization Potential of Generative AI

Generative AI is a rapidly growing field with the potential to revolutionize many industries. According to IDC's August 2023 *Generative AI ARC Survey* (n = 1,364), 19% of the survey respondents worldwide have reported that GenAI has already disrupted their business operating model, and another 18% expect significant impact over the next 18 months.

While technology is a source of advantage, it is the business model that will help businesses monetize generative AI and drive the most lasting competitive advantage. To level set, business models are the systems of linked activities through which a company creates, delivers, and captures value. It's important to think beyond the core value proposition and customer segments as value and advantage in a business model come from different areas. Successful monetization will involve broader considerations such as defining the play in the value chain, cost model needed to ensure attractive returns, forming strategic partnerships, prioritizing privacy and security, and emphasizing customer relationships and outcomes.

Generative AI can be used to generate revenue from data: for example, businesses can use synthetic data generated by GenAI to improve the accuracy and performance of machine learning models, which can lead to increased revenue for businesses, and they can also sell synthetic data to third-party companies. Generative AI can aggregate market data to test concepts, ideas, and models. Stitch Fix, which uses algorithms to suggest style choices to its customers, has experimented with DALL·E to visualize products based on customer preferences regarding color, fabric, and style.

Monetizing generative AI extends beyond pricing considerations. The position of an organization in the ecosystem can guide the commercial partnerships it should pursue (e.g., with relevant platforms) such as document management system to access a broader market and enhance the value proposition of its generative AI solutions. There is also a big opportunity to enhance customer value management by hyper-personalization, which uses real-time data to provide more customized and contextually relevant experiences. Mondelez India tapped the Rephrase.ai platform (text to video powered by GenAI offering) to record an avatar of an Indian actor, and then used it to create personalized ads in local stores across India.

Associated Drivers

- **AI everywhere —** Generative AI takes the spotlight
- **The digital business imperative —** Competitiveness and outcomes
- **Everything as a service intensifies —** Transforming models to drive change

IT Impact

- There will be increasing pressure on IT to ensure security, privacy, and control across the tech stack from infrastructure to model to application and end-user interface for both internal and external generative AI solutions.
- IT will need to play a critical role in evaluating third-party solutions that the business may benefit from integrating with versus building its own.
- IT will need to reimagine its core functions ranging from software development to technical debt and IT operations to support the rapid rate of innovations powered by GenAI.

Guidance

- Focus on customer outcomes. Create helpful resources, such as buyer's guides, tutorials, and user testimonials, that can assist customers in understanding the benefits and value of the solution, thus ensuring the "stickiness" of the product.
- Evaluate the choice between embedding generative AI or offering separate add-ons subject to the nature of the use case and customer preferences.
- Strike the right balance between pricing, integrating generative AI, and addressing broader commercial considerations like partnerships, cost allocation, channel strategy, privacy, and security.

Prediction 9: By 2028, 30% of G2000 Companies Will Experiment with AGI (Currently Speculative) Systems That Will Have a Transformative Effect on Society and Create Significant Opportunities and Threats

What does the next few years hold for AI? And how close are we to AI that surpasses human intelligence? Artificial general intelligence (AGI) refers to a much broader form of machine intelligence than what we have today, which is very specialized and limited to a particular domain. While we await a single, formally recognized definition of AGI, OpenAI refers to it as "highly autonomous systems that outperform humans at most economically valuable work." Alternate definitions include "systems that demonstrate broad capabilities of intelligence including reasoning, planning, and the ability to learn from experience, and with these capabilities at or above human level."

A group of industry AI researchers recently made a splash when they published an academic paper titled "Sparks of Artificial General Intelligence: Early experiments with GPT-4." The researchers noted that GPT-4 can solve novel and difficult tasks that span mathematics, coding, vision, law, psychology, and more, without needing any special prompting, exhibiting "strikingly close to human-level performance." They concluded that GPT-4 "could reasonably be viewed as an early (yet still incomplete) version" of AGI.

A key subtlety on "how close is AGI?" is that intelligence exists on continuum and therefore assessing whether a system displays AGI will require considering a continuum. Multiple groups are working toward AGI and as it progresses, AGI will be transformative, impacting everything from the labor market to how we understand concepts like intelligence and creativity. AGI can also broaden access to services that previously were accessible to only privileged.

Associated Drivers

- **AI everywhere —** Generative AI takes the spotlight
- **Dynamic work and skills requirements —** New work mode era
- **The drive to automate —** Maximizing efficiency and new opportunities

IT Impact

- Recent advances in AI have made AI a C-suite and board-level priority. This has already created additional pressure on the IT organizations to balance the risks and value of these systems. IT will be compelled to double down collaboration with the different business units to identify and prioritize the strategic use cases.
- IT will have to take the overarching leadership to manage business reputation, mitigate risks to intellectual property, and insist that providers of AGI systems maintain transparency.
- IT will have to collaborate with business to set up data access controls across the value chain and establish rigor for evaluation metrics.

Guidance

- Identify the use cases that has the highest level of strategic advantage to experiment with AGI systems, as they are going to be complicated and expensive. Build the structure to be agile so that you can experiment and fail fast.
- Adopt proactive change management to ensure readiness from people and processes aspects to embrace AGI systems.
- Ensure use of the right tools and technologies that help minimize biases and misuse while adhering to your corporate code of conduct and associated regulations.

Prediction 10: By 2027, the Spending on Server Accelerators (GPU, FPGA, and AI ASIC and ASSP) Compared with Server CPUs Will Invert, Achieving a 55/45 Rate

Specialized AI-enhanced workloads running on datacenter infrastructure pull against the general-purpose nature of traditional server microprocessors. How far should an end user go relying on a general-purpose server microprocessor before offloading AI tasks to a specialized accelerator like a server graphics processing unit (GPU), field-programmable gate array (FPGA), or AI ASIC (custom) or AI ASSP (merchant)? AI-enhanced workloads typically require a hardware and software stack tuned to its performance needs, including mixing and matching of the right silicon.

Cloud service providers and hyperscalers have the resources and need to optimize their AI computing infrastructure for the diversifying workloads of their end users. They will continue to invest in direct purchases of merchant processors and accelerators from vendors and will invest in their own in-house custom processors and accelerators.

Like cloud service providers, communications service providers are also buying custom processors and accelerators directly from their vendors. Their move from proprietary ASICs to standard (mostly x86) server processors is shifting an entire hardware and software ecosystem and giving communications service providers new leverage in how processors are designed and sold.

For the foreseeable future, AI workloads will drive increasing combinations of server processors and accelerators, including GPU, FPGA, ASIC, and ASSP. Until AI workloads that require the offload of tasks from server processors to accelerators standardize on algorithms and software stacks tuned to server processors, purchase of accelerators will eat into purchase of server processors. By 2027, the collective revenue among server GPUs, FPGAs, and AI ASICs and ASSPs will reach $37.0 billion compared with total server CPU revenue of $32.3 billion.

Associated Drivers

- **AI everywhere —** Generative AI takes the spotlight
- **The digital business imperative —** Competitiveness and outcomes
- **Everything as a service intensifies —** Transforming models to drive change

IT Impact

- No single technology solution stack can serve all workloads. Instead, AI requires diverse, heterogeneous solution stacks, including among the choices of data processing silicon used in datacenter infrastructure.
- For at least the next five years, end users will explore the price, performance, and power benefits of offloading functions from server microprocessors to accelerators such as GPUs, FPGAs, and AI ASICs and ASSPs. Processors and accelerators purchased directly by cloud service providers and hyperscalers influence semiconductor design and pricing.
- The increasing share of processors procured by cloud service providers and hyperscalers reduces the influence of traditional server system OEMs and distributors that buy in smaller volumes.

Guidance

- Exercise due diligence in assessing the need for AI accelerators as there isn't enough computing capability today to narrow the gap between the volume and sophistication of data being created and the transformation required to make the data into useful and valuable insights for organizations.
- Model the mix of silicon you select according to where that silicon will be used and the associated movement, storage, and processing of data.
- Prioritize AI training on core datacenter infrastructure and latency-sensitive AI inferencing at the edge.

ADVICE FOR TECHNOLOGY BUYERS

To maximize the realization of artificial intelligence and automation initiatives, technology leaders should consider the following:

- **AI strategy:** As AI adoption expands across multiple businesses and IT processes, look for these four broad considerations in your AI strategy:
  - **Scale:** Systems designed for business as usual can be overwhelmed. Look for a flexible AI platform that can quickly scale up and down based on current demand. For example, your conversational AI platform must withstand extreme spikes in call or messaging volume.
  - **Standardization:** Instead of implementing distinct solutions throughout the organization to handle small tasks, embrace the platform approach to support consistent experiences. This will help not only ease the oversight and maintenance, leading to reduced maintenance costs, but also generate significant value by collecting a wider swath of organizational data in one place. Such a system can apply analysis to that data, providing insights that reveal ways to make the entire organization more efficient and effective. A holistic approach has the potential to grow with your business, paying dividends over time.
  - **Compatibility and integrations:** Many AI platforms (conversational, computer vision, or advanced machine learning) are not designed to integrate seamlessly with in-house or third-party applications, tools, and technologies. You'll want a solution that is easy to plug into other systems if you want to maximize your value – not just for the use cases you have today but also for the ones you haven't even thought of yet.
  - **Compliance and data privacy:** Governments and industry oversight committees impose continuously shifting regulations on businesses, and for good reasons. When interacting with customers and their data, your organization will need to think about all the potential points of regulatory exposure involved. In addition to compliance, ensure that you understand how the solution is used and where it houses customer data, as this can also present potential issues given regulations such as GDPR and other regional data privacy laws.
- **Build, buy, partner, acquire:** Assess build versus buy and partner versus acquire decision within the context of different categories of AI software. Consider the business use case value, market understanding (your and technology suppliers' market), data and competitive differentiation, availability of talent, time to market, and cost in your decision and investment strategy.
- **Low code/no code:** Embrace low-code/no-code solutions to accelerate time to market as well as support better collaboration between data scientists and business analysts. Make sure that the solution is flexible, is customizable for a breadth of use cases, and can be well integrated into the organizations' overall AI governance and risk management process. Further:
  - **Generative AI:** To maximize ROI and accelerate time to value, create an environment of experimentation for the right/prioritized use cases. Leverage an iterative approach to prioritize the right set of use cases. Develop policies around responsible use of generative AI and inhibit nefarious scenarios. Partner with trusted technology/solution suppliers and service providers, prioritizing those that strive to provide transparency on training data sets and appropriate model usage, and/or offer their models in open source. Prepare for fine-tuning and prompt-tuning skill sets through either hiring, reskilling, or professional services support.
- **Embedded AI applications/solutions:** Large technology suppliers are integrating AI as an intelligence layer across their entire tech stack. This is hugely important to help businesses

accelerate time to value and address talent gap. However, it is critical that the embedded AI applications/solutions are technically robust and accurate to ensure that AI systems are reliable as intended in the design. The conformity assessment for AI applications should include procedures for testing, inspection, and certification of the algorithms and of the data sets used in the development phase. Businesses should also have easy mechanisms to customize them for their needs and support the organizations' responsible AI requirements.

- **Data foundation:** Select a responsible AI/ML platform with support for all data types. Ensure the platform supports collaboration. Embrace an intelligent data fabric that helps automate and enforce universal data and usage policies across hybrid data and cloud ecosystems; automate how data is discovered, cataloged, and enriched for users; and automate how you access, update, and unify data spread across distributed data and cloud landscapes without the need of doing any data movement or replication. In detail:
  - **Enterprise automation:** Automation and digital transformation (DX) are now dynamic duos. Enterprise automation is expanding rapidly in terms of use cases, technologies, and benefits. Consequently, the challenges to managing all this automation are growing exponentially, such as security risks, scalability, and manageability. Enterprise automation requires a robust governance framework to address these challenges and democratize automation rapidly. This can be done through an automation COE to centralize governance and principles but decentralize delivery of automation. Enterprise automation also requires a strong integration strategy to bring disparate data from applications and processes together. For transformation of customer and employee experiences, embrace an AI-powered automation strategy.
  - **Governance and risk management:** The adoption of AI has profound implications for everyone engaged in business management. The current and proposed legal frameworks regulating AI are rapidly evolving. For global businesses, the task of ensuring that their AI technology complies with applicable regulations will be complicated by the differing standards that are emerging from China, the European Union, and the United States. It is essential to build trust with users and other stakeholders so that such technologies are operated responsibly and with appropriate controls in place. Further:
    - **Establish clear policies** with respect to data privacy, decision rights, accountability, and transparency.
    - **Have proactive and ongoing risk management and enterprise data governance** practices performed jointly by IT and those in business and compliance functions. Establish clear escalation and governance processes and offer recourse if customers are unsatisfied.
    - **Establish holistic metrics.** Do not just goal yourself on revenue but think about the social outcomes. Try hard to measure the leading indicators of social outcomes.
    - **Have diversity** in your AI data teams. Work to have a representation of the population of where you would be deploying the algorithms. Avoid the blinders of the homogeneous teams. Diverse teams will work to have diverse training data, more thoughtful feature sets, and less bias in the data.
    - **Charter data teams** to be the "conscience" officers.
- **Talent:** Your organization may not have all the expert data scientists needed to develop new AI algorithms, or you may have needs for general-purpose deep learning capabilities for image or audio analysis. Consider third-party services for these needs while focusing your internal efforts on more proprietary requirements. Establish a portfolio of expertise that extends beyond data science. AI is largely about data and algorithms. However, "math" alone does not

ensure a successful implementation of new AI-based technology solutions or desired outcomes. Enterprises need to consider the role of data scientists along with industry expertise.

EXTERNAL DRIVERS: DETAIL

AI Everywhere — Generative AI Takes the Spotlight

- **Description:** With intelligence becoming the primary source of value creation, we are on the verge of the "Intelligence Revolution," in which artificial intelligence (AI) and automation-oriented technology will be the main accelerators of business change. In the realm of "AI everywhere," generative AI (GenAI) emerges as a transformative force, potentially revolutionizing the future. This branch of artificial intelligence enables a machine-driven autonomous creation of new content, from images to music to even written text, with remarkable accuracy. Early applications of GenAI have showcased its potential in fields such as creative arts, content and code generation, and personalized recommendations. However, it also raises concerns regarding bias and privacy: AI algorithms can inadvertently perpetuate biases and pose threats to personal data. As a result, regulation becomes crucial to ensure responsible and ethical use of GenAI. Despite these challenges, the possibilities are vast, ranging from improved customer experiences to innovative problem solving. Harnessing the power of GenAI and navigating its associated complexities have the potential to shape the future of industries and drive advancements in the AI-driven world.
- **Context:** Businesses are already jumping to get a piece of the AI pie, afraid to miss out on the opportunities it presents. Although we are in the early days, monetization and scale of AI solutions are expected to evolve rapidly. However, this comes during a time of relative economic uncertainty and increasingly constrained IT budgets. Furthermore, AI is not without risks, especially when it comes to ethical AI and data privacy, and companies need to carefully consider the best use cases in order to implement AI effectively.

The Drive to Automate — Maximizing Efficiency and New Opportunities

- **Description:** Broader automation use cases – beyond just generative AI – are now ubiquitous. Now that data is embedded in the core of strategic capability for every organization, automation is critical to scaling a digital business and is evident in three domains: IT automation, process automation, and value stream automation – leading to autonomous operations, digital value engineering, and innovation velocity. Industrial organizations have spent the past few years evolving toward the Fourth Industrial Revolution (Industry 4.0) through the use of industrial automation and intelligence. Thoughtful implementation is more important than ever as data becomes embedded in the strategic core of every organization. Automation technologies such as robots and drones are being used increasingly in the military and healthcare sectors. Given this boost in automation, data is increasingly precious, and privacy must be prioritized and security enhanced. In some cases, automation has also led to concerns over the future of work – whether it will enhance or take away.
- **Context:** Businesses are rethinking how to employ automation to maximize operational efficiency – from automating assembly in manufacturing to identifying opportunities for food waste reduction in hospitality to improving customer experience (CX) in digital banking. IT will need to continue to assess new technologies and approach automation investments strategically, both within the walls of the organization and in the field. Among industrial organizations, IT/OT convergence will necessitate shared responsibility across teams for automation priorities and implementations.

Geopolitical Turbulence — Sovereignty Under Emerging Fragmentation

- **Description:** The geopolitical landscape in today's world is continuously evolving. The Russia-Ukraine War accelerated this division in 2022, with many companies finding themselves in a situation where they were forced to side along geopolitical lines according to customer and government demands. Other conflicts continue to brew, with the United States–China tensions/conflict in Taiwan being perhaps the most closely watched. Domestic-level conflicts are also brewing that can have an impact on the IT sector. In several countries around the world, there are future upcoming elections that can have vast implications on businesses, such as their effect on populist protectionism and implementation of ESG, for instance. In addition, the energy crisis in Europe was sparked by the Russia-Ukraine War, but unstable costs of energy remain a challenge to various countries around the world and will only increase in importance in light of these geopolitical challenges as well as factors such as climate change. Today's geopolitical turbulence therefore has the potential to further advance the emergence of an ever-increasing multipolar order.
- **Context:** The uncertain geopolitical landscape is multiplying risks for businesses as they navigate trade fragmentation and a plethora of new regulations and deal with factors, such as populist protectionism, that reduce their ability to access markets. These conflicts and trade tension lead to supply chain disruptions, restricted imports, and increased cybersecurity risk — and therefore increased costs for IT companies reliant on global manufacturing and distribution networks for essential components. Given this, the IT sector should aspire for diversification of supply chains, bolstered cybersecurity measures, and emerging market solutions.

The Digital Business Imperative — Competitiveness and Outcomes

- **Description:** A digital business sees value creation based on the use of digital technologies for both internal and external processes, including customer engagement, employee experience, and product and services development. Building and leading a digital business is imperative for organizations to be competitive. While certain operational aspects may always have a nondigital component, digital businesses prioritize a digital-first strategy that aligns all parts of the business and IT landscape with digital workflows to drive value and growth. The development strategies for both digital and nondigital assets now require leveraging multiple channels for the digital business to obtain support or funding. This places a strong emphasis on providing digital experiences for customers and citizens, employees, and partners and necessitates a shift toward fully digital operating models and resilient supply structures enabled by digital technology. The focus of a digital business is increasingly on delivering measurable outcomes. Businesses that have recognized the value of digital anticipate maintaining or even increasing their investment in technology, even in times of economic uncertainty.
- **Context:** As more and more enterprises embrace digital strategies and technology, they prioritize technology investments that drive innovation or allow for competitive differentiation. Technology is no longer viewed as a tool to keep the business running, but it is the foundation for building new revenue-generating experiences and products. Laggards will need to adapt quickly and develop their digital road maps and embrace a digital business platform. Identifying top digital revenue opportunities that deliver value will be crucial for overall business success and implementation of organizational digital-first strategies.

Everything as a Service Intensifies — Transforming Models to Drive Change

- **Description:** The concept of "everything as a service" ("XaaS") is driving change across all sectors and ecosystems, affecting both the supply side and the demand side of businesses.

Organizations are adopting as-a-service models at different speeds out of necessity, but the complex delivery strategies make requirements more complicated and the impacts less predictable. The shorter decision cycles enabled by on-demand services allow industry leaders to approach things differently, but the commitment models are fundamentally changing. On the supply side, there is a rising demand and higher customer expectations. As a result, suppliers are compelled to convert and enable their offerings more quickly using a secure services-based model. Buyers are now making decisions based on commitments to measured outcomes in terms of optimization, reliance, and financial models. Architecture and solution strategies are now critical to the service provider, where proprietary systems that are being maintained or migrated can materially impact the efficacy of the as-a-service solution. Leaders face the challenge of finding new financial, operational, and governance models that support a successful transition to an as-a-service approach. Critical factors for organizations to thrive through the as-a-service change landscape include solution control, contractual clarity on roles and responsibility, and accountability alignment including geoeconomic assurance and data.

- **Context:** To deliver optimally, companies are looking to better manage their as-a-service offerings. They're changing product design, delivery, and pricing alongside adjusting management and operations to best optimize as-a-service technology. This is the new model for the tech industry, and suppliers and buyers will need to adapt accordingly. This includes developing cloud-based control to manage provisioning, continuous usage tracking, and architectural efficiency.

Dynamic Work and Skills Requirements — New Work Mode Era

- **Description:** In the wake of COVID-19 pandemic-driven accelerated work transformation, enterprises continue to face dynamic work conditions. These range from lack of skilled employees to codifying more flexible ways of working that rely on a broad range of technologies and services. In some regions, most notably in Asia/Pacific, organizations are focused on building more secure and technically sophisticated office environments. In North America, remote and more flexible work models are driving investments in technologies that support collaboration across and within disparate work environments. Across this spectrum of work models, organizations are investing in infrastructure, hardware, software, and services to enable and manage increasingly automated ways of working. These include automated remote onboarding, learning in the flow of work, and use of AI and generative AI to facilitate basic tasks and workflows. While the pandemic drew much needed attention to the employee experience, enterprises have shifted to aligning employee requirements more plainly to strategic business goals. The key challenge around the globe has been to find or upskill/cross-skill employees to scale and meet the demands of complex, automated work processes. Flexible work models continue to change to become even more agile, with digital workspaces highlighting skills, workforce management, automation, changing demographics, and as-a-service talent resourcing.
- **Context:** New modes of working are now intrinsic to leadership and organizational resilience and go well beyond traditional staff planning methods. They are also having an impact on frontline workers who have historically been neglected in favor of higher-paid front- and back-office peers. New work models require agile cross-functional teams — including HR, IT, LOB, finance, facilities management, and operations — to engage top talent and meet client brand expectations. While headlines debate the fate of environmental, social, and corporate governance (ESG) initiatives, it's clear that environmental concerns will be an embedded element of workplace design and implementation of flexible work models. C-suite leaders and their teams must collaborate to recalibrate work culture, augmentation, and space/place planning to enable more secure, dynamic, and refined work models of the future.

Shifting Tech Regulatory Landscape — Navigating Risk and Opportunity

- **Description:** With frontier technologies like generative AI, geopolitical concerns, and cyber-risks, the tech legal landscape is rapidly changing. While the GDPR in the EU is perhaps the most well-known of privacy laws, other countries have enacted legislation to ensure that personal information is protected and ethically used, such as China's Personal Information Protection Law (PIPL) and Japan's Protection of Personal Information Act (APPI). Nations all over the world are considering frameworks to regulate AI, including the EU's AI Act and the United States' AI Bill of Rights. Cybersecurity is top of mind with the United States' CIRCIA Act, Japan's Basic Act on Cybersecurity, and the EU's Cybersecurity Act. And with ongoing chip wars, countries around the world have mandated domestic production for certain parts of semiconductor manufacturing and banned foreign-created semiconductors in some cases — often along geopolitical lines. Tech regulation, however, is not just a blockade but presents an equal amount of opportunity as well. The aforementioned chip laws also incentivize domestic production and innovative chip manufacturing through tax subsidies. Other strategies such as electric vehicle subsidies are accelerating the green transition across many nations. And larger industry verticals are receiving big boosts, such as Saudi Arabia's investments in healthcare technologies.
- **Context:** Businesses must navigate an increasing number of regulatory rules. Even if it's not always the primary focus, tech is often a crucial part of these regulations. Most of these rules are intended to hedge against risks, but some are entrenched in geopolitical divides, so those firms that stay ahead of the game and build upon resiliency will be best equipped to comply with these regulations. Moreover, regulations and policies are not just restraints — they are also often springboards for investment with many regulations proposing tax subsidies and other kinds of incentives.

LEARN MORE

Related Research

- *Critical External Drivers Shaping Global IT and Business Planning, 2024* (IDC #US51057623, September 2023)
- *Market Analysis Perspective: Worldwide Artificial Intelligence Software, 2023* (IDC #US51206323, September 2023)
- *Market Analysis Perspective: Worldwide Conversational AI Tools and Technologies, 2023* (IDC #US50013623, September 2023)
- *Market Analysis Perspective: Worldwide Computer Vision AI Tools and Technologies, 2023* (IDC #US51226823, September 2023)
- *Market Analysis Perspective: Worldwide Search and Knowledge Discovery, 2023* (IDC #US49696923, September 2023)
- *Market Analysis Perspective: Worldwide Connectivity Automation, 2023* (IDC #US48426622, September 2023)
- *Market Analysis Perspective: Worldwide Enterprise Intelligence Services, 2023* (IDC #US50081723, September 2023)
- *Market Analysis Perspective: Worldwide AI Life-Cycle Software, 2023* (IDC #US50008423, July 2023)

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com