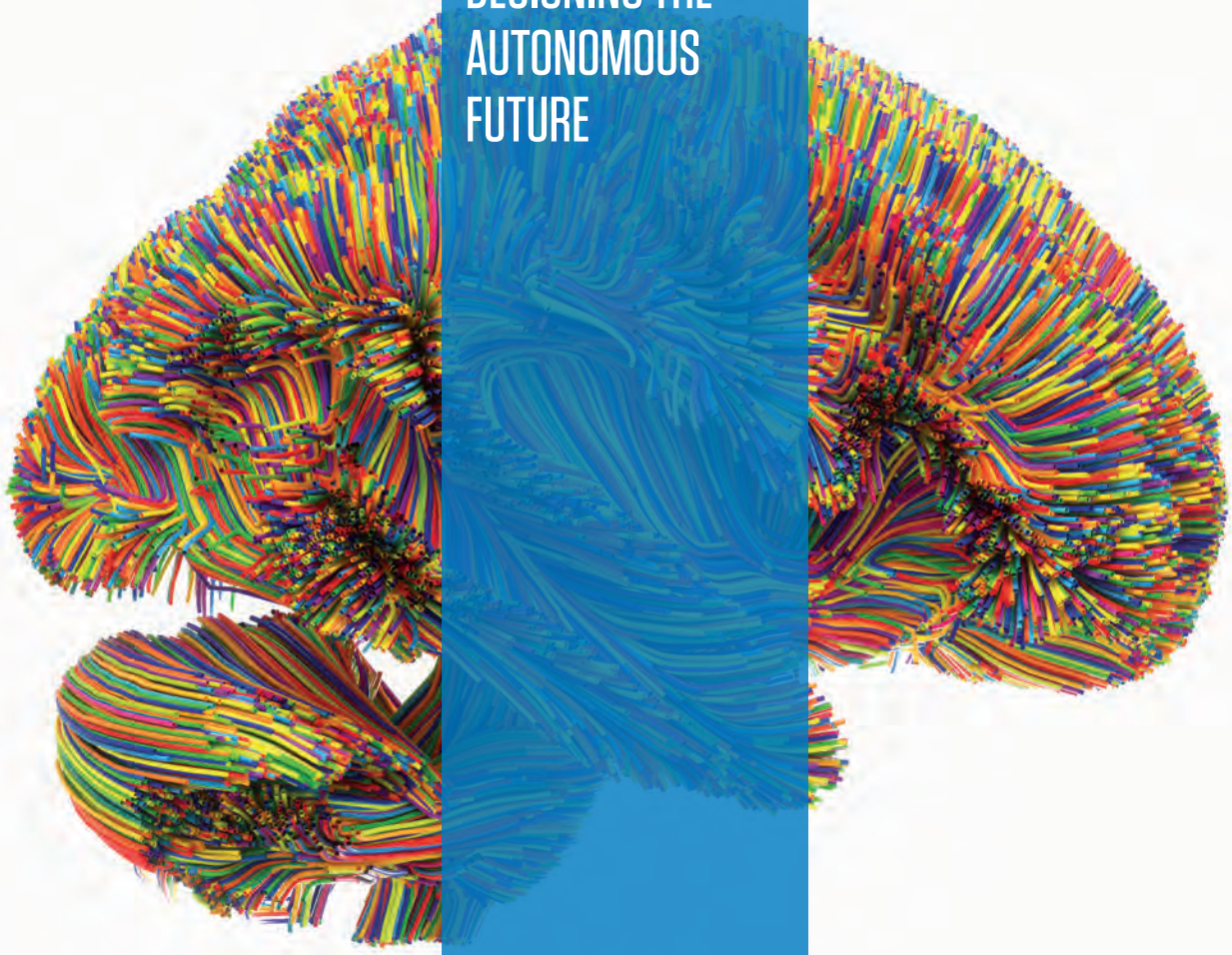
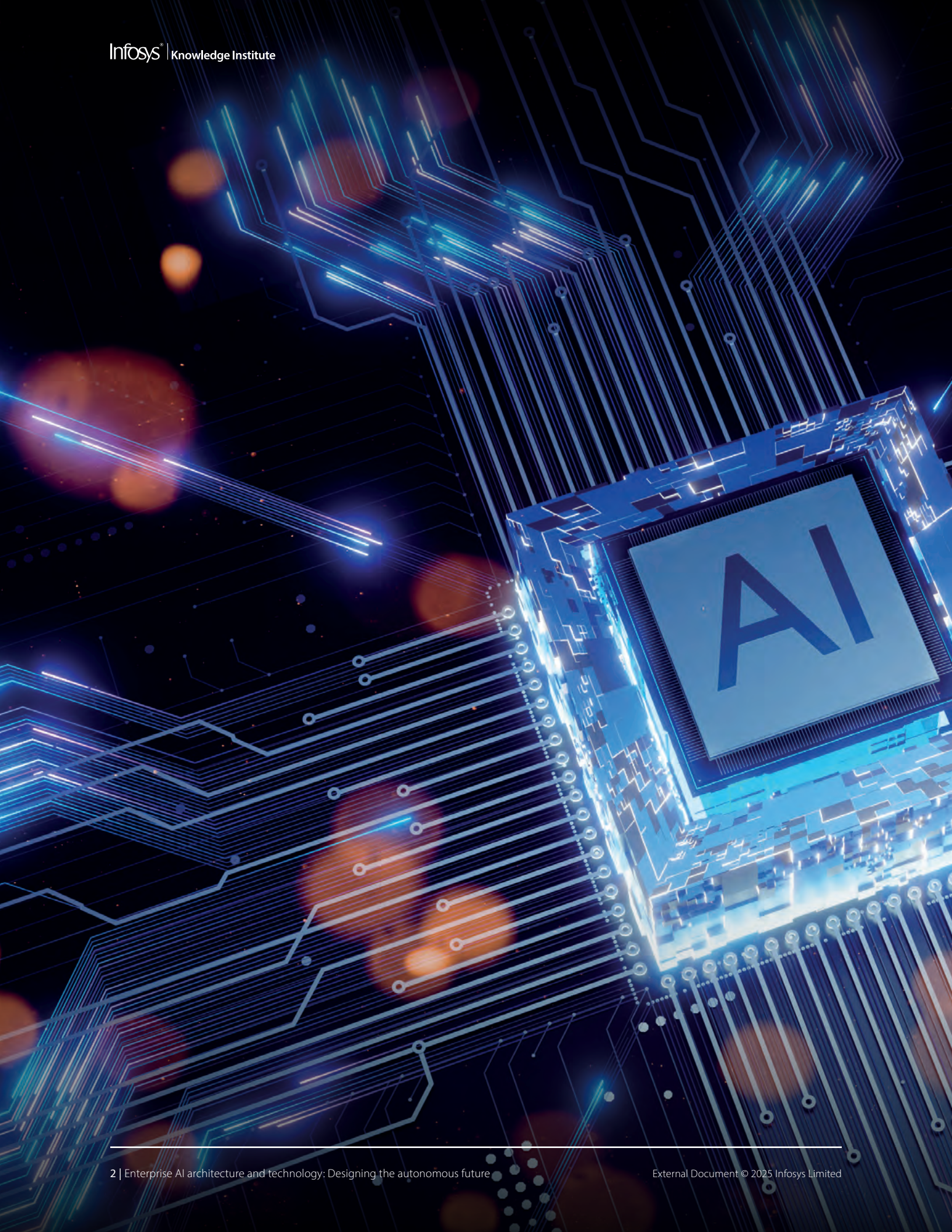
Infosys®
Navigate your next

ENTERPRISE AI ARCHITECTURE AND TECHNOLOGY: DESIGNING THE AUTONOMOUS FUTURE



Infosys® | Knowledge Institute



Contents

Introduction

Artificial intelligence (AI) is accelerating through transformative breakthroughs in agentic systems, multimodal processing, and frontier cognitive architectures — innovations that are reshaping the enterprise landscape as we know it. What began as a promising experiment has now matured into demonstrable business impact. The expanding influence of generative AI is no longer a distant possibility; it's a present force that's empowering organizations to move beyond pilot projects. Businesses are decisively breaking free from isolated AI applications, embedding intelligence systematically across their operational backbone through structured, three-horizon transformation frameworks.

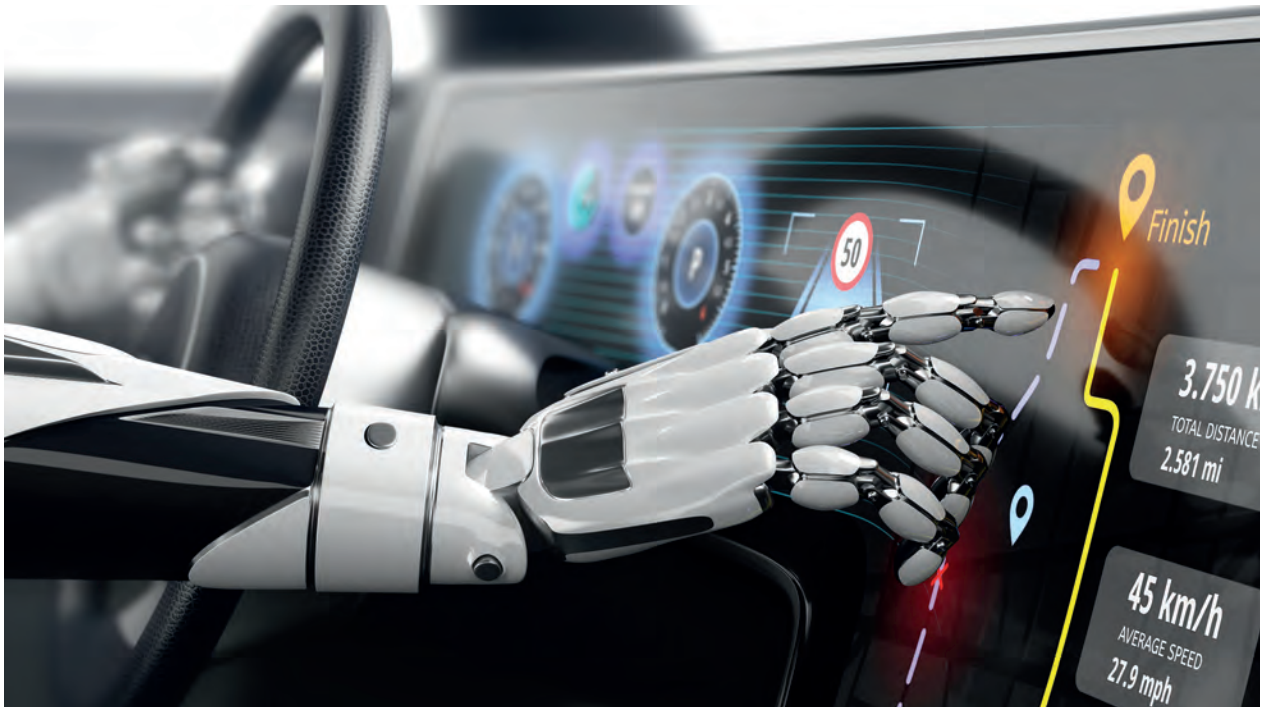
But the future of AI goes beyond mere intelligence. It's contextually adaptive and autonomously collaborative — a force that doesn't just augment human capability but transforms how work is done. Generative AI has evolved from merely blurring the lines between human and machine ingenuity to creating entirely new forms of partnership. Enterprises that integrate AI with comprehensive cloud infrastructures, advanced model engineering, and agentic orchestration are gaining decisive competitive advantages — reshaping industries and unlocking new professional categories in the process.

These AI-first organizations are no longer focused on strategy alone. They deliver operational excellence through robust infrastructures spanning machine learning operations (MLOps), agent operations (agentOps), multimodal understanding, and simulation-based safety testing — all designed to unleash the full potential of autonomous systems.

What sets the most successful AI deployments apart is not just innovation but trust. Secure, contextually aware, and demonstrably reliable, these deployments integrate responsible AI strategies into comprehensive assurance frameworks.

The result? Scalable transformation that amplifies human capabilities, fosters breakthrough innovation, unlocks autonomous efficiencies, accelerates sustainable growth, and cultivates intelligently connected ecosystems — all through a disciplined, horizon-based approach to AI evolution.

The journey from horizon 1 to horizon 3



The evolution of AI is mapped across three interconnected horizons, where each track is assessed for its maturity, adoption, and potential impact.

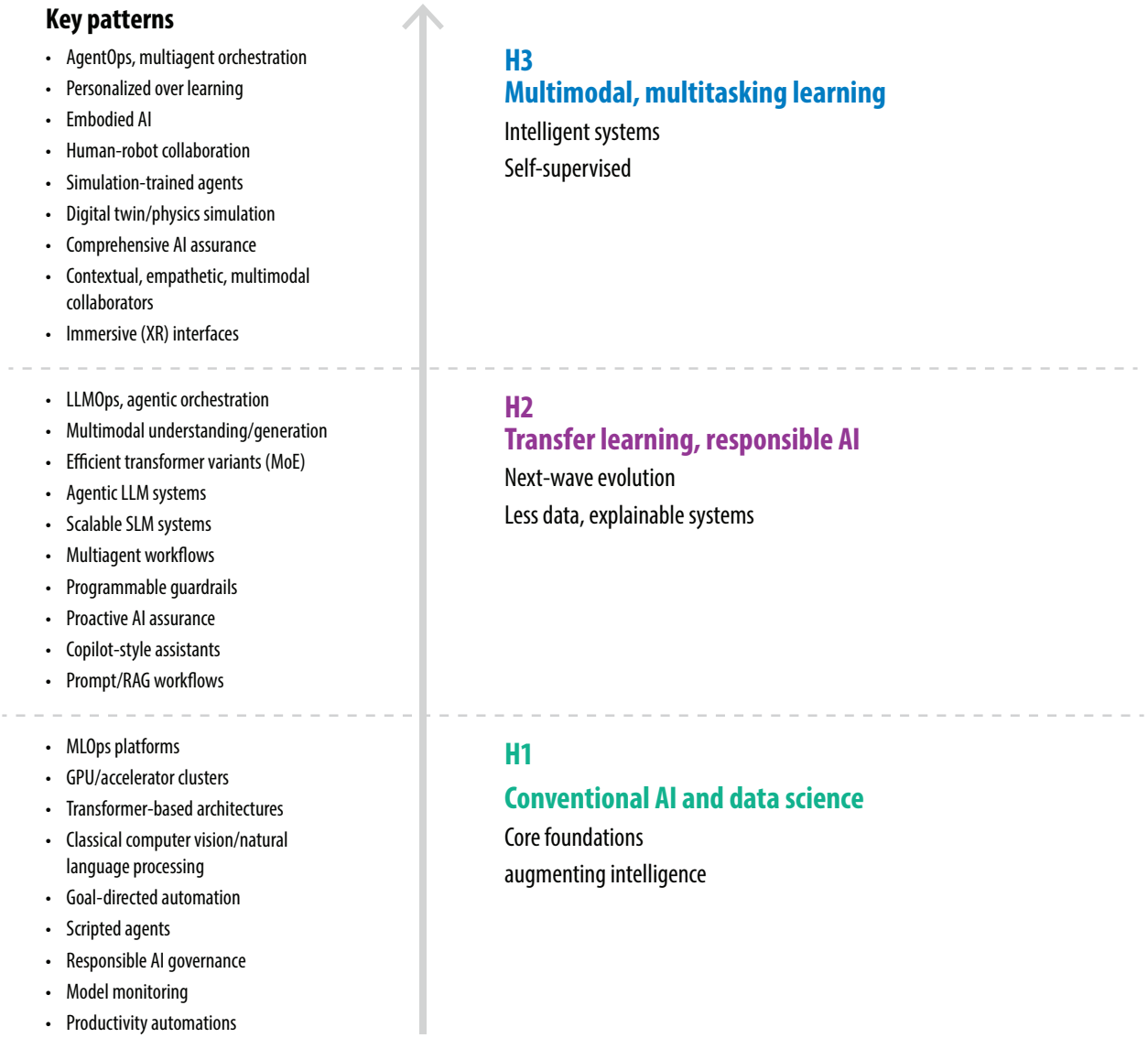
Horizon 1 (H1) laid the groundwork by establishing technical foundations with mature MLOps platforms, production-proven deep learning architectures, and goal-directed automation using classical AI reasoning. During this phase, organizations deployed standardized graphics processing unit (GPU) clusters and Kubernetes infrastructure while embedding AI applications in everyday workflows — powering search, recommendations, and productivity automations across office suites.

Building on these foundations, horizon 2 (H2) marked the present evolution of AI with the introduction of intelligent enterprise blueprints.

These blueprints feature agentic orchestration, multimodal understanding across text, image, audio, and video, and sophisticated large language models (LLMs) that plan, reason, and collaborate as multiagent workflows. This phase emphasizes proactive AI assurance, supported by programmable guardrails and copilot-style assistants that deliver measurable productivity gains.

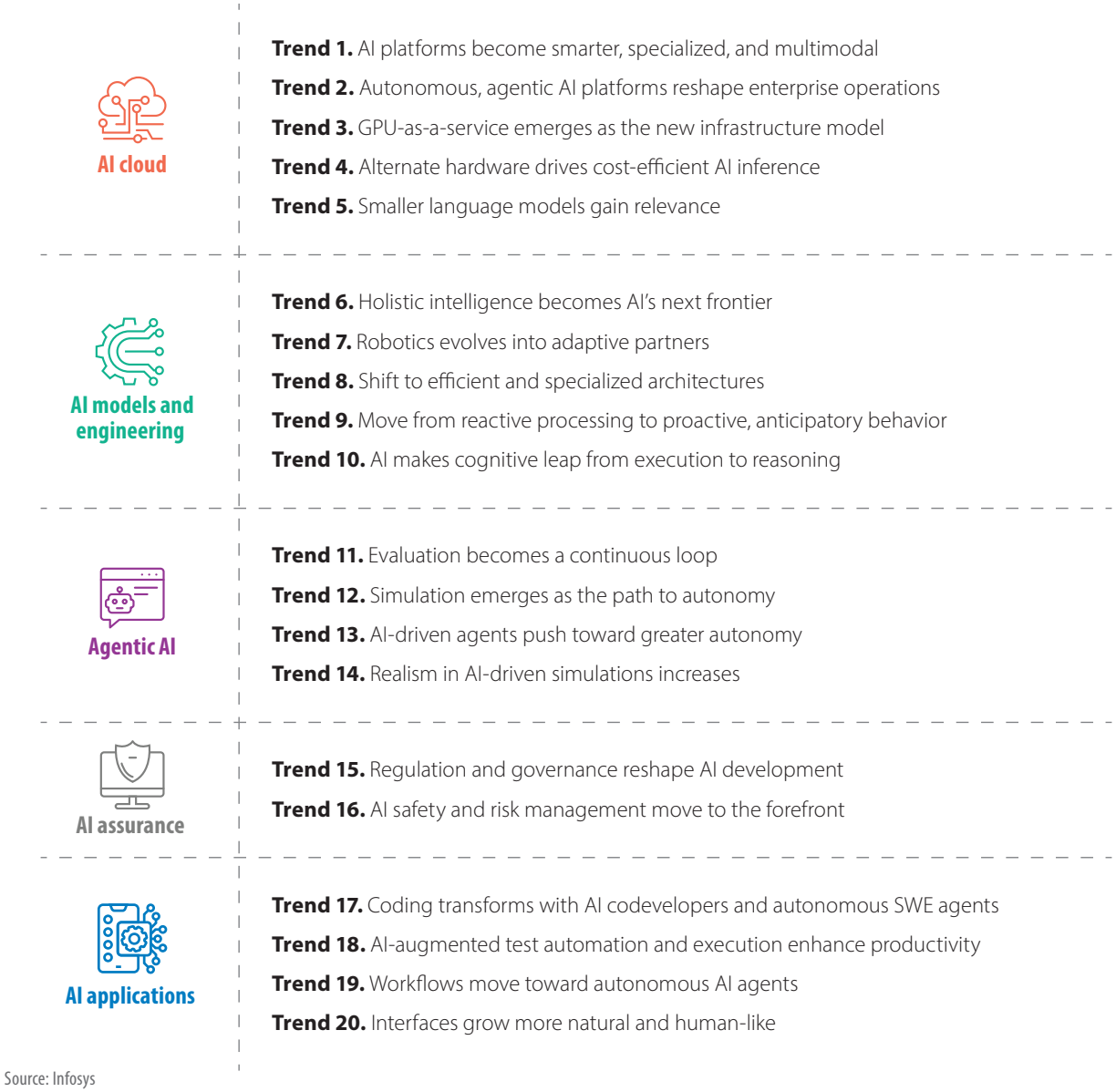
Looking ahead, horizon 3 focuses on the next wave of intelligent systems: autonomous agent swarms, advanced cognitive models beyond transformers, and contextual, empathetic AI collaborators across immersive interfaces. These systems integrate perception, reasoning, and action to enable adaptive real-world autonomy — all while maintaining comprehensive assurance frameworks that ensure trust and reliability.

Figure 1. Market dynamics across the three horizons



Source: Infosys

Figure 2. Horizon journey across subdomains



Source: Infosys

AI cloud

AI cloud is evolving from mature enterprise deployments to increasingly autonomous, real-time systems. Enterprise platforms have consolidated on proven MLOps foundations with cloud-native, standardized GPU and accelerator clusters, plus stable networking and storage. This has enabled large-scale training and inference, reflecting broad, mainstream adoption.

In the near term, the focus is shifting toward LLMOps and agentic orchestration — building robust pipelines for prompting, fine-tuning, evaluation, governance, and observability, coupled with newer inference accelerators. These advancements boost iteration, reduce latency, and lower total cost of ownership as successful implementations expand from early experiments.

Looking ahead, attention is turning to agentOps and multiagent coordination — systems that plan, reason, and collaborate with shared memory and goals — augmented by personalized learning loops that adapt models to users, data, and contexts. Next-generation accelerators and inference silicon on ultra-low-latency, tightly coupled compute aim to enable real-time autonomy and efficiency, as pilots mature toward broader relevance.

Trend 1: AI platforms become smarter, specialized, and multimodal

The AI landscape is evolving from general-purpose platforms toward specialized solutions tailored for specific tasks and industries. This trend encompasses generative AI platforms for content creation, computer vision systems for visual data analysis, and conversational AI frameworks for virtual assistant development. Each specialization offers optimized performance and features aligned with particular use cases, enabling more efficient and effective AI deployment across diverse sectors.

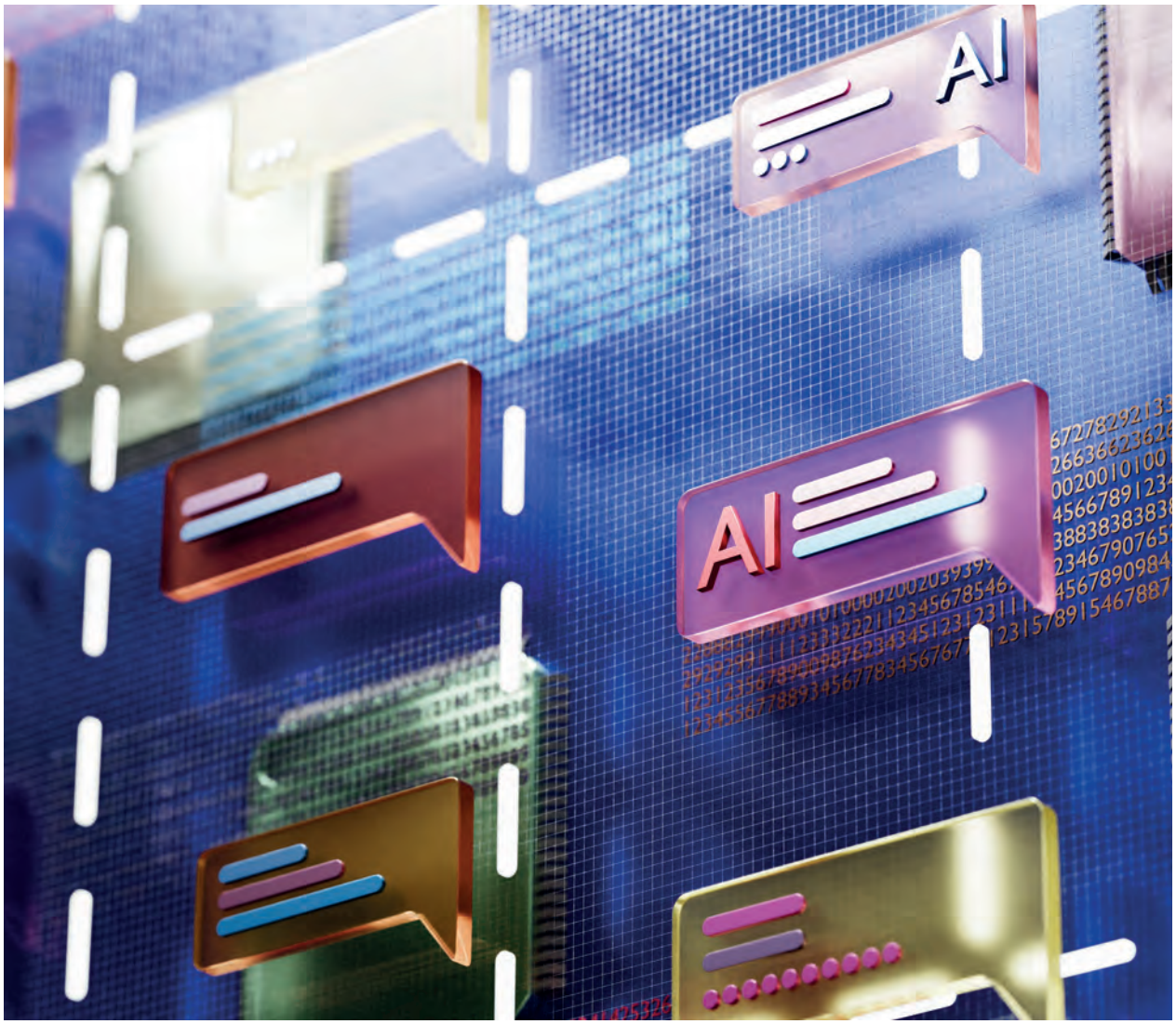
Multimodal AI platforms are gaining significant traction by processing and integrating information from multiple data sources, including text, images, and audio. These comprehensive systems create

more human-like AI experiences by understanding context across different modalities, leading to richer interactions and more nuanced responses.

This shift toward specialization and multimodal integration marks AI's maturation — moving beyond one-size-fits-all solutions to deliver targeted, context-aware capabilities that mirror human thinking.

Trend 2: Autonomous, agentic AI platforms reshape enterprise operations

Agentic AI platforms represent a significant evolution from simple automation to AI agents that autonomously reason, plan, and execute actions to achieve specific goals, bridging traditional rule-based automation with adaptable, decision-making AI capabilities. These systems enable independent operation through informed, context-aware decision-making without constant human oversight, leveraging LLMs, natural language processing (NLP), and multimodal AI to process vast data volumes, understand language nuances, and make real-time contextual decisions. Built to learn continuously, they utilize reinforcement learning to improve performance through feedback and experience. At the same time, they seamlessly integrate with enterprise systems like customer relationship management (CRM) and enterprise resource planning (ERP), enabling agents to access and interpret data across organizational knowledge bases. The platforms support scalable multiagent deployments with specialized agents coordinating through orchestration frameworks to achieve broader objectives, with monitoring and governance to track performance, ensure compliance, and balance autonomy with human oversight.



A global footwear conglomerate, in collaboration with Infosys, co-created an agentic solution to address inventory discrepancies. The staged implementation helped the client transition from a monolithic architecture to one built on poly cloud and poly AI principles of modularity, event-driven design, and seamless integration.

Trend 3: GPU-as-a-service emerges as the new infrastructure model

As AI adoption matures, organizations — especially in regulated industries — are building on-premises GPU-as-a-service (GPUaaS) platforms for training and inference. Key capabilities include: training clusters for batch workloads, serverless or Kubernetes-based clusters with auto-scaling for inference, intelligent GPU resource management for optimal workload scheduling, and multitenant support.

A leading European telecom provider is developing a sovereign GPUaaS platform that delivers secure, high-performance GPU infrastructure to its clients. This initiative has enabled the company to support large-scale model training while eliminating the need for hardware ownership and management.

Trend 4: Alternate hardware drives cost-efficient AI inference

There is increasing demand for ROI-positive, power-efficient compute options for generative AI applications, particularly in regulated environments and lower-throughput use cases.

Recent innovations include Intel Xeon Gen 6 with performance cores (P-cores) and AMX instruction sets and AMD EPYC 5th Gen with AVX-512, DDR5 memory, and PCIe Gen 5 I/O. These CPUs make it feasible to run small- and medium-sized models efficiently, reducing total cost of ownership while broadening deployment options.

Trend 5: Smaller language models gain relevance

While LLMs continue to dominate, small language models (SLMs) are gaining traction as lightweight, cost-sensitive, and domain-specific alternatives.

Using techniques such as pruning, quantization, and knowledge distillation, SLMs deliver strong NLP performance with lower latency, faster inference, and reduced computational overhead. Their compact size enables their deployment on edge devices and embedded systems, extending AI to smartphones, IoT, and other hardware platforms.

Easier to train, fine-tune, and debug, SLMs accelerate prototyping and development, making them a practical choice for enterprises seeking efficient, accessible, and scalable AI.

AI models and engineering

The AI models and engineering segment has advanced rapidly, moving from foundational dual-modality, single-task systems to integrated, intelligent platforms, and now toward autonomous systems with cognitive autonomy, capable of world modeling and seamless human collaboration.

Early development centered on production-proven, foundational AI. Transformer architectures like BERT and GPT handled single-modality tasks reliably within established MLOps stacks. In multimodal AI, tools like CLIP made it possible to check online content and turn text into images. In robotics, the focus was on helper robots and exoskeletons that could follow clear commands — like voice or hand gestures — to do specific, well-defined jobs.

The current phase reflects a shift to integrated intelligence and operational scaling. Multimodal large language models (MLLMs) now analyze dynamic media like video and audio, functioning as tool users and automating workflows through techniques like multimodal retrieval-augmented generation (RAG). LLMOps streamline operations, while transformer architectures have evolved into more efficient and specialized variants like mixture-of-experts (MoE) models and FlashAttention. Robotics has become perception-driven, moving beyond preprogrammed responses to cognitive manipulation.

Looking ahead, AI is advancing toward autonomous cognition. Next-generation architectures like state space models (SSMs) and physics-informed neural networks (PINNs) support embodied reasoning and causal inference, enabling systems to autonomously acquire new skills, collaborate deeply with humans, and achieve situational awareness and empathetic teaming.

Trend 6: Holistic intelligence becomes AI's next frontier

AI is evolving from narrow, single-purpose models into systems capable of cross-modal reasoning and integrated cognition. Foundational models like BERT and GPT, which processed text or images in isolation, have matured into multimodal platforms combining text, vision, and audio. Advanced architectures such as Video-JEPA enable richer contextual understanding and workflow automation. Practical implementations illustrate this shift. Infosys' Lex Knowledge Assistant ingests knowledge from multiple sources — including text, tables, and images — and uses RAG to provide contextually relevant answers.

A leading bank in the Netherlands collaborated with Infosys to revolutionize its enterprise content management landscape. Faced with billions of archived documents, the bank grappled with slow retrieval, rising compliance risks, and operational inefficiencies. Infosys engineered a cloud-native, AI-powered platform leveraging RAG to automate document classification, data extraction, and regulatory compliance. This accelerated processing times, reduced operational costs, and boosted workforce productivity. By converting static content into a dynamic, retrievable knowledge layer, the bank unlocked enterprise agility, strengthened data governance, and delivered faster, more secure customer experiences.

A pharma giant implemented an AI-powered platform that processes text, data tables, visuals, and audio transcripts to streamline scientific document creation. By integrating multimodal information, the system enables automated data extraction and smart content generation, enhancing accuracy, efficiency, and compliance in medical documentation.

Trend 7: Robotics evolves into adaptive partners

Robotics is undergoing a profound transformation, moving from task-specific, command-driven machines to adaptive, perceptive partners. Early systems relied on rigid human-robot interaction models, executing only pre-programmed tasks. The new paradigm is perception-driven intelligence, where robots use advanced simulation environments like NVIDIA Omniverse to learn from their surroundings, acquire dexterous manipulation skills, and adapt to dynamic contexts.

This progression points to robots with autonomous cognition and empathetic collaboration. Equipped with causal inference and theory of mind (ToM), they can recognize the goals, emotions, and beliefs of humans and other machines, anticipate needs, respond fluidly, and interact as true partners rather than tools. Advances in vision-language-action models already allow robots to translate natural language and visual cues into physical action. Self-supervised vision-based tactile sensing extends perception to texture and stiffness, while open-vocabulary scene understanding enables recognition of unfamiliar objects in unstructured environments. These steps collectively lay the foundation for intuitive, situationally aware collaboration between humans and robots.

Trend 8: Shift to efficient and specialized architectures

The transformer architecture has been the backbone of modern AI, but it is now being refined for greater efficiency, scalability, and specialization. Variants like MoE activate only the parameters needed for a given task, offering resource efficiency at scale. Techniques such as LoRA and QLoRA allow fine-tuning without retraining entire models, while innovations like FlashAttention improve context handling and computational performance.

Scaling strategies have also become smarter. Speculative decoding and KV caching enhance inference speeds, while optimized attention mechanisms such as multi-query attention (MQA) and grouped-query attention (GQA) further reduce costs. Models like Claude 3 and Gemini 1.5 Pro demonstrate the capacity to process extremely long sequences, pushing reasoning and context boundaries far beyond earlier limits. Looking forward, entirely new paradigms such as SSMs and PINNs are emerging, designed to combine symbolic reasoning with dynamic world models, enabling new levels of autonomy beyond the transformer's constraints.

Trend 9: Move from reactive processing to proactive, anticipatory behavior

AI is progressing from a mere data processor to an entity that constructs and interacts with a rich, internal model of reality. Early systems operated without a comprehensive understanding of their environment, performing tasks based on explicit data. The next phase, integrated intelligence, saw the beginning of perception-driven systems, where sensor data fusion and learning from simulation environments began to create a rudimentary internal representation. The pinnacle of this trend is the development of sophisticated world models and context-aware systems. This involves building rich 3D spatial intelligence for applications like "Neural Twins of Wildland" and deploying complex models via EdgeAI for deep, on-device context. This capability will enable AI to move from reactive processing to proactive, anticipatory behavior based on a deep, persistent understanding of its surroundings.

Platforms like NVIDIA Cosmos exemplify this shift by generating physics-based synthetic data to train autonomous systems that act proactively, anticipating changes through a persistent understanding of their surroundings.

Trend 10: AI makes cognitive leap from execution to reasoning

The most significant shift in AI is cognitive: from systems built for reliable execution to entities capable of reasoning and autonomy. Initially, AI excelled at completing predefined tasks. With perception-driven intelligence, it advanced toward adaptive collaboration. Now, the frontier is autonomous cognition, where models can perform causal inference, apply ToM, and autonomously discover new skills.

This transformation is already visible in real-world applications. Bank of America's Erica agent goes beyond command-based transactions to provide financial advice, answering complex customer queries with reasoning-based support. In manufacturing, Visual RAG systems reason across text and images to deliver accurate safety information, moving from simple search to intelligent, life-saving insights. These advances signal AI's shift from being a sophisticated executor to a truly cognitive collaborator, capable of deep situational awareness and reasoning-driven partnership.

> A manufacturing giant adopted a visual RAG system to overcome the limitations of traditional search tools. By integrating text and visual data, the system enabled employees to access accurate safety information across manuals, diagrams, and videos. As a result, the company experienced improved safety awareness, faster incident prevention, and reduced operational risks.

Agentic AI

Agentic AI is progressing from rule-bound automation to increasingly autonomous, adaptive systems — without losing sight of governance. Earlier enterprise deployments centered on goal directed automation built on classical reasoning and scripted agents. These agents executed predefined rules and flows with offline testing and tightly limited autonomy, best for deterministic tasks.

Today, adoption is accelerating for agentic LLM systems that can plan, reason, call tools and application programming interfaces (APIs), and collaborate in multiagent workflows. These systems are typically validated in sandboxes, simulations, and API level environments before phased production rollout, delivering early wins while maturing safety, observability, and policy controls. Another innovation has been the A2A and MCP protocols, which enable agents to engage with legacy commercial off-the-shelf systems in various industries, reducing the need for rip-and-replace modernization approaches.

The next phase advances toward autonomously adaptive swarms trained in rich simulation, blending advanced reasoning with embodiment and digital twin or physics environments to achieve real time decisioning and reliable simulation to reality deployment at scale. The emphasis shifts to robustness, alignment, and cost aware efficiency, as agents coordinate complex, end-to-end tasks with minimal supervision, integrate with enterprise data and systems, and continuously learn from feedback to improve performance and resilience.

Trend 11: Evaluation becomes a continuous loop

Teams are shifting from ad hoc to continuous evaluation loops that use golden sets, synthetic edge cases, rubric-based scoring, and human review to gate changes and monitor live performance. Instead of treating evaluation as a one-time step, organizations are building ongoing feedback cycles that catch errors early, improve reliability, and reduce risk in production. Offline and API simulations play a bigger role in hardening behaviors before agents are deployed, while continuous monitoring ensures systems remain accurate, policy-aligned, and resilient to drift over time.

A large bank has replaced its quarterly testing with nightly evaluation suites. These tests check whether AI tools respond within time limits, comply with policies, and retrieve information accurately. Only the updates that meet all thresholds are moved forward, helping the bank maintain consistent performance and customer trust while avoiding costly errors.

Trend 12: Simulation emerges as the path to autonomy

Leading teams are validating complex multiagent behaviors through digital twins before releasing them into production. Simulation trained policies and swarm patterns help reduce risks by allowing agents to face rare or high-stakes situations in a safe, repeatable setting. This approach accelerates learning, improves coordination, and lays the groundwork for real time decision-making and reliable simulation to reality deployments.

An energy utility uses a digital twin of its power grid to train agent swarms for fault management. Here, agents learn to reroute electricity, balance loads, and optimize switching plans under a variety of failure scenarios — from equipment malfunctions to sudden demand spikes. Once proven safe and effective, these policies are carefully promoted to live operations with built-in safeguards, improving grid resilience, minimizing downtime, and handling unexpected events without risking service disruptions.

Trend 13: AI-driven agents push toward greater autonomy

AI is transforming simulations from static, rule-based models into adaptive ecosystems where agents learn, coordinate, and optimize in real time. Self-learning robots and digital agents trained in simulated environments now adjust to fluctuating conditions with minimal human intervention — improving flexibility, resilience, and efficiency.

Reinforcement learning and multiagent systems let agents discover strategies via trial and error; LLMs and advanced dialogue systems add context awareness and natural interaction; digital twins fuse simulations with live internet of things (IoT) signals for continuous adaptation; and federated learning enables privacy-preserving collaboration across distributed sites.

This convergence is pushing enterprises toward intelligent operations that generalize from simulation to the physical world, reducing risk while accelerating innovation.

A global logistics provider partnered with Infosys to deploy a multiagent simulation trained via reinforcement learning that dynamically rerouted shipments and tuned inventory under real-time disruptions such as bad weather and labor shortages. The program cut delivery delays by 20% and lifted peak-season satisfaction.

Trend 14: Realism in AI-driven simulations increases

AI is pushing simulations toward lifelike fidelity, enabling virtual worlds that closely mirror physical complexity across automotive, aerospace, gaming, real estate, and retail. Deep learning boosts 3D visuals by improving textures, lighting, and physics; reinforcement learning creates more adaptive, human-like agents; synthetic data platforms like NVIDIA Omniverse generate diverse training data safely; and

physics-informed AI speeds up complex simulations such as fluid dynamics with higher accuracy. Enterprises are integrating digital twins and real-time IoT feeds to continuously align simulations with live conditions, improving design quality, safety, and time-to-market.

Cross-functional collaboration between data science, engineering, and domain experts is becoming essential to tailor virtual environments for industry-specific needs and to operationalize insights across the life cycle — from concept and testing to operations and optimization.

Infosys uses NVIDIA Omniverse and extended reality (XR) to help enterprises visualize and simulate architectural designs in photorealistic environments, optimize layouts for customer experience and energy efficiency, and cut rework by testing scenarios virtually. Alongside the Infosys-Cambridge AI Centre, large-scale multiagent simulations accelerate research-to-enterprise adoption for trustworthy, fast insights.

AI assurance

AI assurance is evolving from foundational governance to advanced, demonstrable trust. Early efforts standardized responsible practices with life cycle governance, enforceable bias and privacy controls, model monitoring, and consistent evaluation metrics, embedding guardrails into development and production to ensure safe, reliable operations, at scale. The focus is now shifting from reactive checks to proactive assurance. LLMOps-driven evaluations, programmable guardrails, and continuous monitoring for drift, hallucinations, and toxicity. Risk, security, and compliance are operationalized with policy-as-code and auditability to align with regulatory expectations across diverse model stacks.

Looking ahead, next-generation assurance will encompass agentic and multimodal systems using simulation-based safety testing, formalized governance, attestations, and policy enforcement, reinforced by advanced privacy and security methods such as confidential computing and federated evaluation. The end state is measurable trust — provable safety, resilience, and compliance. This will enable certification-ready AI and support high-stakes autonomous applications where verifiable assurances are essential to satisfy regulators, customers, and markets.

Trend 15: Regulation and governance reshape AI development

The surge in AI regulation is reshaping the industry — setting boundaries, fostering accountability, and restoring public trust. Growing concerns around ethical AI, risk, and societal harm have prompted major legislative efforts like the EU AI Act and a wave of Responsible AI frameworks. This signals a shift where AI development must balance innovation with transparency, compliance, and ethical rigor.

Enterprises are now embedding governance across the AI life cycle, ensuring systems are auditable, fair, and accountable. While regulation leads the charge, AI techniques enable enforcement. Explainable AI (XAI) is essential for transparency, a key regulatory demand.

Algorithmic auditing tools help detect bias, performance drift, and ethical violations. NLP and knowledge graphs support regulatory interpretation and automate compliance. The Workday lawsuits, alleging AI-based hiring discrimination, underscore these issues. As mandates like New York's Local Law 144 take effect, companies must adopt robust governance to avoid legal and reputational fallout.

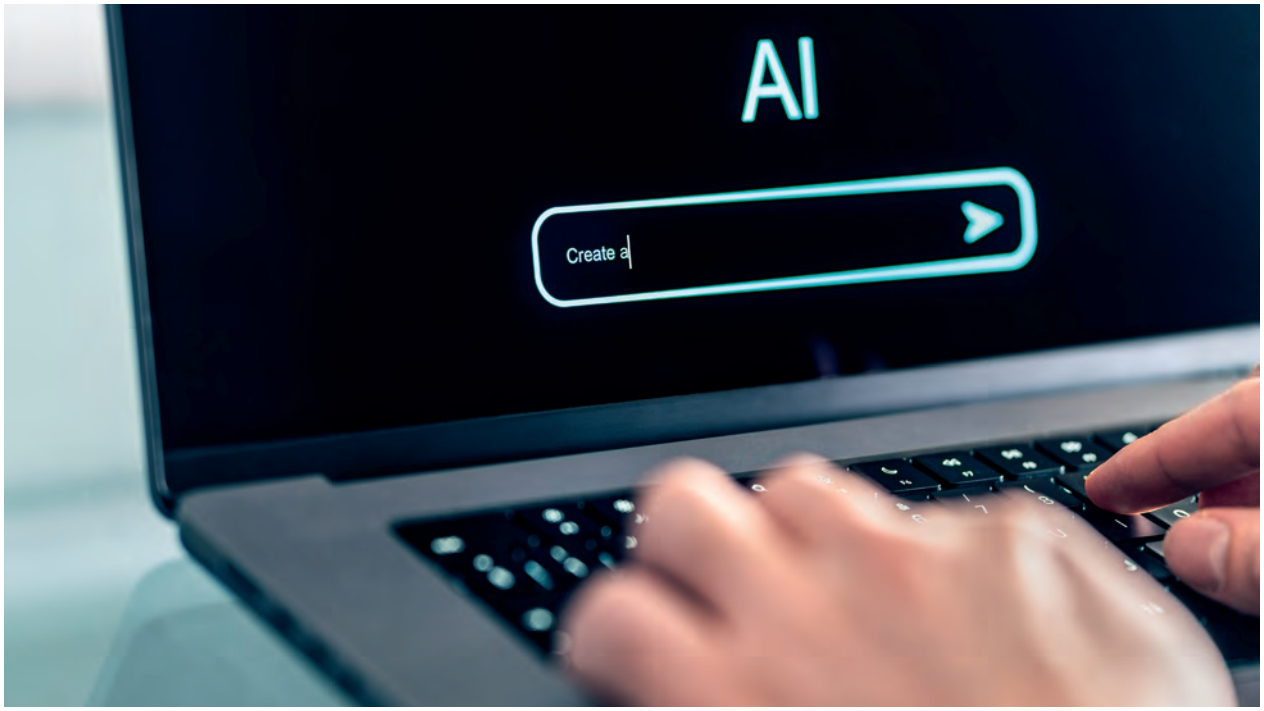
A global enterprise deployed Infosys' responsible AI toolkit to proactively identify and remediate security vulnerabilities across its AI ecosystem. The toolkit not only strengthened compliance with regulatory and ethical standards but also embedded guardrails to ensure that all future AI solutions are developed securely, transparently, and in line with responsible AI principles.

Trend 16: AI safety and risk management move to the forefront

AI safety and risk management are reshaping the industry by prioritizing reliable, secure, and ethical system design. As AI powers critical applications — like autonomous vehicles, trading, and healthcare — safety is no longer optional. This trend drives a shift from theory to practical safeguards against system failure, bias, and adversarial attacks. The future lies in AI systems with built-in safety features, rigorous testing, and real-time monitoring. Techniques such as adversarial machine learning (AML), anomaly detection, and formal verification are central to preventing manipulation and error. Reinforcement learning with human feedback (RLHF) helps align model behavior with human values, while XAI provides critical transparency into decisions. Companies must integrate risk by design, conduct threat modeling, monitor systems continuously, and ensure data quality. A lawsuit against Character.AI spotlights the dangers of unsafeguarded generative AI, reinforcing the urgent need for content moderation, ethical guardrails, and proactive protections — especially for vulnerable users.

A major technology company, with the help of Infosys, is leveraging AI Red Teaming services to assess its AI use case using a curated golden dataset and proprietary adversarial prompts. This engagement uncovered key vulnerabilities, enabling the company to proactively strengthen its AI solution's security and resilience.

AI applications



AI is progressing from embedded utilities to proactive collaborators across work and development. Mature applications — search, recommendations, computer vision, NLP, and productivity automations — are now standardized in office suites and software development toolchains, woven into everyday flows across functions with predictable reliability.

In the near term, the focus is on copilot-style assistants for workplace and software engineering (SWE). These tools generate, summarize, and automate tasks, while enabling team-aware workflows with prompts, RAG, and evaluation loops. Multimodal and voice interfaces begin delivering measurable productivity, albeit with an uneven enterprise-wide scale.

Looking ahead, AI evolves into contextual, empathetic, multimodal, and proactive collaborators spanning text, image, audio, and video. These systems integrate with immersive XR environments and emerging neural or thought-based interfaces, orchestrating end-to-end development and knowledge work. With safe autonomy, real-time assistance, and continuity across devices and environments, they move from tool-centric helpers to trusted coworkers, elevating creativity, speed, and decision quality while aligning outputs to organizational context and governance.

Trend 17: Coding transforms with AI codevelopers and autonomous SWE agents

AI-powered coding assistants like GitHub Copilot and Amazon CodeWhisperer suggest code completions, identify potential coding errors, and generate entire code blocks based on comments and context. This reduces development time, improves code quality, and helps developers focus on higher-value designs and architecture tasks. However, agentic AI speeds up software development, owing to its ability to manage open-ended problems, multistep processes, and its skill to improve over time. Agents we've tested can achieve between 80% and 90% improvement in database code generation; between 60% and 70% improvement in generating APIs and microservices; and up to 60% improvement in generating user interface code.

Infosys built an autonomous SWE agent workflow for a hi-tech major to streamline application error handling. It automatically collects, classifies, deduplicates, and organizes errors from telemetry data. Using LLMs, raw logs are converted into structured information for the SWE agent to resolve issues and generate pull requests, which are validated by humans before deployment — significantly accelerating error resolution and time to market.

Trend 18: AI-augmented test automation and execution enhance productivity

AI-based test automation tools are revolutionizing QE workflows. These tools predict potential failure points through pattern recognition. Using AI to generate, run, and analyze test cases and complex testing scenarios happen autonomously. Currently used in advanced QA labs and pilots, there are self-healing test environments and digital twins for simulating real-world production behavior which has huge potential for enterprise testing pipelines. QE teams deploy AI to maintain test stability during rapid development cycles with self-healing test scripts that automatically adapt to user interface (UI) changes.

Trend 19: Workflows move toward autonomous AI agents

AI is advancing from task-level support to managing entire workflows. Autonomous agents can now handle customer inquiries, process returns, recommend personalized solutions, and even automate human resource processes such as screening and recruitment. AI agents are already enabling organizations to reimagine and re-engineer business processes — when the ROI is justified — and on a smaller scale, drive process automation and simplification. Beyond re-engineering, ongoing experiments pave the way for vibe coding and exploration for business users to continuously innovate.

Many large enterprises have found that implementing intelligence at scale is easier said than done. They need an architecture-first approach, and they need to treat AI as a transformative program. At the same time, it is important to set up the right hub-and-spoke operating model, backed by a value- and outcome-driven AI investment approach. On the technology front, building AI runways that power platform and democratization capabilities across all layers of the architecture is proving key to effective AI scaling.

An e-commerce platform adopted AI agents to autonomously manage customer support — resolving queries, processing returns, and providing tailored product recommendations, enhancing customer experience and reducing manual support efforts.

Trend 20: Interfaces grow more natural and human-like

Interfaces are becoming more natural and intuitive, using gesture control, voice recognition, and gaze tracking to enable safer, handsfree access to devices and data. These capabilities are particularly valuable in high-stakes environments such as manufacturing, healthcare, and logistics.

With the advent of agentic UI design, AI can bring an entirely new and personalized approach to user engagement in a process. Rather than having to rely on a monolithic process design, user queries can be delivered through independent agents that autonomously route queries, working with each other to understand, gather, and validate the correct information needed for a response.

A manufacturing company introduced gesture-controlled systems for its technicians, allowing them to safely access technical information without keyboards in hazardous environments. This improved safety and ensured uninterrupted workflows on the factory floor.

Advisory council

Rafee Tarafdar

Priyapravas Avasthi

Kaushal Desai

Kamalkumar Rathinasamy

Rajeswari Ganesan

Allahbaksh Asadullah

Shyam Kumar D

Syed Ahmed

Producers

Ramesh N
Infosys Knowledge Institute

Pragya Rai
Infosys Knowledge Institute

Harry Keir Hughes
Infosys Knowledge Institute

Contributors

Akshatha Kamath

Chetana Amancharla

Dr. Puranjoy Bhattacharya

Elangovan Thanasekaran

Ganesh Prabhu

Jayarama Nettar

Lokesha Mahadevaswamy

Muppali Govindraj

Nampuraja Enose

Neelmani Verma

Prasanna Gawade

Ramjee R

Rani Malhotra

Rohit Kumar Bathala

Sameer Choudhary

Shakthi Varman

Suchindran Muthu

Ujjwal Tiwari

Venkateshprasanna H. M.

Vishal Manchanda

Vishal Parikh

Vishwa Ranjan

Mayank Gupta

Shreya Bansal

About Infosys Knowledge Institute

The Infosys Knowledge Institute helps industry leaders develop a deeper understanding of business and technology trends through compelling thought leadership. Our researchers and subject matter experts provide a fact base that aids decision-making on critical business and technology issues.

To view our research, visit Infosys Knowledge Institute at infosys.com/IKI or email us at iki@infosys.com.

For more information, contact askus@infosys.com