DW DIGITALISATION WORLD

ISSUE IV 2026

AN ANGEL BUSINESS COMMUNICATIONS PUBLICATION

DIGITALISATIONWORLD.COM

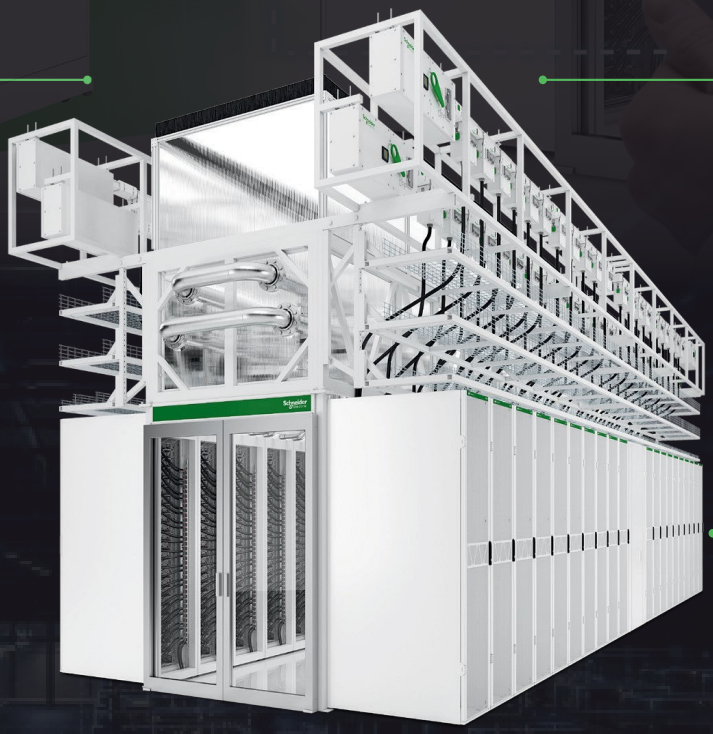
DESIGNING SECURITY FOR REAL-WORLD BEHAVIOUR

Speed Meets Certainty: The Fastest Path to AI-Ready Infrastructure

Accelerate AI Deployments with Prefabricated & Pre-integratedPodand Rack Solutions

EcoStruxure™ Pod and Rack Solutions from Schneider Electric™ deliver a faster, lower-risk method to deploy AI and accelerated compute infrastructure at scale. How? By shifting complexity off-site.

Our solutions arrive fully engineered, factory-assembled, and compute-ready, compressing deployment timelines from months to days, enabling rapid, predictable, and globally scalable AI capacity without comprimising reliability.



Prefabricated for Speed
Factory-built and tested pods and racks reduce onsite work and accelerate time-to-compute.

Deploy in days...
Not weeks.
Not months.

Global Expertise, Delivered Anywhere
Engineering and service experts to support fast, high-density rollouts across 100+ countries.

AI-Ready Power & Cooling
Engineered for the next generation of liquid-cooled AI architectures, with integrated technical water and high-density power pathways built to evolve as workloads grow.

Scalable Pod Architecture
Flexible by design, allowing pod infrastructures to scale from initial deployments to multi-megawatt buildouts with consistent quality across global sites.

Pre-integrated Racks
EIA, ORV3, and NVIDIA MGX racks ship integration-ready for rapid deployment and reliable performance.

Discover the IT Pod

Schneider Electric™

VIEWPOINT

BY PHIL ALSOP
EDITOR

AI – full speed ahead?

Across this month's news cycle, one theme emerges with striking consistency: AI adoption is accelerating faster than the governance, infrastructure and operational discipline needed to support it. Organisations are no longer debating whether to deploy AI. The question now is whether they can do so sustainably, securely and economically.

The challenge is particularly acute in Europe, where the pursuit of AI leadership is colliding with structural inefficiencies in cloud strategy. Insight's findings on cloud waste are sobering: organisations are reportedly losing almost a quarter of their cloud capacity through overprovisioning, poor visibility and inactive resources. At a time when AI workloads are sharply increasing infrastructure costs, that wasted spend represents far more than operational inefficiency — it is becoming a strategic handicap. The emerging "digital sovereignty trilemma" encapsulates the balancing act now facing enterprises and governments alike: controlling costs while maintaining resilience and meeting increasingly stringent sovereignty and governance demands.

This issue of sovereignty runs throughout the broader AI narrative. Dell's research into public sector adoption of agentic AI highlights how governments are shifting from experimentation to implementation, but only where governance, privacy and infrastructure requirements can be satisfied. Similarly, ShareGate's findings reveal that many organisations remain overconfident about their governance readiness, even as AI tools expose long-standing weaknesses in data management and access control. The result is a growing disconnect between AI ambition and organisational preparedness.

At the same time, AI is profoundly reshaping cybersecurity risk. The rise of "shadow AI" — employees using AI tools outside formal oversight — has become a recurring concern across multiple studies. Lenovo, ISACA and Keeper Security all point to the same underlying reality: AI adoption is expanding faster than security frameworks can adapt. The consequences are already visible in identity sprawl, unmanaged machine accounts, increased attack surfaces and the growing inability of organisations to determine whether they have already suffered AI-enabled attacks.

Cybercriminals, meanwhile, are embracing AI with remarkable efficiency. Fortinet's latest threat intelligence demonstrates how AI-enabled offensive tooling is industrialising cybercrime, compressing attack timelines and lowering the skill barrier for attackers. Barracuda's analysis of device code phishing further illustrates how attackers are exploiting legitimate authentication mechanisms to bypass traditional defences. The common thread is that conventional security assumptions are no longer sufficient in an AI-driven threat landscape.

Yet the conversation is not solely about risk. There is also growing evidence that organisations are struggling to measure AI's true value. Harness' research into software engineering productivity highlights the emergence of an "AI productivity paradox," where gains in coding efficiency coexist with hidden burdens such as code review, debugging and cognitive overload. Existing performance frameworks were not designed for AI-assisted workflows, and many organisations are discovering that traditional productivity metrics no longer tell the full story.

What becomes clear across all these developments is that AI maturity is no longer defined by experimentation alone. The winners in the next phase of digital transformation will not necessarily be those deploying AI the fastest, but those capable of building disciplined governance, resilient infrastructure, transparent measurement frameworks and integrated security models around it.

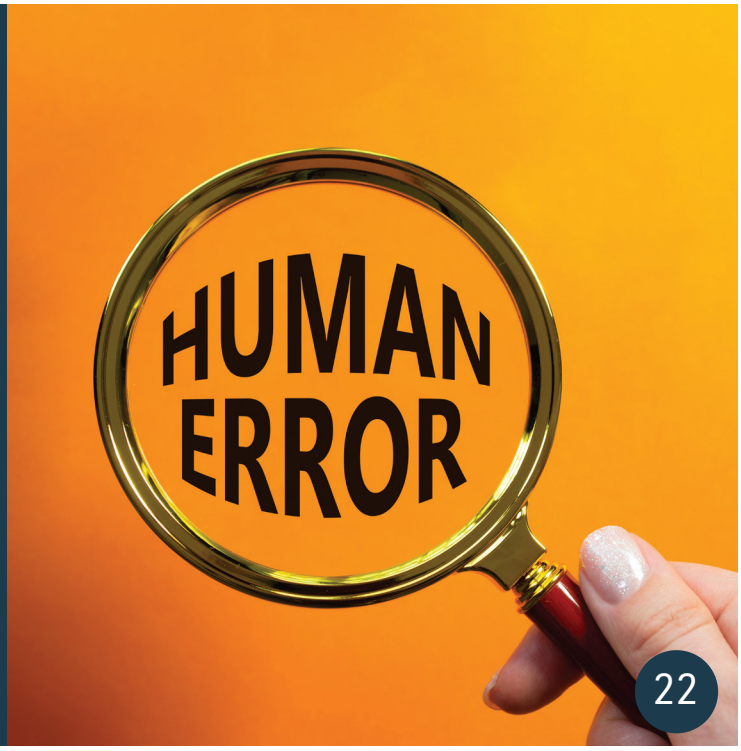
AI is rapidly becoming foundational to economic competitiveness and operational resilience. But without corresponding investment in governance, visibility and strategic infrastructure design, organisations risk creating environments that are more expensive, more fragmented and ultimately more vulnerable than the systems AI was intended to improve.

Contents

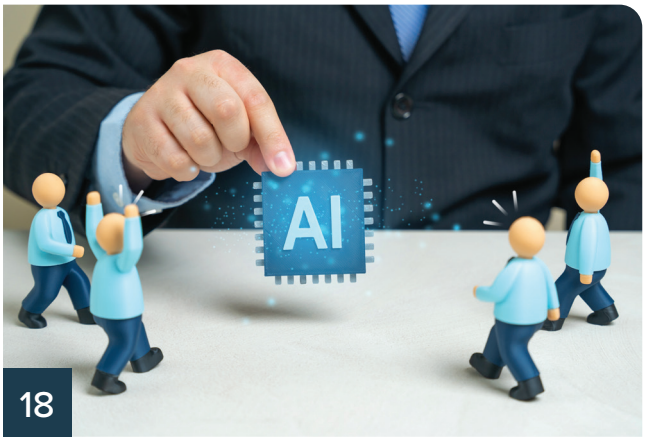
Cover Story

Designing security for real-world behaviour

One of the most persistent misconceptions in cyber security is the belief that human risk is primarily a people problem. In reality, it is a design problem and increasingly, boards, regulators, and threat actors alike recognise it as such



22



18

NEWS

DIGITALISATION WORLD

Editor
Philip Alsop
+44 (0)7786 084559
philip.alsop@angelbc.com

Senior B2B Event & Media Executive
Mark Hinds
+44 (0)2476 718971
mark.hinds@angelbc.com

Director of Logistics
Sharon Cowley
+44 (0)1923 690200
sharon.cowley@angelbc.com

Design & Production Manager
Mitch Gaynor
+44 (0)1923 690214
mitch.gaynor@angelbc.com

Graphic Design & Multimedia Assistant
Harvey Watkins
harvey.watkins@angelbc.com

Publisher
Jackie Cannon
+44 (0)1923 690215
jackie.cannon@angelbc.com

Circulation & Subscriptions
+44 (0)1923 690214
circ@angelbc.com

Directors
Scott Adams: CTO
Sukhi Bhadal: CEO

Published by: Angel Business Communications Ltd, 6 Bow Court, Burnsall Road, Coventry CV5 6SP
T: +44 (0)2476 718970 E: info@angelbc.com

Angel BUSINESS COMMUNICATIONS

Digitalisation World is published 10 times a year on a controlled circulation basis in Europe, Middle East and Africa only. Subscription rates on request. All information herein is believed to be correct at time of going to press. The publisher does not accept responsibility for any errors and omissions. The views expressed in this publication are not necessarily those of the publisher. Every effort has been made to obtain copyright permission for the material contained in this publication. Angel Business Communications Ltd will be happy to acknowledge any copyright oversights in a subsequent issue of the publication.

ISSN 2396-9016 (Online)

Why millions in annual cloud waste is stalling European AI ambitions

Nearly half of European organisations spend up to €5 million a year on cloud – yet a quarter of capacity sits idle.

EUROPEAN ORGANISATIONS are entering the AI era with a built-in disadvantage. New research from Insight shows that cloud-first strategies have created a persistent efficiency tax, with organisations wasting an average of 24% of annual cloud capacity — capital that could otherwise fund sovereign, resilient infrastructure to support AI at scale.

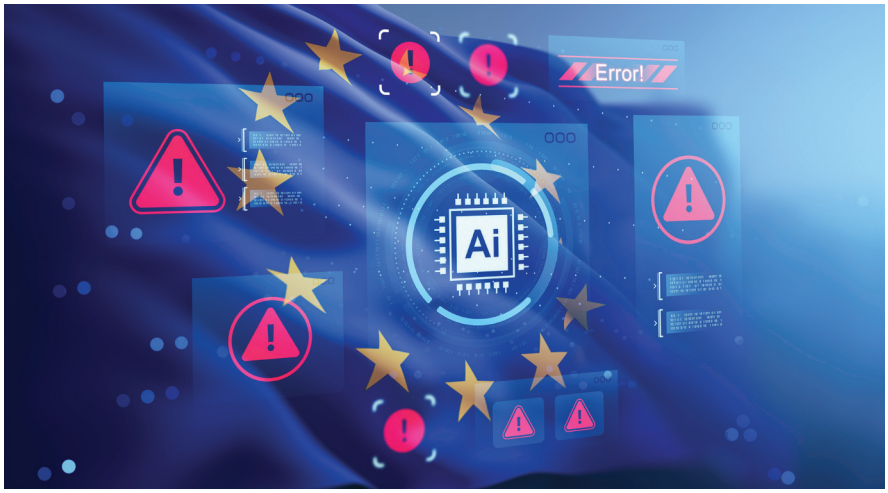
Across EMEA, almost half of organisations spend up to €5 million annually on cloud services. For an organisation with an average cloud spend of €3.75 million, this equates to almost €901,000 in wasted spend each year, limiting investment in AI platforms, data sovereignty controls and long-term infrastructure resilience.

Defining the Digital Sovereignty Trilemma

These findings sit at the heart of Insight's Digital Sovereignty Trilemma report — a research-led framework describing the three competing pressures organisations must balance as AI adoption accelerates:

- **Economic efficiency:** With nearly a quarter of cloud capacity wasted, funding for high-cost AI innovation and advanced data platforms is being eroded.
- **Operational resilience:** To guarantee availability, 47% of organisations over-provision infrastructure, embedding costly "just-in-case" architectures.
- **Data sovereignty:** Increasing regulatory demands around data residency and AI governance are driving workloads towards dedicated or sovereign platforms, often forcing trade-offs between control, cost and performance.

What was once the accepted cost of cloud agility has become a structural



constraint. AI alone is driving a 12% year-on-year increase in hosting costs, while 67% of organisations already view digital sovereignty as a critical strategic priority, rising to 82% within three years. However, over-provisioning (47%), limited visibility (47%) and inactive resources (46%) continue to inflate spend and restrict flexibility.

Gernot Hofstetter, Co-CEO of Yorizon, said: "Insight's Digital Sovereignty Trilemma reflects the reality many European organisations now face: sovereignty, resilience and cost efficiency often pull in different directions. Without deliberate infrastructure design, organisations risk lock-in at a time when digital foundations are increasingly linked to economic and societal stability."

Despite rising costs, 56% of organisations do not carry out total cost of ownership (TCO) assessments before major workload decisions, while 41% remain constrained by legacy applications, making it difficult to rebalance and optimise cloud estates.

Adrian Gregory, President of Insight EMEA, said: "Organisations are wasting nearly a quarter of their cloud capacity just as AI is pushing infrastructure costs sharply higher. To scale AI sustainably, infrastructure must be treated as a strategic asset — reducing waste, applying rigorous TCO discipline and deliberately balancing performance, sovereignty and long-term economic efficiency."

As a result, organisations are increasingly adopting sovereignty-aware hybrid architectures, with 85% already evaluating or deploying dedicated infrastructure for AI. Insight's research highlights a clear opportunity to reclaim wasted cloud spend by improving visibility, removing inactive resources and aligning cloud strategies with AI and sovereignty requirements. In the UK, digital sovereignty is moving rapidly from consideration to expectation. Today, 78% of organisations say it is important, rising to 90% within the next one to two years and 94% longer term. As cloud estates expand and AI pushes infrastructure costs higher, UK organisations are increasingly reassessing workload placement to reduce waste while maintaining control, resilience and flexibility.

Surge in AI-enabled cybercrime

Fortinet leverages threat intelligence to disrupt global cybercrime, transforming awareness into actionable insights.

Fortinet leverages threat intelligence to disrupt global cybercrime, transforming awareness into actionable insights. Fortinet has released the 2026 Global Threat Landscape Report from FortiGuard Labs. Derived exclusively from FortiGuard Labs telemetry, the latest annual report is a snapshot of the active threat landscape and trends from 2025, including a comprehensive analysis across all tactics used in cyberattacks, as outlined in the MITRE ATT&CK framework.

The data reveals that cybercrime no longer functions as a series of isolated campaigns—it operates as a system, with malicious hackers operating across an end-to-end life cycle and compressing the attack life cycle with shadow agents.

Attack Techniques and Targeted Sectors in Today's Threat Landscape
Modern cybercrime crosses borders and sectors, and even traditional definitions of crime itself. As attacks grow more sophisticated and interconnected, key findings from the latest FortiGuard Labs Global Threat Landscape Report reveal:

Velocity defines risk as time-to-exploit (TTE) shrinks: As AI accelerates reconnaissance, weaponization, and execution, fortiguard intelligence shows that TTE as 24–48 hours for critical outbreaks, a sharp increase from earlier reports that revealed a TTE of 4.76 days. Real-world incidents reflect how minutes can define outcomes: Active exploitation attempts were made within hours of the React2Shell vulnerability public disclosure.

Ransomware victims skyrocket: FortiRecon adversary intelligence identified 7,831 confirmed ransomware victims globally, skyrocketing from approximately 1,600 identified victims in the Fortinet 2025 Global Threat Landscape Report. Availability of crime service kits like WormGPT, FraudGPT, and BruteForceAI contributed to this 389% increase year-over-year (YoY). The top three targeted sectors include manufacturing (1,284), business services (824), and retail (682). Geographic concentration includes the U.S. (3,381), Canada (374), and Germany (291).

Identity sprawl defines cloud exposure: FortiCNAPP intelligence confirms that throughout 2025, most confirmed cloud incidents originated from stolen, exposed, or misused credentials rather than from infrastructure exploitation. Sector analysis shows hospitals/physician clinics and retail establishments as the #1 target. Large identity populations, federated access models, and complex cloud integrations make these prime targets for malicious hackers.

Inside the Habits of Modern, AI-Enabled Cybercriminals
As FortiGuard Labs Cyberthreat Predictions for 2026 projected, the most capable threat groups function as semi-autonomous enterprises, supported by shadow agents, access brokers, and botnet operators who provide services on demand. Key findings from the 2026 Global Threat Landscape Report show:

Shadow agents reduce operator skill requirements while increasing workflow speed. FortiRecon dark web signals captured AI-enabled offensive tooling advertised as services and products, including enhanced versions of WormGPT and FraudGPT, and novel services like HexStrike AI, an offensive AI tool with automated reconnaissance attack path generation; and BruteForceAI, a penetration testing tool that integrates large language models (LLMs) for intelligent form analysis and can execute sophisticated multi-threaded attacks.

With AI, criminals work smarter, not harder. FortiGate IPS telemetry recorded a 22% decrease in brute force attempts YoY, pointing to efficiency gains: With optimized, intelligent brute force techniques, threat actors are making fewer attempts against better-selected targets, increasing success probability per credential tested.

This activity translates into about 67.65 billion brute force events globally, with approximately 185 million attempts per day; 1.3 billion attempts per week; and 5.6 billion attempts per month. At the same time, intelligence revealed a 25.49% increase in global exploitation attempts YoY.

Stolen datasets are more popular than leaked credentials. In the 2025 Global Threat Landscape Report, FortiGuard Labs observed a 500% increase in logs available from systems compromised by infostealer malware. In 2026, FortiRecon intelligence found an additional 79% increase and revealed a shift toward theft of more comprehensive data sets, enabled by agentic AI. Within dark web "database" activity, stealer logs dominated advertised and shared datasets (67.12%), exceeding combolists (16.47%) and leaked credentials (5.96%). Stealer logs reduce attacker effort by bundling identity material with contextual artifacts, including browser-resident data, enabling immediate replay and faster conversion than brute force or password spraying.

Decoding the AI productivity paradox in software development

Harness report finds that AI coding tools are being widely adopted in software engineering, with productivity gains alongside emerging challenges in performance measurement and visibility.

THE INTRODUCTION of AI coding tools is influencing the software development landscape, with engineering teams reporting changes in productivity alongside increasing challenges for existing measurement frameworks to keep pace with these developments.

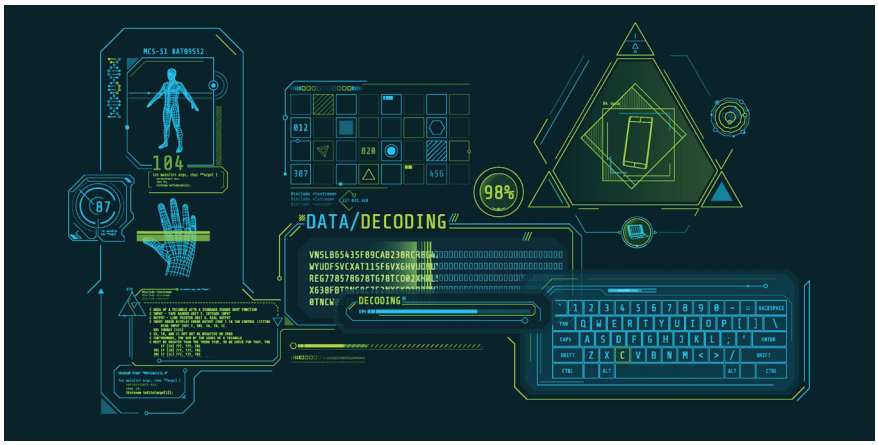
According to The State of Engineering Excellence 2026 report by Harness, which gathered insights from 700 engineering practitioners and managers across multiple countries, AI adoption is now widespread in software engineering. However, organisations are assessing how to interpret reported productivity changes and their associated costs.

A reported 89% of engineering leaders indicate improvements in developer productivity, while 88% report increased developer satisfaction. At the same time, 81% of respondents note that developers are spending more time on manual work, particularly code review. The report also highlights that approximately 31% of developer time is now spent on tasks such as reviewing AI-generated code, fixing bugs, and switching between tools, much of which is not formally tracked.

The findings suggest limitations in existing measurement frameworks:

- 89% of leaders say their current metrics reflect AI's impact, while 94% also indicate that key factors are not captured in those same systems
- Only 6% believe current frameworks are sufficient to address these gaps

The study identifies measurement itself as a key challenge, particularly in assessing productivity, code quality, and return on investment, with organisations relying on dashboards designed for earlier stages of software development.



The report also notes differences in perception between leaders and developers regarding AI-related metrics. While 15% of managers report no concerns about how AI productivity data is used, this compares with 4% of practitioners.

Concerns about monitoring and evaluation are also present, with 54% of respondents expressing concern about performance evaluations based on AI-generated data. Additionally, 55% of developers call for a clearer separation between improvement data and performance evaluation, along with greater transparency and involvement in defining metrics.

While established frameworks such as velocity, DORA, and cycle time remain in use, the report suggests they may not fully capture the effects of AI on development workflows.
The report outlines several approaches, including:

- Expanding metrics to include code quality, validation time, cognitive load, and burnout indicators
- Treating AI performance as a separate measurement area with defined benchmarks
- Increasing developer involvement in how metrics are defined and applied

Overall, the findings indicate that organisations are adapting measurement approaches to account for changes introduced by AI coding tools, particularly in how productivity and engineering output are assessed.

Concerns about monitoring and evaluation are also present, with 54% of respondents expressing concern about performance evaluations based on AI-generated data. Additionally, 55% of developers call for a clearer separation between improvement data and performance evaluation

Public sector's transition to agentic AI: challenges and opportunities

Public sector leaders globally are evaluating agentic AI for autonomous task completion as workforce pressures and data governance requirements shape adoption strategies.

PUBLIC SECTOR leaders across the globe are moving from exploring artificial intelligence (AI) toward implementation, with agentic AI becoming an increasing area of focus.

A study commissioned by Dell Technologies and conducted by International Data Corporation found that 71% of government decision-makers believe agentic AI could accelerate AI adoption in government operations.

The findings come as public sector organisations address workforce shortages, skills gaps, and ongoing modernisation efforts. As a result, decision-makers are placing greater emphasis on the conditions and requirements surrounding AI deployment.

Agentic AI and Workforce Operations: The study found that 51% of leaders plan to invest in agentic AI within the next 12–18 months, with organisations considering autonomous systems for administrative and analytical tasks alongside existing workforce responsibilities.

Skills Gap Challenges: Around 66% of public sector organisations reported that technology is evolving faster than workforce capabilities, contributing to operational and skills-related pressures.

Conditions for AI Adoption: Approximately 44% of respondents said stronger safeguards, including data security, privacy, and sovereignty protections, would influence the pace of AI adoption.

Public-Private Partnerships: The research found that 61% of leaders view public-private partnerships as important for accessing the expertise and technology required for AI implementation.

From Exploration to Deployment: The findings suggest governments are increasingly focused on the practical requirements for deploying AI systems at scale. The study indicates that confidence in infrastructure, governance, and operational readiness may influence adoption timelines.

The report also found that 58% of government leaders identified sovereign data governance, data quality, and control as among the most important platform requirements for sovereign AI deployments, highlighting the role of data management and governance in implementation strategies.

For public sector organisations, AI deployment may require data sovereignty measures, privacy protections, governance frameworks, and supporting infrastructure from the outset. The study suggests these factors are likely to influence how governments approach agentic AI implementation at scale.

Enhancing cybersecurity against industrialised device code phishing

Barracuda research reveals how attackers leverage device code authentication for persistent access, highlighting the need for improved security measures.

RECENT FINDINGS from Barracuda examine the use of device code authentication in cyber attacks, where the technique is used to gain persistent access to services such as Microsoft 365 and Entra ID. Barracuda reports 7 million device code phishing attempts over a four-week period, with the activity associated with phishing-as-a-service tools such as the EvilTokens kit.

Device code authentication allows users to sign in on one device by entering a short code on another device, often used for devices with limited interfaces such as TVs, printers, or command line interface (CLI) tools. Device code phishing involves attackers encouraging users to enter a valid sign-in code on a legitimate login page, which results in authorising the attacker's device.

In this method, attackers request a legitimate device code from Microsoft and use it in phishing messages to

prompt users to authenticate via a real login page. Once authentication is completed, OAuth access and refresh tokens are issued, which can be used by the attacker.

Comparison with traditional phishing approaches includes:

- **Legitimate links:** The method uses official authentication URLs rather than fake websites.
- **Multi-factor authentication:** Because the victim completes the authorisation, standard MFA and conditional access controls may not prevent token issuance.
- **Persistent access:** Refresh tokens can allow continued access even if the user changes their password.
- **Use of familiar workflows:** The process relies on users entering short verification codes, which are commonly used in device linking.
- **Session access:** The attacker gains access through the authenticated session.

Device code phishing can enable access to cloud-based email and identity systems without password theft or triggering some traditional alerting mechanisms.

Barracuda notes that this technique is being used in phishing-as-a-service models, which can increase the scale of such activity. The report also highlights mitigation measures including email filtering, identity protection controls, monitoring, restricting device authorisation flows, and user awareness around entering verification codes only in trusted contexts.

Challenges loom as AI governance struggles to keep pace

AS ARTIFICIAL intelligence continues to be integrated into organisational frameworks, the gap between adoption and governance is becoming more evident. According to a recent study by ShareGate, 29% of organisations have unintentionally exposed sensitive data through the use of AI tools.

At the same time, nearly 93% of IT and security leaders report confidence in their Microsoft 365 governance capabilities to manage AI responsibly. This raises questions about whether confidence levels align fully with existing governance realities and potential blind spots. The types of data unintentionally exposed include customer records (36%), sensitive internal documents (31%), personal data and PII (30%), HR records (30%), financial data (25%), and proprietary intellectual property (21%). Despite these figures, only 51% of organisations have completed a comprehensive governance review since the introduction of tools such as Microsoft 365 Copilot.

Governance teams are also experiencing increased workload pressures. Over 70% report that AI has increased their governance responsibilities, while nearly 80% express moderate concern about AI accessing information that has not recently been reviewed for permissions. The pace of AI development relative to governance processes may increase the risk of exposure to sensitive information that is not fully monitored or controlled.

Rather than creating new governance challenges, AI tools such as Copilot are highlighting existing limitations within organisational systems. In many cases, they make it more visible when information management practices are inconsistent or when visibility over data access is limited. From a financial perspective, AI-related costs are becoming a more significant part of IT budgets.

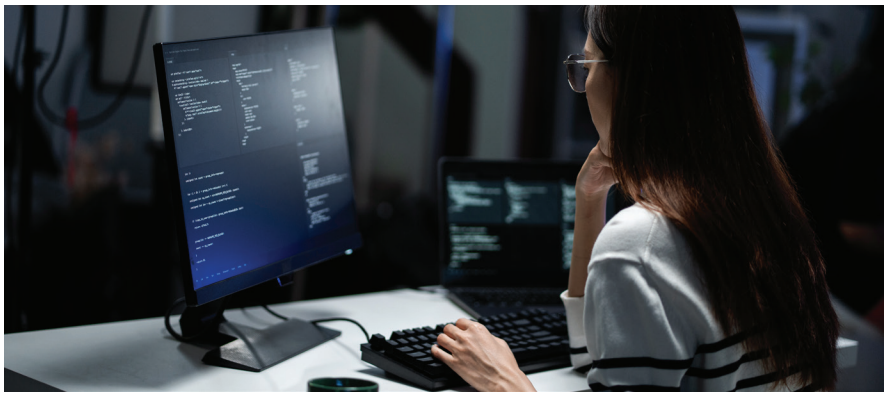
Organisations exposed to attack as one in three rely on adapted IT tools

63% report operational downtime while manual IT/OT coordination continues to slow response.

SOC-as-a-service provider, e2e-assure, has unveiled research revealing that a third of surveyed organisations are relying on IT cybersecurity processes and standards, despite operational technology (OT) requiring a specialist approach, resulting in a preparedness gap that leaves them at increased risk of a cyber attack. The findings show that 32 per cent of surveyed IT Decision Makers admit they are currently relying on detection platforms originally built for IT and "adapted" for OT. This puts organisations at risk, as many are still trying to secure industrial environments with tools that were not designed to understand them.

This is concerning given that 63 per cent of IT decision makers also cited that cyber incidents in the past 12 months resulted in direct operational downtime or impacted critical OT/ICS systems.

The research points to structural weaknesses in how incidents are managed across converged environments, as 28 per cent of surveyed respondents still rely on manual or ad hoc coordination between their IT and OT security teams, while 37 per cent of organisations have a shared platform for both IT and OT environments, but full technical integration needs to become a priority. Richard Groome, OT Cybersecurity Specialist at e2e-assure, commented: "Most adapted IT platforms struggle in OT because they're still thinking like IT tools. They can identify anomalies, but they often have no understanding of the business impact they have. OT downtime isn't just a network problem; it's a process problem, and if you can't interpret what an alert means for a running plant or production line, you're not preventing downtime, you're just creating noise."



While extending IT platforms into OT is an obvious route to take, it creates a critical preparedness gap where organisations may have large volumes of data but lack the visibility needed to understand what it means in an operational context.

Without clear insight, teams are unable to interpret alerts or assess their impact on live environments, limiting their ability to act decisively. This is compounded by the fact that only 15 per cent have deployed passive visibility tools specifically designed for industrial control systems, leaving many organisations without the real-time visibility required to translate data into actionable intelligence and reduce operational risk.

The challenge is becoming more acute as connectivity expands, as 70 per cent of organisations have now fully or largely integrated cloud-connected environments into their IT/OT security strategies. However, without improvements in visibility and coordinated response, increased connectivity risks widen the gap between exposure and resilience. At the same time, many organisations are unable to measure the effectiveness of their risk reduction measures, as 28 per cent of businesses still rely on manual or ad hoc coordination between IT and OT teams, and only 37 per cent operate a shared platform to deliver alignment and visibility across teams. "The volume of data being ingested is often not understood or actionable, meaning incidents may still be missed. More connected does not automatically mean more secure, particularly where exposure increases faster than coordinated response capability", added Groome.

Encouragingly, organisations are beginning to recognise that the challenge is not simply a lack of technology, but how effectively it is used. Sixty-three per cent of leaders are increasing budgets for workforce training and role clarity, the highest prioritised budget area.

The research also highlights shifting priorities across OT security programmes, with supply chain risk emerging as a key area of investment following recent breaches. Investment now is critical, given that previously shared findings found the financial consequences of these preparedness gaps are rising, with almost a quarter (23%) of the most severe OT downtime incidents costing over £1 million, and 6 per cent of incidents exceeding the £5 million mark.

AI adoption is accelerating identity sprawl

Keeper Security has released its latest global insight report, "Identity Security at Machine Speed."

THE STUDY examines the challenges cybersecurity decision-makers face as identity ecosystems expand to include humans and a growing number of Non-Human Identities (NHIs), and finds that legacy tools and unchecked Artificial Intelligence (AI) adoption are widening security gaps that attackers exploit.

Conducted with 3,200 cybersecurity decision-makers and senior IT leaders across Europe, the United States, Asia-Pacific and the Middle East, the research explores how the rapidly expanding identity ecosystem, spanning employees, contractors, third parties and machine accounts, is reshaping enterprise security strategy.

Among the key findings:
Identity sprawl is a near-universal challenge: Nearly nine out of ten (89%) senior UK IT leaders report that managing the growing identity footprint is challenging, which falls in line with the global figure, and reflects the scale and complexity of modern security environments. This consensus masks a specific UK pressure point: more than half (52%) of UK respondents cite AI-driven attacks as a key driver of increased security pressure, the highest figure among European markets surveyed.

Control is fragmented, not consolidated: Identity authority is often distributed across systems, with no single cybersecurity control plane. Globally, 96% cited disconnected or poorly integrated security tools as creating exploitable gaps. In the UK, 67% of respondents identify this to a moderate or great extent, above the global figure of 63%, which points to integration complexity as a persistent challenge for UK security teams.

Detection is improving, but exposure windows remain: UK organisations lead European peers on real-time detection,

with 33% identifying credential misuse within minutes – above the global average of 28%. A further 51% detect within hours. However, 14% still take days or longer to identify unauthorised privileged access, representing a meaningful residual risk.

As AI adoption accelerates, new governance gaps emerge:
AI usage is multiplying NHIs: 43% of respondents globally identify AI-related NHI management and security as a top identity governance gap, a figure matched closely by UK respondents at 40%. As AI agents and machine accounts proliferate within UK enterprise environments, the absence of unified governance over non-human identities is creating an expanding attack surface.

Employee AI use is a top concern: Over half (56%) of respondents are concerned about employees inadvertently exposing sensitive information to AI systems, with 55% of UK respondents identifying this as a leading AI security gap. UK organisations also register the highest concern among European markets about AI-driven social engineering and impersonation at 40%, well above the global average of 35%, reflecting heightened awareness of AI-assisted deception as a threat vector.

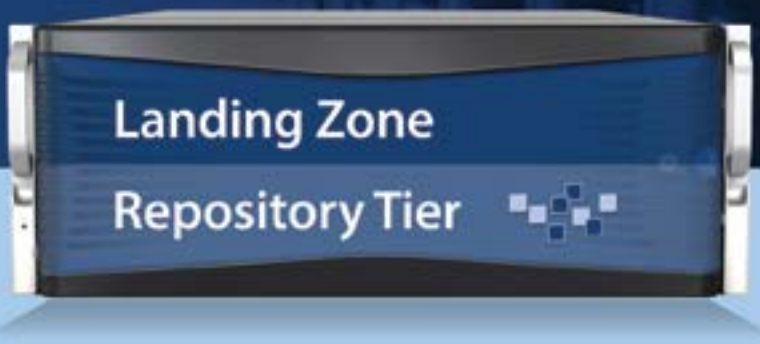
Shadow AI creates blind spots: A lack of visibility into the AI tools employees use was identified as a significant governance gap by 42% of organisations. This sits alongside a broader picture of third-party risk: 34% of UK respondents identify incidents involving third-party vendors or suppliers as a source of increased security pressure, above both the global average of 28% and the figures recorded in Germany and France, highlighting the supply chain dimension of identity risk for UK enterprises.

UK respondents present a picture of above-average threat awareness combined with growing but uneven defensive capability. Over a quarter (27%) report attacks occurring at least weekly. Investment intent is ahead of many markets: 50% of UK respondents are prioritising AI security tools over the next 12 months and 38% plan investment in passwordless or passkey authentication, the highest figure among European markets in the study.

"AI agents, service accounts and machine identities radically outnumber human users in many environments.

EXAGRID
Tiered Backup Storage

The future is here.
Tiered Backup Storage



FASTEST BACKUPS

FASTEST RESTORES

SCALABILITY FOR FIXED-LENGTH BACKUP WINDOW

COMPREHENSIVE SECURITY WITH RANSOMWARE RECOVERY

LOW COST UP FRONT AND OVER TIME

MSP CHANNEL AWARDS
2025 WINNER

- BACKUP & DR INNOVATION OF THE YEAR
- STORAGE HARDWARE INNOVATION OF THE YEAR
- STORAGE VENDOR OF THE YEAR

Thank you so much to all who voted, and congratulations to our fellow MSP Channel Awards 2025 winners!

Visit our website to learn more about ExaGrid's award-winning Tiered Backup Storage.

LEARN MORE >

Gartner identifies six steps to manage AI Agent sprawl

Gartner predicts that by 2028, an average global Fortune 500 enterprise will have over 150,000 agents in use, up from less than 15 in 2025, generating significant agent sprawl, IT complexity and management challenges.

GARTNER has identified six steps to help organizations reduce the risks of AI agent sprawl. Speaking at the Gartner Digital Workplace Summit in London recently, Max Goss, Sr. Director Analyst at Gartner said: "As CIOs and IT leaders see an explosion of AI agents across their organizations, many are contending with an ungoverned sprawl of agents that expose their organizations to a range of risks, including misinformation, oversharing and data loss.

"Organizations need to find a balance where they can govern agents and manage sprawl, but also safely empower employees to innovate with these tools." – Max Goss, Sr. Director Analyst at Gartner "Many organizations resort to blocking or restricting the use of AI agents, but this is not a long-term solution.

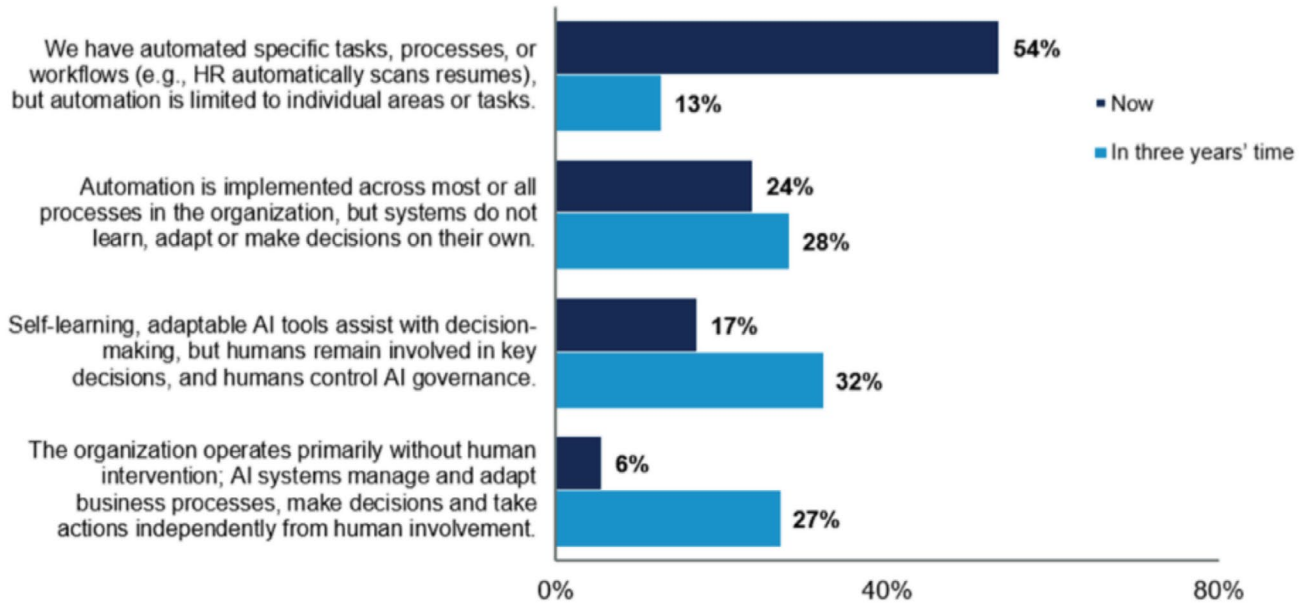
If employees are unable to work in the sanctioned tools, they will likely go around the organization's controls and start using shadow AI which presents far greater risks. Organizations need to find a balance where they can govern agents and manage sprawl, but also safely empower employees to innovate with these tools."

Gartner identified six steps to help CIOs and IT leaders establish governance and guardrails to reduce the risks of agent sprawl.

- **Establish agent governance and policies:** Set clear rules for when and how agents are built, who can create and share them, and what connectors are permitted.
- **Build centralized agent inventory:** Organizations can use AI trust, risk, and security management (AI TRiSM) tools to help discover and categorize agents across applications, both from sanctioned tools, and from shadow AI solutions. Once organizations have an agent inventory, they can start to build adaptive controls to enforce the right policies based on the level of risk the agent presents.
- **Define agent identity, permissions and life cycle model:** Manage the agent identity, permission model and access controls, review, and retire redundant agents to prevent uncontrolled sprawl.
- **Develop AI information governance:** Govern what information the AI tool or agent has access to and ensure that there is a process in place to keep the data current, manage its permissions to prevent oversharing, and archive the data when it is obsolete.
- **Monitor and remediate agent behavior:** Establish ongoing visibility into agent usage, ensure policy compliance, detect anomalous behavior, and correct agents that exceed their intended scope or risk tolerance.
- **Foster a culture of responsible AI usage:** Support the workforce with training programs and a community of practice to drive adoption and amplify best practices on agent management across the organization.

80% of CEOs say AI will force operational capability overhauls

Eighty percent of CEOs expect AI to force a high to medium degree of change to their operational capabilities, shifting the focus from digital business



n = 252, Wave 3 respondents

NQ9. Which of the following most closely describes how autonomous your organization is now and what do you expect it to be in three years' time (i.e. by end of 2028)?

Source: 2026 Gartner CEO & Senior Business Executive Survey, conducted March – November 2025 among 469 global CEOs and senior business executives. Fielded by Gartner using an external panel.

ID:

to autonomous business, according to a survey from Gartner.

"Autonomous business is a strategy where self-learning software agents and machine customers make decisions, take action and create new types of value for organizations. CEOs see this shift as an immediate operational goal," said Don Scheibenreif, Distinguished VP Analyst at Gartner. "While digital business changes what the organization does, autonomous business changes how the organization does it."

The Gartner CEO and Senior Business Executive Survey of 469 CEOs and other senior business executives worldwide was conducted across three quarters, ending in the fourth quarter of 2025.

The survey found that 54% of CEOs said their automation was limited to specific tasks; by the end of 2028, only 13% expect to remain at this level. Conversely, 32% of CEOs expect their organizations to deploy self-learning and adaptable AI tools to assist with human decision-making, while 27% expect their organizations to operate primarily without human intervention, signalling a move to autonomous business ecosystems (see Figure 1).

"CEOs are realizing that AI is not simply another layer of automation. It is a catalyst for rebuilding the enterprise itself." – David Furlonger, Distinguished VP Analyst at Gartner

"CEOs are realizing that AI is not simply another layer of automation. It is a catalyst for rebuilding the enterprise itself," said David Furlonger, Distinguished VP Analyst at Gartner. "This transition to autonomous business requires CEOs to have a capabilitiesfirst mindset that prioritizes how work gets done and how value is delivered in an increasingly autonomous economy."

While automation and autonomous business can provide efficiency gains, they can also become a competitive threat.

Transactional Revenue Is at Risk from AI

Some CEOs expect AI to have a negative impact on their profit models. Twenty-eight percent of CEOs surveyed said transactional revenue was most at risk from AI, as AI agents could bypass existing intermediated systems or their ability to conduct real-time pricing and negotiation.

"As AI agents automate purchasing, pricing, and negotiation, they remove the extra steps and inefficiencies that transaction fees were designed to cover. This is forcing CEOs to rethink profit models and pivot toward recurring, outcomebased revenue models to avoid losing profit," said Furlonger.

Customer base remains unchanged

Only 17% of CEOs expect significant changes to their customer base due to AI, compared to 39% during the digital era. Instead, business leaders are primarily using AI to deepen relationships with

➤ Figure 1: CEO Anticipated Adoption of Automation and Autonomous Capabilities

Source: Gartner (April 2026)

existing customers and, increasingly, machine customers.

Gartner predicts that through 2026, the number of large companies that have a dedicated business unit or sales channel to access fast-growing machine customer markets will double versus 2024.

For CIOs, this underscores the need to build systems that support both human and machine decisionmakers, with trust, accuracy, and data integrity at the core.

"To prepare for this inevitable future, CEOs and CIOs must lead their organizations to rebuild their operational foundations and reengineer their people, assets, and financial structures," said Scheibenreif.

40% of organizations deploying AI will use AI observability

Forty percent of organizations deploying AI will implement dedicated AI observability tools by 2028 to monitor model performance, bias and outputs, according to Gartner, Inc., a business and technology insights company.

"AI is everywhere, but most organizations are still figuring out how to monitor and trust these systems," said Padraig Byrne, VP Analyst at Gartner. "That visibility gap makes scaling risky and that's why observability matters. Unlike traditional software, AI's decision making is often hidden, making it hard to explain or trust, yet errors can cause substantial financial loss, reputational damage and regulatory scrutiny."

Gartner defines observability as the characteristic of software and systems that enables them to be understood based on their outputs and enables questions about their behavior to be answered. AI observability requires dedicated tools that manage and assess the behavior, decision-making and risks of an AI solution, such as model drift, bias and LLM logic.

"The shift to specialized AI observability tools is accelerating due to executive concern over risk management in complex AI models and agentic AI, not just for infrastructure or application health," said Byrne. "There's a growing need for predictive issue detection and real-time actionable insights in AI models. Failure to adopt these tools exposes organizations to significant governance risks." According to Gartner research, AI observability also includes the ability to monitor the availability, performance and accuracy of the AI platforms beyond risk and trust, which becomes essential as enterprises increasingly rely on AI-driven outcomes for decision-making.

"Without clear, standardized model telemetry, infrastructure and operations (I&O) teams face prolonged incident resolution times for AI applications, which will require complex manual efforts to trace and debug the behaviors of opaque deep learning models," said Byrne. "Dedicated AI observability provides the necessary mechanisms to monitor and mitigate algorithmic risk, establishing the technical foundation for widespread enterprise AI trust and adoption."

"Unlike traditional software, AI's decision making is often hidden, making it hard to explain or trust, yet errors can cause substantial financial loss, reputational damage and regulatory scrutiny."
– Padraig Byrne, VP Analyst at Gartner

Gartner recommends I&O leaders factor the following steps into their AI platform strategies:

- Establish mandatory AI model monitoring policies for all production deployments, requiring continuous tracking of fairness, drift and data quality metrics.
- Standardize monitoring frameworks across data science, MLOps and engineering teams to ensure consistency and control. This mitigates organizational silos and streamlines issue resolution.
- Prioritize infrastructure capable of ingesting and analyzing high-volume model telemetry, focusing on specialized solutions that support distributed tracing of AI inference calls.
- Ensure IT strategies include provisions for future monitoring of AI platform performance, detection of shadow IT activity and cost management to address these challenges as the technology matures.

Autonomous business and AI layoffs may create budget room, but not deliver returns

Among organizations piloting or deploying autonomous business capabilities, approximately 80% report workforce reductions, according to a survey by Gartner. However, those reductions do not appear to translate into return on investment (ROI).

The survey found that workforce reduction rates were nearly equal among respondents reporting higher ROI from autonomous technologies and

those experiencing only modest gains or negative outcomes.

Gartner surveyed 350 global business executives in the third quarter of 2025 to understand the current state of autonomous business at enterprises. Qualifying organizations reported enterprisewide annual revenue of at least $1 billion or equivalent, and they had been piloting or had already deployed at least one of the following: AI agents, intelligent automation or autonomous technologies.

Using technologies such as AI agents, intelligent automation, RPA, digital twins and tokenized assets, autonomous business will move organizations from simple augmentation and automation to true autonomy, where both machines and people have more autonomy. This does not mean humanless business; rather, it means human-amplified business.

"Many CEOs turn to layoffs to demonstrate quick AI returns; however, this disposition is misplaced," said Helen Poitevin, Distinguished VP Analyst at Gartner. "Workforce reductions may create budget room, but they do not create return. Organizations that improve ROI are not those that eliminate the need for people, but those that amplify them by aggressively investing more in skills, roles and operating models that allow humans to guide and scale autonomous systems."

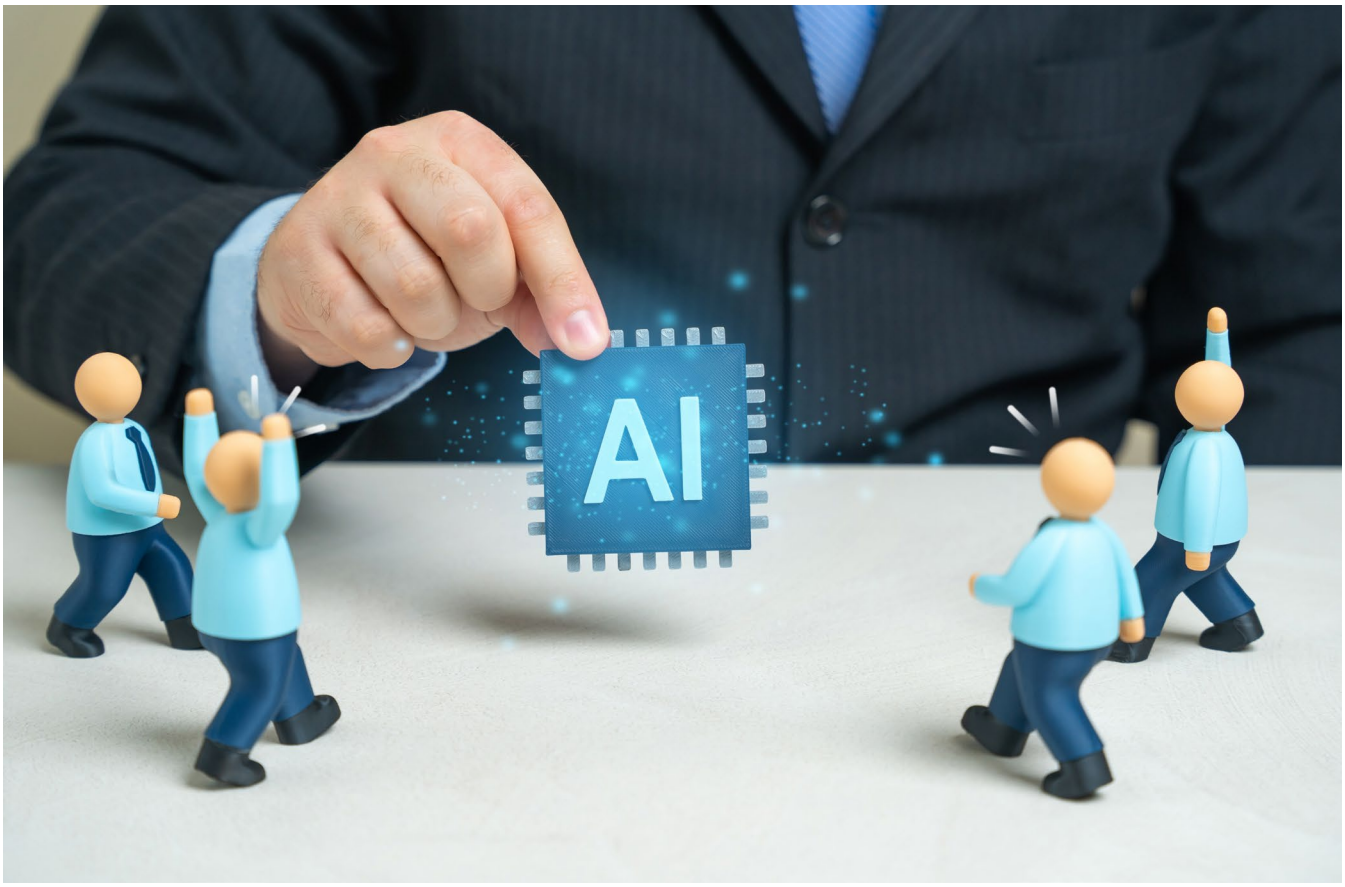
"Long term, autonomous business will create more work for humans, not less." – **Helen Poitevin, Distinguished VP Analyst at Gartner**

Long Term: Autonomous Business Will Create More Work for Humans

Autonomous business will continue to increase with the growing adoption of AI agents. Gartner forecasts AI agent software spending will reach $206.5 billion in 2026 and $376.3 billion in 2027. This is up from $86.4 billion in 2025.

Because autonomy will increase for both machines and people, and the need for people will go up, not down, Gartner predicts that autonomous business will be a net-positive job creator by 2028 to 2029, driven by new forms of work that AI cannot absorb.

"Long term, autonomous business will create more work for humans, not less. Lasting structural factors such as demographic decline and high-stakes, trust-dependent consumer moments will ensure human talent remains central to running, governing and scaling autonomous business," said Poitevin.



$22.5 trillion in value creation ahead

Enterprise AI adoption and agent-driven models scaling toward the end of the decade.

INTERNATIONAL DATA CORPORATION (IDC) has shared key research insights unveiled at Directions, its flagship client event, outlining how artificial intelligence is reshaping the global economy, transforming enterprise decision-making, and redefining how organizations build, buy, and deploy technology.

The research highlighted at Directions focused on five areas: the economic impact of AI, the rise of the agentic buyer lifecycle, the expansion of the AI model landscape beyond LLMs, new frameworks for measuring AI business value, and the emergence of AI agents as a new application model reshaping enterprise software and services.

Together, these trends signal an AI supercycle defined by two phases: infrastructure buildout and enterprise adoption.

"We are entering the strongest technology spending cycle in nearly 30 years, driven by AI and the rise of agents," said Meredith Whalen, Chief Product & Research Officer at IDC.

"But this is not just a buildout story. The real value comes from adoption, and most enterprises are still in the early stages of that shift. The market reaches an inflection point closer to the end of the decade, as AI becomes embedded into how work actually gets done."

AI economic impact: Trillions in value ahead

IDC forecasts that AI will generate $22.5 trillion in cumulative global economic value by 2031, driven by productivity gains, new revenue models, and business transformation.

However, the timeline remains uncertain. Near-term value depends on how quickly organizations move from experimentation to operational deployment, with workforce transformation, upskilling, and AI agents playing a central role.

IDC also highlighted that while the war in the Middle East will stress test the economy via energy volatility, infrastructure resiliency, and supply chain, it will not disrupt the trajectory of the market.

The agentic buyer lifecycle: AI reshapes how decisions are made

IDC research shows buying processes are shifting from human-led journeys to AI-mediated decision systems, where agents shape discovery, evaluation, and selection.

This shift is driving:

- Zero-click, agent-driven discovery
- Reduced brand control over customer relationships

- Increased importance of structured data and agent visibility (AEO)

Beyond LLMs: A multi-model, multi-agent future

IDC introduced new research showing that enterprise AI is rapidly evolving beyond general-purpose models into a multi-model, multimodal, and multi-agent landscape. This shift marks the end of the "one model fits all" approach and introduces a new layer of complexity in how enterprises design, govern, and optimize AI systems. Organizations are adopting "model choice" strategies and must now manage increased complexity in model selection, governance, and orchestration.

Agents as apps: A reset of the enterprise software model

IDC also introduced new research, "Agents as Apps: The Rise of Agents — A Vendor Business Model Reset," which examines how AI agents are redefining enterprise software and services.

The research finds that AI agents are shifting the application model from tools that require user interaction to systems that execute outcomes autonomously at scale. In this model, competitive advantage moves away from user interfaces and toward agents that can reliably deliver results with trust, performance, and economic efficiency.

IDC's research outlines 10 critical moves that enterprise software vendors and service providers must take to remain competitive in the agentic era, warning that without immediate strategic adaptation, organizations may face stagnation or decline.

Measuring what matters: From AI ROI to business value

IDC introduced its Agentic Business Value Maximization Framework to help organizations measure and scale AI impact.

With 42% of organizations struggling to assess AI ROI, the framework emphasizes strategy, use-case prioritization, value mapping, and continuous optimization to move from experimentation to measurable outcomes.

Expanding data and research to track the AI economy

At Directions 2026, IDC also announced a series of new data products and syndicated research programs designed to help organizations track and navigate the rapidly evolving AI market.

These include:

- Robotics data products tracking market share across emerging categories, from commercial cleaning and delivery robots to humanoid systems
- AI infrastructure trackers covering the buildout of AI capacity, including semiconductors, data centers, and sovereign AI environments
- New syndicated research programs focused on critical growth areas such as satellite technologies and agentic AI platforms

These offerings are designed to provide organizations with greater visibility into the technologies, markets, and competitive dynamics shaping the AI economy.

From Experimentation to Execution

IDC emphasized that while AI investment is accelerating, most enterprise adoption remains early and uneven.

The inflection point is expected by 2029, when AI shifts from training to inference at scale and agent deployments reach the billions, embedding AI into enterprise operations. IDC also emphasized that the rise of AI agents and agent-driven systems will accelerate this transition, reshaping how applications are built, deployed, and monetized across the enterprise.

"The next phase of the AI market will be defined by execution," Whalen added. "The opportunity is clear, but execution is now the constraint."

European AI spending to reach $290 billion by 2029

According to a new forecast from the International Data Corporation (IDC) Worldwide AI and Generative AI Spending Guide, European spending on artificial intelligence will reach $290 billion in 2029, growing at a compound annual growth rate (CAGR) of 33.7% over the 2025–2029 forecast period.

Spending will be driven by large spending from banking, retail and software and information services, but also from accelerating industries such as healthcare. Generative AI (GenAI) solutions are already pervasive across enterprise deployments and are expected to account for nearly 54% of the total market by the end of the period.

What is happening in the European AI and GenAI market in 2026?

IDC Worldwide AI and Generative AI Spending Guide forecasts a healthy CAGR through 2029 as enterprises move from AI experimentation to strategic deployment across all major European markets, despite geopolitical tensions and supply chain disruptions.

European AI market at a glance

- Total European AI spending by 2029: $290 billion
- Forecast CAGR (2025–2029): 33.7%
- GenAI share of market by end of period: ~54%
- Largest technology segment: Software (58.5% of total spending in 2026)
- Fastest-growing technology segment: Software (42.9% CAGR, 2025–2029)
- Largest industry: Banking (12.5% of market in 2026)

- Fastest-growing industry: Healthcare Provider (39.7% CAGR, 2025–2029)

"Despite geopolitical tensions and supply chain disruptions, the AI market remains dynamic and is rapidly transitioning from experimental to operational and strategic for enterprises," said Carla La Croce, research manager, Data and Analytics, IDC. "Organizations are no longer treating AI as a standalone tool — they are repositioning it as a strategic asset to transform their business models. The emergence of agentic AI tools has made this transformation more urgent and more profound than many anticipated."

Market dynamics & outlook

Why is European AI spending surging despite macro headwinds?
AI platforms and GenAI solutions deliver measurable returns in cost efficiency, customer experience, and risk management.

As a result, enterprises are accelerating budget reallocation toward AI, shifting from experimental pilots to mission-critical, multi-agent deployments. Software is leading this charge, growing at a 42.9% CAGR with AI Platforms at 52.5%, driven by the explosion of agentic components across industries.

What is IDC's outlook for the European AI market?

European AI spending will maintain strong double-digit growth through 2029, sustained by AI Platform expansion, cloud-native development, and industry-specific AI embedding into enterprise strategies.

Agentic AI is the key growth catalyst. Risks include regulatory fragmentation from the EU AI Act, persistent AI talent shortages, and cloud cost optimization pressures — all of which will create incremental demand for AI governance and assurance services.

> With 42% of organizations struggling to assess AI ROI, the framework emphasizes strategy, use-case prioritization, value mapping, and continuous optimization to move from experimentation to measurable outcomes

Industry Trends: Where AI Investment Is Concentrated

- Banking leads European AI spending (12% of the market), with use cases such as fraud analysis, threat intelligence, contact centers, and customer self-service, with institutions shifting from pilots to mission-critical multi-agent automation and increased focus on FinOps, sovereign cloud, and AI governance.
- Software and information services follow, with most spending on AI infrastructure provisioning to support agentic workloads via PaaS and IaaS.
- Retail ranks third, investing in digital commerce, AI-enabled customer service, planning, personalization, pricing, and supply-chain optimization.

Fast-growing industries

- Healthcare Provider: Fastest-growing industry at 39.7% CAGR (2025–2029) across all five major markets. Primary use case: Clinical Workflow and Resources Optimization.
- Media & Entertainment: 37.3% CAGR, driven by GenAI for content creation, video production, and audience personalization.
- Also, above-average growth: Professional and Personal Services, Utilities, and Life Sciences.

MANAGED SERVICES SUMMIT

BENELUX
LONDON
NORDICS
MANCHESTER

CREATING VALUE with MANAGED SERVICES

managedservicessummit.com

MANAGED SERVICES SUMMIT BENELUX

benelux.managedservicessummit.com

30 JUNE 2026

MANAGED SERVICES SUMMIT LONDON

london.managedservicessummit.com

09 SEPTEMBER 2026

MANAGED SERVICES SUMMIT NORDICS

nordics.managedservicessummit.com

05 NOVEMBER 2026

MANAGED SERVICES SUMMIT MANCHESTER

manchester.managedservices summit.com

17 NOVEMBER 2026

Designing security for real-world behaviour

One of the most persistent misconceptions in cyber security is the belief that human risk is primarily a people problem. In reality, it is a design problem and increasingly, boards, regulators, and threat actors alike recognise it as such.

BY SIMON SEYMOUR-PERRY, CEO OF LOGICA SECURITY

RESEARCH consistently shows that the vast majority of cyber incidents involve human error. Yet most organisations continue to respond by increasing training, tightening policies, and adding layers of control. Despite decades of investment, why are incident levels still so stubbornly high?

The explanation is uncomfortable but clear: many security failures occur not because people are careless, but because the environments in which they operate are misaligned with how work actually gets done.

When security slows execution, interrupts workflow, or makes the secure path harder than the alternative, behaviour adapts predictably. Shortcuts emerge. Informal practices normalise. Controls are bypassed — sometimes unintentionally, sometimes deliberately.

Resilience rarely collapses suddenly. It erodes. And when it does, the consequences are operational as much as technical: disrupted services, financial loss, regulatory scrutiny, and damaged trust.

Forward-looking organisations are recognising a critical truth: **Security that works in theory but fails in practice is not resilience, it's exposure.**
By designing controls around real workflows, decision points, and incentives, these organisations reduce risk while simultaneously improving operational performance. Well-aligned security minimises disruption, supports productivity, protects revenue, and strengthens confidence in the organisation's ability to operate under stress. Security, in this model, becomes not just protective but economically enabling.

Security as friction is a structural risk

Across industries, a familiar pattern persists. Complex password requirements drive insecure storage and credential reuse. Authentication processes disrupt workflow continuity, encouraging shortcuts. Approval chains designed to control access instead teach employees how to route around them when urgency rises. On paper, these environments appear controlled. In reality, they are fragile.

The gap between documented control and operational behaviour creates the conditions for both unintended error and deliberate misuse.

The issue is not awareness alone. Most professionals understand what is expected of them. The deeper problem is structural: security is too often experienced as friction — competing with productivity, service continuity, and commercial outcomes.

Faced with this tension, people respond rationally. They prioritise delivery. Over time, workarounds become embedded in the operating model. Vulnerabilities accumulate quietly until they surface as incidents.

Poorly aligned security therefore creates a dual cost. Not only does it elevate cyber risk, but it also suppresses operational efficiency. Organisations that redesign controls so the secure path is also the easiest path achieve something strategically powerful, they reduce exposure while improving execution. Security stops being organisational drag and starts enabling performance.

Accountability has changed the conversation

The shift underway is not driven solely by attackers. It is being accelerated by regulators. Supervisory expectations have moved beyond demonstrating that controls exist. Increasingly, regulators are asking a far more demanding question: Can the organisation continue to operate securely when conditions are no longer normal?

On the frontline, this includes scenarios where:

- Operational pressure intensifies
- Decision velocity increases
- Systems degrade
- Suppliers fail
- Human error rises
- Malicious behaviour is attempted

This question reaches far beyond cyber tooling. It interrogates how organisations behave under stress and whether important business services remain within tolerance when disruption occurs.

For boards, this marks a governance inflection point. Cyber resilience is no longer a technical matter that can be delegated downward. It is now directly tied to operational continuity, financial stability, regulatory confidence, and enterprise value.

Leading organisations understand that resilience is not merely defensive, it is commercially material and becoming a performance characteristic. Those that design security to function in real conditions experience fewer operational disruptions, lower incident costs, faster recovery, stronger execution under pressure and ultimately, greater stakeholder confidence.

From behaviour correction to environment design

The organisations responding most effectively are no longer attempting to "fix people." They are redesigning the environments in which decisions occur.

Rather than relying primarily on vigilance, they embed security directly into workflows, tooling, and operational processes. This not only reduces reliance on individual effort but also strengthens guardrails against misuse. Controls are aligned to real roles meaning security supports decisions in real time. Put into action, this ensures operational pressures are designed for, not ignored.

This shift is particularly critical in highly regulated sectors such as financial services and critical national infrastructure, where resilience extends well beyond corporate IT estates.

Large portions of the workforce operate across branches, control rooms, operational sites, and data centres, all environments where access decisions are simultaneously physical and digital, and where hesitation carries real-world consequences for everyday citizens. When resilience is designed only through a traditional cyber lens, organisations often default to manual processes, shared access, inconsistent safeguards, or locally developed workarounds.

The result is predictable: A widening gap between policy and practice — and rising operational risk. By contrast, organisations that align security with the realities of delivery streamline execution, strengthen accountability, reduce avoidable delay, and protect revenue-generating services. Security becomes less about restriction and more about enabling reliable performance.

Critically, these organisations validate their designs, meaning assurance shifts from theoretical to observable. Through scenario testing, operational exercises, and real-world simulation, they generate evidence that controls hold under pressure.

The emergence of human centric resilience

Out of this shift has emerged a more mature operating philosophy: Human Centric Resilience.

And its premise is straightforward:

Organisations are resilient when they are designed to operate securely in the real world, not just in control frameworks.

This requires anchoring security to important business services rather than abstract control sets, understanding where human judgement materially affects outcomes, and shaping environments that guide behaviour toward secure action while constraining unsafe or malicious activity.

Just as importantly, it requires evidence, not assumptions, that

services can remain within tolerance during disruption. Organisations adopting this approach recognise that resilience is both protective and economically significant.

By removing the structural conditions that drive unsafe behaviour, organisations can: lower incident frequency, reduce operational drag, protect revenue and improve execution consistency, strengthening stakeholder trust in the long term.

The most resilient organisations do not simply recover faster; they fail less often.

Through deliberate design and continuous validation, they reduce exposure before it materialises — enabling more predictable operations and supporting long-term value creation.

Why boards are paying attention

For boards, this evolution presents both a strategic challenge and a material opportunity. Organisations that embed resilience into their operating model do more than satisfy regulatory expectations, they perform with greater consistency and confidence.

By supporting secure behaviour and constraining misuse, they minimise disruption, protect critical services, and strengthen organisational trust. Those that fail to adapt face a growing gap between perceived resilience and actual performance under stress

Controls that appear robust on paper can falter rapidly in live conditions, particularly when human behaviour intersects with poorly aligned systems.

The debate over whether human factors matter is over. The real question now is whether organisations continue attempting to correct behaviour or redesign the systems that shape it. Because in 2026, resilience is not defined by policies. It is defined by performance under pressure.

Organisations that design for reality will be better positioned to operate securely, respond decisively, and sustain enterprise value in an increasingly volatile environment.

Preparing cryptography for the Quantum Era: why waiting is the biggest risk

Quantum change is coming – you don't get to choose when, only how prepared you'll be.

BY MICHAEL FASULO, SENIOR DIRECTOR OF PORTFOLIO MARKETING, COMMVAULT

DATA you consider safe today may already be compromised; you just won't know for another decade. That's the uncomfortable truth few leaders want to acknowledge. The race toward quantum advantage is accelerating, and while timelines remain uncertain, the threat to today's cryptography is very real, very structural, and already underway.

For years, public key cryptography has quietly underpinned global digital trust, from banking transactions and identity systems to billions of secure connections every day. Its strength relies on the computational difficulty of certain mathematical problems. But quantum computing flips that equation. Once sufficiently powerful systems emerge, the cryptographic foundation of the modern internet becomes instantly obsolete.key cryptography has quietly underpinned global digital trust This isn't a future security issue. It's a now problem.

The threat is already here

There is growing evidence that threat actors are engaging in "harvest now, decrypt later" campaigns, stealing encrypted data today with the expectation they'll decrypt it once quantum capabilities arrive.

And that should raise alarms. The most valuable data for cybercriminals – financial records, intellectual property, medical histories, operational intelligence, etc. – is exactly the kind of information that requires long-term confidentiality. In other words, quantum computers don't need to exist to break your encryption; your adversaries only need patience.

That's the essence of the quantum risk.

Post-quantum cryptography moves from theory to realityquantum Cryptography Moves

The good news: momentum is building. In 2024, NIST released its first standardised post-quantum cryptographic (PQC) algorithms – a milestone that begins moving PQC from academic concept to practical implementation. The mindset of cryptoagility is also shaped by hard learned lessons, after several promising algorithms ultimately failed to deliver adequate protection.quantum

cryptographic algorithms
The bad news: implementing new cryptographic standards across real world, highly entangled IT environments is slow and complex, often requiring five to ten years. Organisations that wait for quantum systems to arrive will be far too late.

Where leaders must start

Preparing for PQC isn't as simple as swapping one algorithm for another. It requires understanding how deeply cryptography permeates your digital ecosystem, across applications, data storage, networking, identity, and external integrations.

And here's the part many CIOs underestimate: cryptography isn't a background control; it's a bedrock dependency that defines your resilience.

Not all data needs the same handling. A risk-based prioritisation model helps determine where PQC must be implemented first, especially for data with long-term sensitivity and retention.

Migration will need to occur in phases, starting with your most critical systems.based prioritization model helps determine where PQC must be implemented first, especially for data with longterm sensitivity.

Equally important is developing crypto-agility, the ability to pivot algorithms as standards mature and threats evolve. The quantum timeline is opaque; your cryptographic strategy can't be.

This work cannot happen in isolation

Your suppliers, partners, service providers, and integrations also rely on cryptography. Any weak link can undermine an entire chain of trust. PQC adoption requires ecosystem alignment, not siloed technical updates.

The window for preparedness is shrinking

Organisations that will thrive in the post-quantum future will be the ones that start transitioning now, with urgency and clarity. Those who wait will face emergency-mode cryptographic overhauls, heightened risk exposure, and potential long-term data compromise.quantum future will be the ones that start transitioning now, with urgency and clarity. Those who who wait will face emergencymode cryptographic overhauls, heightened risk exposure, and potential longterm data compromise.

Quantum change is coming – you don't get to choose when, only how prepared you'll be. I always recommend starting small: identify where cryptography sits in your environment, prioritise the data that must remain confidential the longest, and apply post-quantum approaches in low-risk areas to build confidence.

From there, expand steadily, strengthening crypto-agility and treating cryptographic readiness as a business imperative rather than a one-time upgrade. The organisations that take these early, incremental steps now will be the ones standing on solid, resilient ground when the quantum era arrives.

When AI hacks AI, the victims are still human

With AI agents now a key part of many organisations' 'attack surface', AI systems are a crucial focus for cyber defence.

BY JORGE MONTEIRO, CEO OF ETHIACK

FOR DECADES, enterprise cybersecurity was built around an axiom: users are human.

That assumption has been smashed as individuals and organisations introduce a new type of user - agentic AI - into their IT systems.

We are quickly moving beyond chatbots to autonomous "action agents", AI systems that can log into platforms, manage subscriptions, process invoices, interact with SaaS tools and run operational workflows on behalf of employees.

Their purpose is efficiency. They remove manual work and automate routine processes. But from a security perspective, they fundamentally change what a user is.

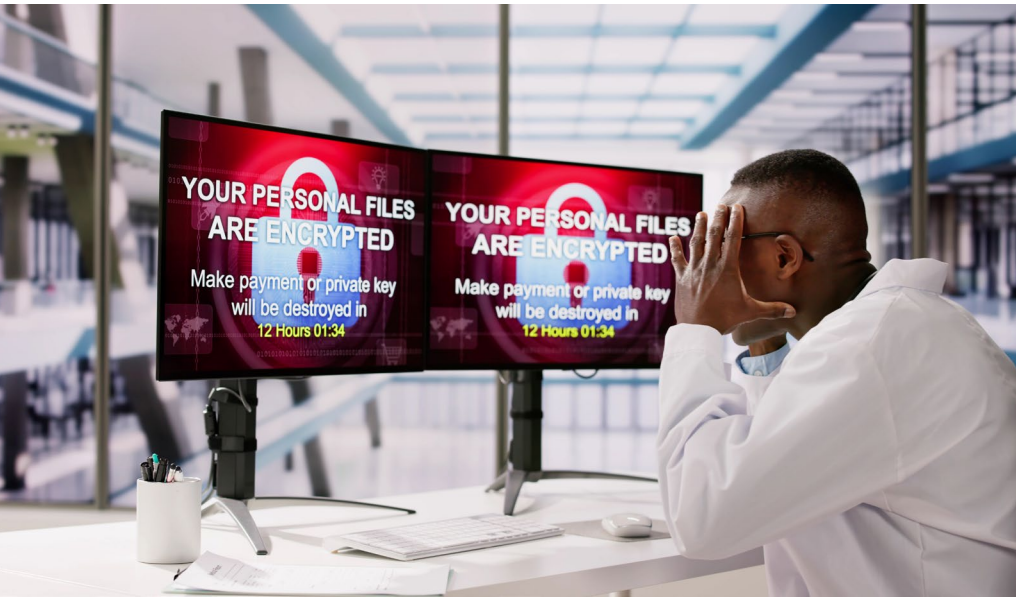
The scale and speed of this change recently came to worldwide attention with the OpenClaw saga. Previously called Clawdbot, and then Moltbot, OpenClaw is an open source AI assistant that integrates with more than 50 popular apps. Once installed, it can access and run its owner's email, social media and messaging apps, all with a breezily simple UX - you just "DM it like a friend."

Sure enough, it quickly became hugely popular. But its huge power was soon revealed as a huge vulnerability. When Ethiack's AI pentesting agent, Hackian, tested OpenClaw's open-source code, it struck gold in less than two hours: it found a 0day, a previously unknown critical vulnerability that a cybercriminal could exploit to take over the account of anyone using OpenClaw - and with it all their connected apps. The security flaw was so serious it was given a 'high severity' score of 8.8 on the CVE register.

What if the weak link in your IT is an AI rather than a human?

Previously cybercriminals often targeted people - whether through phishing emails, malware or just plain human error - as a way to steal credentials and gain access.

But a hostile hacker who seizes control of an AI system like OpenClaw doesn't



need to break in, as the agent comes with a full set of keys.

More worrying still, an AI agent that has been captured by a threat actor is likely to go unnoticed. This is because the conventional wisdom - that users are human - has led many authentication and fraud detection systems to be predicated on how people usually behave. When activity deviates from the 'normal' pattern, red flags are raised and security teams investigate.

Security monitoring relies heavily on behavioural analysis, for instance, flagging unusual login locations, strange working hours or activity inconsistent with a user's history.

But AI agents undermine these assumptions. Identity is no longer tied to a person, behaviour is no longer human, and when an AI agent is compromised, credentials are not stolen.

A captured AI system may operate continuously and at machine speed, processing thousands of actions per hour. Yet malicious prompts, poisoned data sources or compromised third party platforms could lead the agent to cause huge damage while still operating within its authorised parameters.

From the perspective of identity verification and audit logs, everything will appear routine; if no employee account has been compromised, the activity may not be technically unauthorised.

While OpenClaw was created for individuals, an enterprise-level AI agent might have access to finance systems so it can pay invoices, reconcile accounts or manage subscriptions. In doing so, the company effectively creates a privileged operator, equipped with extensive delegated decision-making power - inside its environment. The scale of this threat is set to grow rapidly. Not just because of the increased adoption of AI in multiple business settings, but also because AI systems continuously interact with other automated services.

In the near future, cyberattacks made via a vulnerable AI may target workflows rather than people, with hackers who seize control of one AI agent able to influence multiple systems - all seemingly without making unauthorised access. At this stage, the challenge for cybersecurity teams will be about verifying rogue intent rather than detecting intrusion.

Threat actors are using AI to hack AI systems. The best defence? AI

The cybersecurity front line is seeing an unprecedented AI arms race. A 2025 report by the UK's National Cyber Security Centre concluded that all types of cyber threat actor – state and non-state, skilled and less skilled – are routinely using AI tools to penetrate IT systems.

With AI agents now a key part of many organisations' 'attack surface', AI systems are a crucial focus for cyber defence.

However, existing security models do not fully address this threat. Zero Trust architectures verify identity and device integrity, but a compromised AI agent will sail through those checks if it authenticates correctly and uses approved accounts.

A better telltale to watch for is not authentication, but authority - and whether an action should have been performed by a non-human actor at all.

Organisations need to adapt their security controls for an AI native future. AI agents should be treated as privileged identities, but with minimal permissions. Critical actions such as payments, supplier changes or access provisioning should still require a human validation failsafe.

Security monitoring should focus on what the agent does, not simply whether it is logged in correctly. Integrations between systems should be isolated, and organisations must be able to audit why an automated decision occurred, not just record that it did.

Cybersecurity teams should make use of AI tools as well. For example, Ethiack's Hackian is a hackbot able to continuously scan vast attack surfaces, including AI agents, learning and locating potential weaknesses.

Used ethically, transparently and under human control, next-generation penetration testing can help organisations stay ahead of AI-enabled attackers, by finding vulnerabilities in their AI systems early and closing them fast.

The next wave of cyber incidents may not involve breached networks or stolen passwords. Instead, they could see trusted AI systems doing exactly what they were allowed to do, but not what the organisation intended.

Security has always been about trust. AI agents don't remove that problem; instead they are moving the cybersecurity front line from the organisation's outer wall to its operational core. Businesses are not just deploying software tools anymore, they are introducing non-human operators into the heart of their systems, and security strategies must evolve accordingly.

Preparing for the quantum threat

Protecting organisations against quantum-based threats requires a proactive, scalable strategy that accounts for live traffic, long-lived data and operational continuity.

BY DAVID SPILLANE, SYSTEMS ENGINEERING DIRECTOR, FORTINET

While still in their infancy, quantum computers already pose a significant threat to data security. This is due to their ability to bypass traditional encryption protocols via tactics such as harvest-now, decrypt-later – allowing cybercriminals to access encrypted data previously considered secure.

It is critical organisations begin implementing quantum-safe encryption now, so any harvested data remains secure once quantum computers arrive. This starts with selecting quantum-safe solutions that secure data via encryption.

Yet, with quantum-safe technologies having the ability to affect system performance and infrastructure, organisations need to carefully consider the type of solution they adopt. This means evaluating performance impact, hybrid operation and adherence to formalised industry standards. This will allow organisations to safely implement quantum-safe solutions now while being quantum-ready in future.

Quantum is rapidly evolving

The current landscape for quantum computing is changing at speed, with the UK government previously establishing a 10-year National Quantum Strategy aiming to accelerate the adoption of quantum computing across the economy. This includes committing £670 million towards applications in clean energy, AI and healthcare and developing quantum computers capable of outperforming conventional supercomputers by 2035.

With adoption set to accelerate, the National Cybersecurity Centre (NCSC) has published guidance for organisations looking to transition from current cryptography to quantum-safe encryption. While primarily targeting

high-risk sectors such as finance, energy and telecoms, this advises developing a plan for migration by 2028 and completing the transition across systems and services by 2035.

While the above guidance provides a clear timeline, organisations need to understand how this can be put into practice and what they can do now to prepare.

Becoming quantum-safe

One way organisations can protect themselves is by adopting quantum-safe solutions. These are advanced cryptographic techniques designed to protect against the advanced cybersecurity threats quantum computing poses.

This includes Post-Quantum Cryptography (PQC), which are algorithmic, software-based solutions that use complex mathematical problems to withstand quantum attacks. It also encompasses Quantum Key Distribution (QKD), physical, hardware-based technology that uses quantum mechanics to secure cryptographic keys. But what should organisations be looking for when it comes to selecting quantum-safe solutions?

Choosing the right solution

The first factor to consider is minimal performance impact. Quantum readiness cannot come at the expense of network performance, particularly in enterprise environments where modern architectures such as Software-Defined Wide Area Networks (SD-WAN) demand high throughput and ultra-low latency.

An effective quantum-safe solution must integrate high-performance processing to ensure quantum-safe encryption does not degrade network performance.

Next, assess the solution's ability to work in hybrid mode. It's not hugely practical to have new PQC algorithms working flawlessly across platforms from the onset. Solutions should require the simultaneous use of both a classical algorithm, such as DH, and a PQC algorithm, such as ML-KEM, during a single key exchange. This ensures multi-layered protection should one system fail, while also supporting a smoother, controlled migration for organisations to safely test performance and reliability in a live environment.

> While still in their infancy, quantum computers already pose a significant threat to data security. This is due to their ability to bypass traditional encryption protocols via tactics such as harvest-now, decrypt-later – allowing cybercriminals to access encrypted data previously considered secure

Next, make sure the solution adheres to the algorithms formalised by the National Institute of Standards and Technology (NIST). This includes approved PQC algorithms, such as the Module-Lattice-Based Key-Encapsulation Mechanism Standard (ML-KEM) and Hamming Quasi-Cyclic (HQC). These have ultimately undergone years of public, global scrutiny by cryptographers, helping to guarantee interoperability and compliance.

Finally, an optimal solution should give security teams the flexibility to deploy the right tool for the right job without replacing existing infrastructure. While this means offering both PQC and QKD, their applications differ. PQC is ideal for large-scale, cost-effective deployments across diverse environments, including clouds and data centres. It can also be integrated into existing appliances, firewalls and VPN gateways to secure active traffic. QKD on the other hand is suitable for high-assurance, mission-critical connections, such as securing government or financial networks, where the highest level of assurance is compulsory.

Looking to the future

Protecting organisations against quantum-based threats requires a proactive, scalable strategy that accounts for live traffic, long-lived data and operational continuity. By safely adopting quantum-safe solutions, organisations can remain resilient, secure and protected in the years ahead.

Protecting mission-critical networks from next-generation threats

By combining automation with human expertise and embedding security throughout the architecture, telecom providers can ensure the networks we all depend on remain trusted, resilient and dependable against next-generation threats.

BY FERNANDO RIONEGRO, VICE PRESIDENT, CLOUD AND NETWORK SERVICES, EUROPE AT NOKIA

MOBILE DEVICES now allow people to navigate, communicate and transact from almost anywhere in the world. A simple tap can pay for a coffee, give instant access to public transport or send money to a friend in seconds. Connectivity has become a constant of everyday life.

But mobile infrastructure underpins far more than convenience. What were once more simple communications networks now serve as mission-critical systems, supporting everything from emergency response teams and complex healthcare systems to financial platforms and essential public services, shaping how we live, work and do business.

However, with greater connectivity also comes greater exposure and as networks become more interconnected, the attack surfaces expand. Every interface, protocol and software layer introduces a new potential entry point for attack.

Over the past year alone, telecom providers have faced everything from large-scale espionage campaigns targeting core network infrastructure to data breaches exposing hundreds of thousands of customer records.

The rising number of such incidents underlines a troubling reality: telecom providers are now prime targets for cybercriminals and state-backed groups due to the sensitive data and critical national infrastructure they manage. Against this backdrop, four threats in particular stand out for their speed and impact.

Stealthy campaigns target the telco core

Adversaries are becoming more

coordinated and deliberate in their targeting of telecom networks. Rather than opportunistic strikes, attackers are increasingly focusing on the telco core itself with coordinated, infrastructure-level campaigns. Over the past year, 63% of telecom providers experienced at least one so-called "living-off-the-land" intrusion, while nearly a third reported four or more such attacks.

These attacks reflect a clear shift in strategy. Attackers are no longer looking for quick wins; instead, they are embedding themselves within networks by blending into routine administration, misusing trusted tools and exploiting configuration drift. By operating in ways that appear legitimate, they can move laterally across critical systems, from orchestration layers and mobile core signalling to subscriber databases and lawful interception paths, without triggering traditional security alerts. When legitimate activity becomes the disguise, the telco core itself becomes the attack surface.

The Salt Typhoon campaign is a prime example of this shift. By exploiting long-standing entry points, attackers compromised lawful interception systems. They maintained long-term, privileged access across networks in more than 80 countries, demonstrating just how deeply they can now embed themselves within telecom environments. More recently, the breach at British operator Brsk, where around 230,000 customer records were reportedly stolen and auctioned online, further highlighted the scale of data and trust at stake.

The evolution of DDoS attacks

DDoS attacks no longer resemble slow-building traffic floods. Instead, they strike fast and at scale, overwhelming networks before traditional defences have time to respond. Traffic peaking at 5–10 Tbps has become a daily reality, with 78% of DDoS attacks now ending within five minutes, and 37% wrapping up in under two minutes.

This shift is driven by a more mature and accessible attack ecosystem. Residential proxy networks built on more than 100 million hijacked home devices combined with Mirai-derived botnets, are giving attackers instant access to enormous bandwidth. The result is multi-terabit floods that can be launched and withdrawn within minutes. In this environment, resilience hinges on sub-minute detection and mitigation, ideally triggered across multiple vantage points before the first wave hits.

DDoS attacks no longer resemble slow-building traffic floods. Instead, they strike fast and at scale, overwhelming networks before traditional defences have time to respond.

AI on both sides of the battlefield

AI is shaping both sides of telecom security. Adversaries are using automation and AI to move faster through networks, making phishing and social engineering more convincing, and adapting malware and exploits to telecom-specific systems. The latest threat intelligence data shows that phishing and social engineering remain the leading root cause of major cyber incidents globally, cited in 25.6% of cases, while 55% of telecom providers report malware engineered specifically for telecom protocols with 45.1% encountering custom-built toolkits.

In response, telecom security leaders are increasingly turning to AI to defend against rising stealthy attacks and rapid DDoS campaigns. More than 70% now rely on AI/ML-based threat analytics, including predictive models, instant context and governed automation to strengthen network resilience.

Hidden implants and protocol abuse
Attackers are now pushing deeper into telecom infrastructure, targeting management planes and telco-native protocols that sit at the heart of network operations. These intrusions can remain hidden for long periods of time, often in areas where traditional IT security has blind spots, lying dormant until remote commands activate them.

When attackers gain control at this level, the consequences extend far beyond data theft. Service integrity can be disrupted, recovery becomes more complex, and confidence in the network itself is undermined. In fact, 44.4% of operators rank reputational damage as the most serious consequence of a breach, ahead of both financial loss and technical disruption.

Building resilience into the core
As network threats continue to evolve, resilience will increasingly depend on how effectively security is integrated into every layer of network architecture. Instead of treating security as an overlay, telecom providers must embed protection across every layer, from infrastructure and operations to governance. Achieving this requires adopting continuous monitoring, zero-trust principles, and a security-by-design mindset at the core.

A key part of this approach is the ability to detect early indicators of abnormal behaviour and contain potential threats before services are impacted. Continuous monitoring across core network domains, supported by anomaly detection and trust validation designed specifically for telecom traffic, enables operators to spot subtle changes that may otherwise go unnoticed.

Limiting attacker dwell time is equally important. Closing identity gaps through regular credential rotation, strong authentication for network devices and tighter controls on shared accounts reduces opportunities for persistence and lateral movement, helping contain incidents at an early stage.

AI is becoming a powerful enabler of this shift, supporting faster and more precise detection and response while improving visibility across complex environments. When deployed with clear governance, explainability and human oversight, AI-driven systems can strengthen proactive threat hunting and decision-making without undermining accountability.

By combining automation with human expertise and embedding security throughout the architecture, telecom providers can ensure the networks we all depend on remain trusted, resilient and dependable against next-generation threats.

Why multi-cloud success requires more than connectivity

The future of multi-cloud is a unified, globally connected fabric, not a patchwork of one-off bridges. It will be cloud-agnostic, operationally consistent and cost aware.

BY DMITRY PANENKOV, CEO AND FOUNDER OF EMMA, THE CLOUD MANAGEMENT PLATFORM

AWS and Google Cloud have introduced a multi-cloud interconnect that speeds up private connectivity between their platforms. What once took days to configure can now be set up in minutes, with built-in security and monitoring included. For organisations requiring direct links between these two clouds, this is a practical and welcome improvement. At the same time, Nutanix is pushing forward with sovereign cloud strategies, helping businesses meet data residency and compliance requirements while still preserving flexibility across diverse environments.

While these advancements mark visible multi-cloud progress, they only scratch the surface of what's truly needed. Multi-cloud success isn't about simply connecting providers together, but about creating a unified, seamless ecosystem. Until hyperscalers tackle portability and large-scale management, multi-cloud will remain fragmented. Announcements like these, while useful, feel more like incremental steps than groundbreaking progress.

Connectivity is just the starting point

Connecting multiple providers is only the first step in a multi-cloud strategy. The real challenge lies in delivering consistent, scalable operations across increasingly complex multi-cloud environments – where gaps in governance, limited workload portability, cost optimisation challenges and operational complexity continue to slow progress.

These announcements address just one piece of a much larger puzzle, and they arrive years after enterprises began calling for real interoperability.

True multi-cloud isn't about stitching together isolated bridges between providers. It's about enabling a cohesive, cloud-agnostic ecosystem in which workloads can move seamlessly based on business requirements, not platform constraints.

Today's enterprises aren't building a single bridge; they're building a whole system of roads – a global network spanning AWS, Azure, Google Cloud, sovereign clouds, on-prem and edge environments. In this landscape, the most significant risks emerge during day-two operations: policy drift across multiple clouds, segmentation gaps, observability blind spots and unpredictable costs that can quickly

derail budgets. A bilateral interconnect, no matter how fast or secure, does little to address these systemic, operational challenges.

Why a standardised foundation matters

Hyperscaler solutions often treat multi-cloud as a set of point-to-point connections. That approach doesn't scale. Every new region, provider or service adds complexity. Without a standardised global foundation, teams end up managing a patchwork of environments with inconsistent policies and uneven security.

A multi-cloud fabric addresses this challenge by providing a global control layer that enforces connectivity, governance and observability across all environments. Policies are defined once and apply universally, ensuring compliance and performance remain consistent across all clouds. This is the difference between operational chaos and operational consistency.

Without this unified layer, organisations face operational fragmentation: inconsistent policies across clouds, uneven security postures and complex troubleshooting across regions. A globally enabled fabric simplifies operations, reduces risk and creates a repeatable model for scaling multi-cloud deployments.

The hidden barrier: Multi-cloud economics

Speed is important but cost often decides whether a multi-cloud strategy ultimately succeeds. Moving data between clouds is expensive, and egress fees combined with data gravity can quickly turn even well-architected designs into financial dead ends.

Connectivity is no longer the bottleneck. The real test is operating multiple clouds together securely, efficiently and sustainably, while still delivering the agility and resilience enterprises expect

For multi-cloud to be viable, predictable cost structures are essential. Without them, common patterns such as cross-cloud disaster recovery or active-active resilience may be technically feasible but financially impractical. Leaders must plan for these costs early, not after architectures are already in place. By embedding cost analytics and data-placement strategies from the outset, enterprises can avoid financial traps and ensure long-term sustainability.

Beyond connectivity: Operational consistency is the real differentiator

Hyperscalers productising bilateral interconnects is meaningful progress, but the real leap lies in moving beyond pairwise connectivity features to repeatable multi-cloud operations. Connecting two platforms is impressive, but long-term differentiation will come from the ability to run workloads consistently across all clouds.

True multi-cloud maturity means embedding governance, observability and cost predictability into the operating model from day one. Organisations need frameworks that enable them to deploy, monitor and optimise workloads across providers without introducing inconsistency or financial risk. These capabilities are what transform multi-cloud from an experiment into a practical, enterprise-ready strategy.

The future: From bridges to global digital fabric

The future of multi-cloud is a unified, globally connected fabric, not a patchwork of one-off bridges. It will be cloud-agnostic, operationally consistent and cost aware.

Today's collaborations between hyperscalers and platforms like Nutanix are an important step, but they're only the beginning. The next phase of innovation will focus on integrating consistent governance, security and cost management across every environment.

Connectivity is no longer the bottleneck. The real test is operating multiple clouds together securely, efficiently and sustainably, while still delivering the agility and resilience enterprises expect.

Sovereignty is no longer about location

Sovereignty has become one of the most frequently used terms in European technology conversations, yet it is often defined too narrowly. For some, it still means data residency. For others, it means hosting in a sovereign cloud region. In practice, neither definition is sufficient.

BY LEONARDO BOSCARO, EMEA SALES LEADER, NUTANIX DATABASE

ACROSS EUROPE, regulatory frameworks such as the EU's Digital Operational Resilience Act are now in force, placing explicit obligations on financial institutions to demonstrate resilience, test critical services and manage third-party dependency risk. The EU Data Act introduces clearer requirements for data portability and provider switching. In parallel, national governments continue to invest in sovereign cloud initiatives, while geopolitical tensions and cross-border data access debates remain part of the strategic landscape.

These developments have changed the nature of the sovereignty conversation. Today's organisations do not simply ask where data resides. They increasingly ask who controls the operating model, how easily workloads can be moved, and whether operational governance remains consistent if commercial or regulatory conditions change.

Beyond geography

Early sovereignty debates focused primarily on data location. Keeping data within national or regional boundaries was seen as the primary safeguard against external jurisdictional risk. While location remains important, it does not resolve the deeper issue of dependency.

An organisation may run workloads in a sovereign region, yet still be tightly coupled to a single provider's tooling, lifecycle constructs and operational workflows. In such situations portability exists in theory but becomes difficult in practice. Controls must be rebuilt, recovery processes revalidated, and governance models adjusted each time infrastructure choices evolve. Many large organisations have responded by building internal automation frameworks designed to run workloads consistently across on-premises and cloud environments.

These initiatives can successfully introduce infrastructure portability. However, operational complexity often reappears at the data layer, where database provisioning, patching and recovery processes remain dependent on environment-specific tooling or internally maintained scripts.

True sovereignty requires more than regional hosting or infrastructural abstraction. It requires control of the data operating model itself.

Regulation is redefining dependency DORA, for example, goes beyond traditional availability metrics. It requires financial institutions to assess and manage concentration risk, including reliance on critical third-party providers. Supervisory authorities are increasingly focused on whether organisations can demonstrate resilience independently of any single infrastructure provider.

Similarly, the EU Data Act introduces measures designed to make switching between cloud providers more feasible over time. While implementation will evolve, the direction is clear: policymakers expect greater flexibility and reduced lock-in.

These frameworks do not instruct organisations to abandon public cloud. They do, however, raise expectations

around operational autonomy. Leaders must demonstrate that resilience, governance, and recovery processes are not dependent on proprietary architectures that cannot be replicated elsewhere.

The data operating model as the control point

This is where the conversation moves from infrastructure to operating model. When database provisioning, lifecycle management and recovery are governed through a consistent data operating model, sovereignty becomes operational rather than theoretical.

Workloads may already move between environments, but sovereignty is tested when data platforms must be recovered, audited or re-deployed under regulatory pressure. If database operations differ across infrastructures, governance must effectively be rebuilt each time environments change.

A consistent database operating model allows organisations to apply lifecycle policies, guardrails and recoverability standards once and enforce them uniformly across on-premises and public cloud environments. Infrastructure becomes a deployment choice rather than a governance constraint.

Without this layer of standardisation, hybrid strategies can increase complexity rather than reduce risk. Dependencies do not disappear; they simply shift from external providers to internally developed automation frameworks that require continuous maintenance and specialised expertise.

Open source without operational sovereignty

The growing adoption of open-source databases across Europe reflects another important dimension of the sovereignty discussion. Organisations increasingly select open technologies to reduce dependency on proprietary platforms, licensing constraints and provider-specific services.

However, open-source adoption alone does not automatically deliver sovereignty. When each environment requires different provisioning models, upgrade procedures or recovery practices, operational dependency persists.

Complexity moves from the provider to internal platform teams responsible for maintaining automation and operational consistency.

For highly mature organisations, internally developed database automation can provide flexibility. For many others, industrialising these capabilities across multiple environments proves difficult to sustain over time.

Sovereignty therefore depends not only on open technology choices, but on the existence of a consistent database operating model capable of delivering lifecycle automation, governance and recoverability out-of-the-box across infrastructures.

Hybrid without fragmentation

Many EMEA organisations will continue to operate in hybrid and multicloud environments. Public cloud delivers elasticity and access to innovation while On-premises environments provide control, proximity and in some cases regulatory assurance. The strategic objective is not to favour one over the other, but to operate coherently across both.

This coherence depends on whether the data operating model travels with the workload. If provisioning, patching and recovery processes differ materially between environments, operational risk increases. When these processes are standardised through a common operational layer, consistency is preserved even as infrastructure evolves.

In this context, sovereignty becomes the ability to adapt without rebuilding governance from scratch. It is the confidence that regulatory scrutiny, commercial renegotiation or geopolitical shifts will not force a complete redesign of database operations.

Commercial and geopolitical resilience

Recent global events have shown how quickly commercial and geopolitical conditions can evolve, with licensing models changing, provider strategies shifting and regulatory expectations tightening as concentration risk becomes an increasing supervisory focus.

Organisations that treat sovereignty purely as a hosting decision may find themselves reacting to these changes. Those who treat sovereignty as operational autonomy are better positioned.

By controlling the data operating model, they retain flexibility and can consolidate, migrate or rebalance workloads while preserving consistent governance and recovery discipline.

Independent analysis from Forrester's Total Economic Impact study supports what many organisations are already recognising in practice. When database operations are standardised within a unified operational layer, resilience improves and operational friction declines. The result is not only efficiency, but greater control over how and where critical services run.

Sovereignty as leadership discipline

For today's CIOs, CTOs and CISOs, sovereignty ultimately means retaining operational control, regardless of where workloads run. It requires ensuring that governance, recovery and lifecycle management remain consistent even as infrastructure strategies evolve.

In regulated environments, credibility depends on evidence. Leaders must demonstrate that resilience testing, recovery execution and governance controls remain reproducible across infrastructure choices.

That consistency is achieved not through geography alone, but through disciplined ownership of the data operating model.

Sovereignty, in this sense, is operational autonomy. It is the ability to make infrastructure decisions without compromising control, compliance or recoverability.

Expanding network reach beyond borders with remote peering

The shift from direct to remote peering with a specialist partner provides a fast, flexible and cost-efficient way to grow on a global scale, without limitations.

BY MARK DALEY, DIRECTOR OF DIGITAL STRATEGY AND BUSINESS DEVELOPMENT AT EPSILON TELECOMMUNICATIONS

DIGITAL ECONOMIES across the globe are expanding at a rapid pace, powered by AI, e-commerce, digital banking, fintech, and other mission-critical business applications that rely on the Internet for performance. At the same time, rising consumer demand for streaming, gaming and social media is accelerating content delivery requirements, driving significant growth in Internet traffic. According to TeleGeography, international bandwidth demand surpassed 6.4 Pbps in 2024 – triple the demand of 2020, which was already at an all-time-high due to the pandemic.

This growth places significant pressure on service providers as they evolve their networks to keep pace with digital transformation. It also impacts enterprise connectivity with the growth of cloud services, edge computing, SaaS platforms, real-time applications and distributed workforces. Consistent, low-latency connections to partners, platforms, and services are now crucial for both service providers and enterprises alike, while ensuring bandwidth can support increasingly data-intensive applications and services.

Connectivity to cloud providers, content delivery networks, and other global partners and platforms comes with the challenge of balancing performance, cost, and global reach. Service providers need to efficiently manage growing traffic, quickly scale network reach regionally or globally, and maintain SLAs for multiple customers, all while controlling operational costs. Legacy network infrastructure is no longer enough to juggle this growing set of requirements across multiple markets.

Exchanging traffic at IXPs

By peering and exchanging traffic at Internet Exchange Points (IXPs), organisations can gain more control over how data flows across networks, while improving performance and reducing transit costs. This approach is increasingly important for meeting changing connectivity demands.

Peering is not a new concept for many service providers, who have already been using it to extend their reach on a global scale. However, it has seen substantial growth in recent years thanks to surging traffic demands. As of November 2025, there are 1,019+ active IXPs globally – a growth of almost 53% in the last five years (Internet Society Pulse).

Peering is also gaining popularity among enterprises looking to enhance their digital transformation strategies, particularly in high-demand sectors like finance, gaming, broadcasting, research, and travel.

Peering itself has changed a lot over the years. Traditional 'direct' peering was typically used for large networks with heavy traffic at specific exchange points, requiring direct connections and the physical deployment of infrastructure such as routers and cross-connects at the IXP. This delivers

strong performance, but comes with large costs, long setup times, and operational complexities.

Today, remote peering uses a third-party provider's infrastructure to provide a virtual presence at one or more IXPs. The chosen provider carries traffic from the organisation's point of presence (PoP) to the IXP over its own backbone, which takes away the difficulty of installing and managing physical hardware locally.

While direct peering takes weeks or months of time for installation, plus upfront investment for on-site equipment, remote peering is totally virtual with a service-based model, which translates into cost reduction and a much faster setup process.

Simplifying peering with partnerships

Enterprises and service providers can utilise an expert partner to streamline their peering journey. A remote peering partner can provide extensive global reach, and direct interconnectivity with leading global IXs via a selection of on-ramp locations.

This delivers on-demand access to a wide-reaching network of peering members.

Partnering for remote peering is also highly cost-efficient. Customers pay only for the required bandwidth and connections with short contract terms, and virtual presence removes the need for large upfront infrastructure investments.

Preferred peering partners can also be accessed through a single interconnection port at a convenient location via an experienced peering partner, dramatically reducing deployment times and costs. Multiple IXPs can be accessed remotely and on-demand, with the flexibility to scale bandwidth according to demands.

Remote peering is also a great way to enhance user experiences, delivering optimal performance and ultra-low-latency connectivity for bandwidth-sensitive applications. An expert partner can handle all IX memberships and onboarding to simplify management, with one single contract for connectivity and peering services.

These benefits enable service providers to deliver greater agility and an improved customer experience. For enterprises, they enable global competitiveness and accelerated digital transformation, without physical network constraints.

Local presence. global reach

The continued growth in cloud-first strategies, global internet traffic and AI workloads makes it impossible for many organisations to deploy and manage physical infrastructure across multiple global markets. Remote peering takes away those hurdles, helping enterprises and service providers to scale and respond to changing demands quickly, and without soaring upfront costs.

Service providers and enterprises cannot be held back by the cost and complexity of physical infrastructure in today's digital economy, where seamless, high-performance connectivity across markets is now an expectation. The shift from direct to remote peering with a specialist partner provides a fast, flexible and cost-efficient way to grow on a global scale, without limitations.

By peering and exchanging traffic at Internet Exchange Points (IXPs), organisations can gain more control over how data flows across networks, while improving performance and reducing transit costs. This approach is increasingly important for meeting changing connectivity demands

When boards demand AI ROI, network resilience becomes a governance issue

In a period of CFO scrutiny, the advantage will go to organisations that treat network resilience as a governance priority rather than an engineering afterthought.

BY JEAN PHILIPPE AVELANGE, CIO EXPEREO

BOARDS are not stepping away from AI, but they are tightening the rules, demanding clearer business cases and shorter payback windows. CFO scrutiny is rising, and proof of value is becoming the standard for continued funding. That proof is less about promise and more about repeatable outcomes in daily operations.

The problem is that many companies still treat AI initiatives as standalone tools. In practice, AI is a system with dependencies across data, cloud services, identity, applications, and networks. When outcomes fall short, the failure is typically assigned to the model, even when the real cause sits elsewhere in the technology stack, most frequently, in connectivity and performance.

Recent industry research highlights the tension behind this shift. Technology leaders report that their boards hold unrealistic expectations about how quickly new technology should translate into business performance. That gap between expectation and delivery is exactly where misdiagnosis begins.

Many “AI failures” will be performance failures in disguise

In a production setting, the model is only one component of the user experience. An AI assistant that takes too long to respond or behaves inconsistently across regions is rejected regardless of model quality. ROI then disappears due to lower adoption, longer cycle times, and a growing sense that the technology is unreliable, eroding the business case before the model itself is ever properly tested.

The dependency is already visible in enterprise environments. Organisations consistently report that their networks limit their ability to run large data and AI projects. The biggest constraints are not obscure technical issues, but fundamentals such as networks that cannot scale capacity on demand, and inconsistent application responsiveness caused by latency and performance variation.

The pattern is already clear: as AI expands beyond isolated pilots, the question shifts from whether a model works to whether the service performs consistently everywhere it is needed. In that reality, latency and jitter stop being technical concerns and become business issues, shaping whether AI accelerates a workflow or quietly slows it down.

To prove value under CFO scrutiny, leaders must separate three questions: is the use case valid, is the model performing as expected, and is the operating environment capable of delivering the benefit. Without that distinction, investments hinge on anecdotes, and projects stall for reasons unrelated to AI performance itself.

A practical response is to define performance budgets for each AI-enabled workflow. This means agreeing upfront on acceptable end-to-end response times, the amount of variation that can be tolerated, and how performance should hold across regions and sites. It also means

deciding what will be measured and who owns remediation when the system drifts.

Many organisations are already prioritising networking and connectivity investments ahead of other technology categories because they can see where bottlenecks are emerging. In 2026, the ROI gap will widen between those that address these constraints now and those that wait until performance problems become visible in board metrics.

Resilience moves to the boardroom

Resilience, similarly, now carries clear financial consequences and needs to be treated as board-level material. Connectivity has long been viewed as operational plumbing, but that framing breaks down once AI is embedded into customer journeys, supply chains, and internal decision-making. When performance degrades, business impact is immediate. Research shows organisations reporting financial losses ranging from hundreds of thousands to millions annually from network-related downtime and performance degradation; these costs compound when AI systems are affected. That level of exposure belongs in board risk discussions, not buried in operational reviews.

Network readiness also shapes the probability of harm. Fewer than half of organisations believe their networks are fully ready to support new technology initiatives, and scaling AI on top of stretched infrastructure increases the risk of chronic underperformance - slow enough to erode value, but not dramatic enough to trigger fast intervention. To address this, governance needs to link performance metrics directly to business outcomes. Traditional uptime targets are no longer enough because users can experience “availability” alongside delays or inconsistencies.

Boards should push for controls that capture the full transaction path, not just the health of isolated components. That begins with observability, the ability to see how an AI-driven transaction behaves across applications, cloud services, and network routes.

Without it, incidents become slow and political, with each party defending its own metrics. With it, teams can quickly identify whether issues stem from model throughput, data access, cloud congestion, routing, or local conditions. Testing needs to change. AI workloads are spiky and often unpredictable, especially when tied to automation. Resilience testing in 2026 should routinely include degraded routing, regional congestion, cloud-dependency disruptions, and sudden demand surges, as these are the conditions that undermine user trust and ROI.

Accountability must follow the transaction

As deployments span sites, clouds, and partners, accountability frequently fragments. Each supplier may meet its own targets while the overall user experience still fails. With networking skills scarce and many organisations relying heavily on partners, clear accountability models that follow the transaction end-to-end are essential. Shared objectives, aligned measurement approaches, and defined escalation paths across organisational boundaries are no longer optional.

The test is simple: can the organisation trust AI in production? ROI will depend on delivery discipline, not model sophistication. Leaders should define performance budgets, build end-to-end observability, make resilience testing a standard gate for scale, and align accountability across teams and partners.

In a period of CFO scrutiny, the advantage will go to organisations that treat network resilience as a governance priority rather than an engineering afterthought. The winners will not be those deploying the most AI, but those making it perform reliably wherever the business depends on it.

As deployments span sites, clouds, and partners, accountability frequently fragments. Each supplier may meet its own targets while the overall user experience still fails. With networking skills scarce and many organisations relying heavily on partners, clear accountability models that follow the transaction end-to-end are essential

2026: The year networks take control

AI needs the right network to deliver value, and networks need AI to operate at the speed, scale, and intelligence that modern business demands.

BY MARKUS NISPEL, HEAD OF AI ENGINEERING & EMEA CTO, EXTREME NETWORKS

THE "AI for AI's sake" phase is over. After years of pilots and proof-of-concepts, enterprises headed into 2026 with a reality check: only about one-third of companies have scaled AI beyond experiments. For most, the bottleneck isn't ambition, it's other internal systems that aren't built to scale alongside AI. Networks built for yesterday's workloads with siloed data simply can't keep up with the speed, scale, and adaptability that AI-driven environments now demand.

AI has moved from experiment to operational core. Enterprise networks supporting AI deployments can no longer be reactive; they need to think ahead, adapt in real time, and operate at a scale far beyond what humans are capable. As a result, the entire enterprise network is being fundamentally redefined – opening the door to faster decisions, seamless collaboration, and accelerated innovation across the business.

Networks that will not wait

In the future, autonomous networking won't wait for engineers to identify bottlenecks or diagnose failures. Instead, AI will proactively predict disruptions before they cascade across the enterprise, reconfigure traffic patterns instantly based on changing demand and make thousands of optimisation decisions over its lifetime. It does all this while maintaining strong controls – ensuring human oversight where it is required or desired, and embedding governance, explainability, and auditability so that automated decisions can be understood, trusted, traced and audited.

To achieve this goal, multi-agent systems will be required to manage the complex tasks involved in the entire network lifecycle, from planning to optimisation. Different agents handle very specific tasks and are orchestrated by planning agents that understand user intent and translate this into an execution plan, invoking those specialised agents much like a team manager assigns tasks to experts on their team.

As trust grows through continuous evaluation of system accuracy, controls can be gradually relaxed, allowing processes to move faster and execute changes while human involvement shifts from being "in the loop" to being "on the loop", focusing on supervision and freeing up time for proactive work and strategic tasks. The network learns and adapts on its own, while users oversee it using their expertise.

The real question isn't whether this is coming; it's how prepared your network infrastructure, processes and employees are to take full advantage of it. Thinking about how autonomous driving is evolving offers a useful parallel, not just in terms of technology, but also human acceptance and the willingness to embrace a new way of working and living.

The security equation changes

So, here's where it gets complex. As networks grow and support new AI workloads, they also become more populated with non-human identities. Automated workflows driven by agents across accounting, marketing, engineering, security, and network operations create identities that need

access, permissions, and monitoring. These systems make decisions, interact with other agents, and adapt their behaviour based on outcomes, but traditional security frameworks weren't built for entities that act and learn on their own. Identity and access security must evolve to keep pace.

AI agents need identity- and access-based controls that grant permissions based on purpose, data sensitivity, and context, treating them like any other user with the same or stricter controls. Continuous verification ensures that every interaction is validated, keeping agents accountable and auditable.

To enforce these controls effectively, organisations need full network visibility, which is where network fabric comes in. By mapping traffic and system connections in real time, organisations can use microsegmentation, isolating sensitive systems and keeping a compromised agent from moving laterally or accessing unrelated applications. This means separating operational systems from payment environments, isolating IoT devices from core applications, and containing breaches before they can spread.

Integrated platforms that combine Zero Trust Network Access, cloud NAC, and AI-powered threat detection tie everything together with robust identity and access management systems. By linking identity controls, continuous verification, network visibility, and segmentation, organisations can safely manage AI agents, turning autonomous systems from a potential risk into a secure, high-performing part of the network.

Where it gets real

Look at retail operations in 2026, where AI promises hyper-personalised shopping experiences. Behind every tailored recommendation lies a complex web of IoT sensors, cameras, mobile apps, edge devices and AI platforms. Through each one flows real-time inventory data, customer preferences and behavioural analytics. For example, electronic shelf labels, RFID systems and automated checkouts all need network access, making every connection point a potential vulnerability.

Healthcare is moving into a far more connected future as well. AI will support diagnostics, predictive alerts, and even robotic-assisted procedures, with patient data flowing continuously across EMR systems, monitoring devices, and AI analysis platforms. From wearables to surgical robots, the network becomes a critical pathway, and a point of exposure if networks and access controls aren't designed for it.

In manufacturing, this evolution plays out on an industrial scale, with AI predicting equipment failures, optimising production lines, and coordinating autonomous robots as sensors, industrial controllers, and AI-driven machinery communicate nonstop, multiplying connected endpoints and raising the stakes if networks aren't designed to support distributed AI safely.

> As trust grows through continuous evaluation of system accuracy, controls can be gradually relaxed, allowing processes to move faster and execute changes while human involvement shifts from being "in the loop" to being "on the loop"

Across industries, more automation and autonomy also means more risk. Without a unified enterprise network, strong access controls, and clearly defined human checkpoints, a single misconfigured AI agent could expose sensitive data, disrupt operations, or even put lives at risk. Security and AI-ready networks must be built in from the start to make innovation safe in 2026.

The foundation for autonomy

This brings us to the real challenge enterprises face in 2026: you can't bolt autonomous AI onto outdated, fragmented enterprise networks and expect it to deliver.

Enterprise networks need strong, scalable foundations that can handle these demands with minimal latency. They require edge computing close to where decisions are made, as well as high-bandwidth connectivity that won't collapse under IoT sensor loads or the strain of real-time analytics.

Most importantly, they need network infrastructure built for continuous learning – systems that can take in operational data, identify patterns and refine their models without slowing down. Organisations moving into this need to ask hard questions. Can your current network handle distributed AI agents running at the same time? Do you have enough compute power at the edge to process decisions locally, or are you still routing everything through centralised clouds? Is your security built for thousands of non-human identities or are you still thinking in terms of user credentials?

But the relationship goes both ways. Just as strong networks are critical for AI, AI is becoming essential for networks. Intelligent automation helps networks maintain performance, resilience, and security at scale. Full visibility, centralised data, and AI-driven optimisation ensure that networks not only support enterprise operations but continuously improve themselves, with humans guiding where and how autonomy expands.

In 2026, the winners will be enterprises that understand this symbiosis: AI needs the right network to deliver value, and networks need AI to operate at the speed, scale, and intelligence that modern business demands.

Applied AI, not aspirational AI

The focus this year is on applied AI – systems that deliver real results and actual business value. Impressive demos and speculative use cases don't cut it anymore.

What matters now is production-grade AI that operates safely at enterprise scale.
For businesses willing to prioritise overhauling outdated enterprise networks, the opportunity is massive. AI unlocks entirely new business model horizons, closing competitive gaps faster than manual operations ever could, that is, but only if the foundation is there to support it.

2026 is the year that AI will scale in production. The question is: is your organisation ready?

When less is more: why small language models deserve a bigger role in enterprise

AI has become central to how organisations improve their customer experience and operational performance. While large language models (LLMs) have proved their value across many enterprise use cases, their scale, cost and complexity mean they are not necessarily the right answer to every problem.

BY PRASAD SANKARAN, PRESIDENT, SOFTWARE AND PLATFORM ENGINEERING AT COGNIZANT

SMALL LANGUAGE models (SLMs), particularly those trained on proprietary enterprise data, offer a compelling alternative. They enable organisations to build AI solutions that are differentiated while being more sustainable, easier to govern and better aligned with regulatory expectations. Not only that, they are more cost-effective to run and often more accurate for focused tasks, making them a practical way to accelerate AI adoption without overengineering.

That does not mean that SLMs will replace LLMs. Instead, it is about recognising that different models suit different needs. In practice, this often means using smaller, domain-specific models to fine-tune AI for particular functions, workflows or decisions. By embedding domain knowledge directly into the model, organisations can deliver more precise and business-relevant outcomes, without sacrificing the flexibility of larger models elsewhere.

The many benefits of SLMs

One of the clearest advantages of SLMs is how well they support privacy-sensitive tasks. Their smaller size and lower compute requirements mean they can be deployed on local infrastructure or private servers, rather than relying on external cloud providers. This reduces the need to move sensitive data outside the organisation, lowering the risk of exposure and giving teams greater control over access and usage.

For highly regulated sectors such as healthcare, financial services and government, where confidentiality is essential, SLMs can be a smart alternative to larger, cloud-dependent LLMs.

SLMs also offer a more sustainable approach to AI. As AI workloads grow, large models are placing increasing strain on energy and water resources, with training alone consuming vast amounts of electricity. Smaller, task-

One of the clearest advantages of SLMs is how well they support privacy-sensitive tasks. Their smaller size and lower compute requirements mean they can be deployed on local infrastructure or private servers, rather than relying on external cloud providers. This reduces the need to move sensitive data outside the organisation, lowering the risk of exposure and giving teams greater control over access and usage

specific models provide a far more efficient alternative. Research from UNESCO and UCL shows that SLMs can reduce energy consumption by up to 90% without sacrificing performance, thanks to their lower parameter counts and reduced compute requirements.

Finally, governance is another area where smaller models stand out. SLMs are easier to audit, monitor and explain, making it simpler for organisations to meet regulatory requirements such as Europe's GDPR and HIPAA in the US. Because they can be trained for specific tasks, SLMs also allow organisations to embed their own policies and controls directly into model behaviour, while benefiting from lower training costs, reduced hardware demands and improved accuracy on focused datasets.

In addition to these clear wins, SLMs bring a host of technical benefits that all organisations can appreciate: lower training and equipment costs, for example, as well as accuracy when trained on focused datasets.

Do all these check marks for SLMs mean we should throw out LLMs? Absolutely not.

The case for a hybrid approach

A hybrid, multi-model strategy brings together the strengths of both model types. LLMs remain well suited to complex, open-ended tasks that require broad contextual understanding, while SLMs excel at narrow, clearly defined problems. Used together, they allow organisations to optimise performance, control costs and reduce environmental impact.

As enterprises scale their AI programmes, these trade-offs are becoming more visible. Sharing proprietary data with third-party LLM providers may feel excessive for simple tasks, while hosting large models internally is costly and can quickly undermine return on investment.

At the same time, sustainability commitments are harder to maintain as AI workloads grow. Many organisations are also discovering that some of their most valuable use cases are narrow in scope but critical to the business, making them ill-suited to general purpose models.
This is where SLMs add real value. When blended thoughtfully with LLMs, they provide a more focused and efficient way to address these challenges.

Making SLMs work in practice

Successfully deploying SLMs requires careful planning across the full AI lifecycle. Access to high-quality, appropriately sized datasets is essential, particularly when tuning models for domain-specific use cases. Strong data and model operations are equally important to ensure they remain accurate, relevant and aligned with changing business needs.

Choosing the right model for each task is also essential. SLMs perform best in focused domains, while LLMs are better suited to broader or more context-rich applications. A hybrid approach allows organisations to match each model type to the problem at hand.

Effective orchestration is the final piece of the puzzle. Organisations running both SLMs and LLMs need intelligent routing mechanisms that determine how each query should be handled. Deciding whether a request is best served by a specialised SLM or a general-purpose LLM is key to delivering consistent, high-quality AI experiences.

Small but mighty

SLMs offer organisations a practical way to begin scaling enterprise AI. They deliver faster, safer and more cost-efficient performance, while supporting sustainability and responsible AI goals. For business and technology leaders beginning to see the limits of an LLM-only strategy, a hybrid approach that combines the strengths of both model types may prove to be the smarter path forward.

The enterprise GenAI dilemma: build or buy?

As generative AI cements itself within business strategy, the build-versus-buy dilemma becomes less about the technology and more about prioritisation.

BY CHRIS ACKERSON, SVP OF PRODUCT AT ALPHASENSE

GENERATIVE AI (GenAI) is quickly changing how businesses operate and scale. Three years since the introduction of GenAI tools, nearly 90% of organisations are now regularly using them.

However, with investments growing exponentially, many organisations are betting on seeing a return on investment this year. At the same time, concerns about lack of return on investment from AI are persistent, making it even more important for enterprises to think carefully about how they adopt and integrate these solutions to ensure tangible results are delivered.

With GenAI at the core of many enterprise workflows, including customer service, content generation and financial analysis, executives are left with a choice – do they build their own GenAI solutions, or do they buy off-the-shelf models?

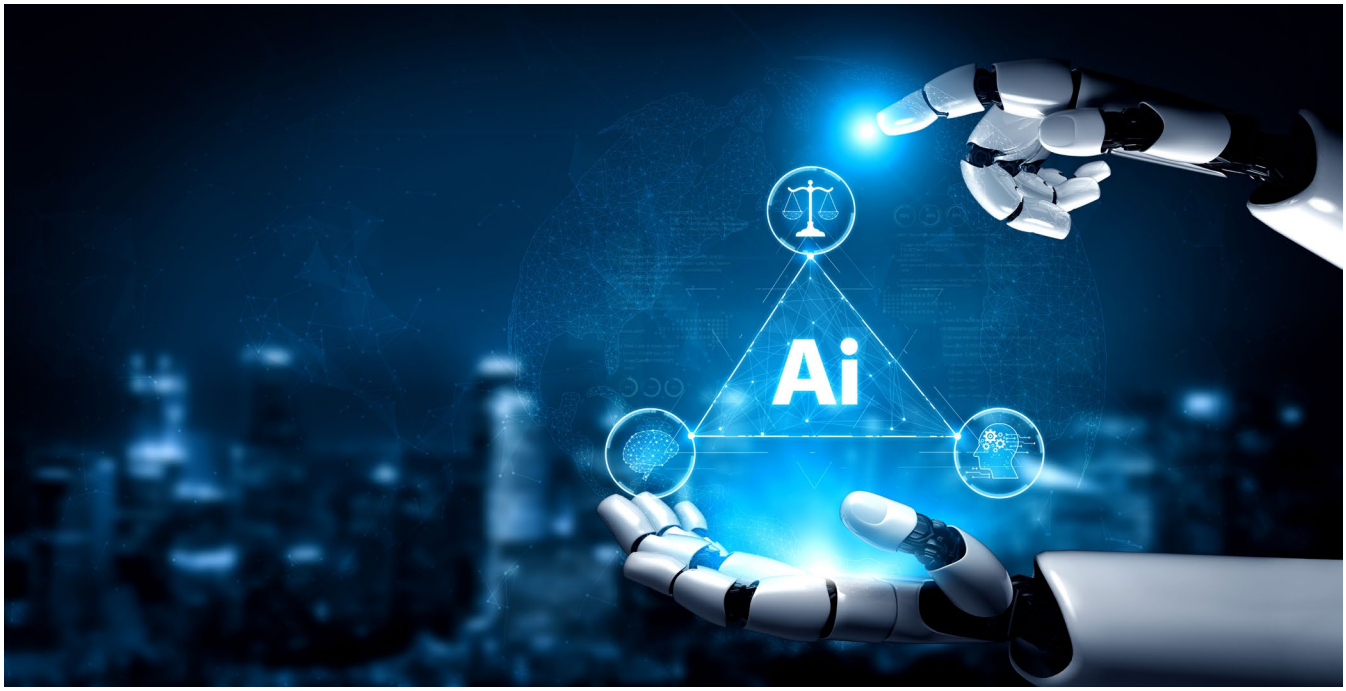
There are clear upsides and downsides to both approaches. The right path depends on budgets, long-term ambitions, and an organisation's appetite for complexity and patience.

The argument for off-the-shelf AI

For many organisations, buying GenAI is the simplest place to start. Off-the-shelf tools offer speed, cost efficiency and far less operational complexity.

In practice, this often means using GenAI to support customer service teams, automate the summarisation of reports and meetings, or improve how employees find and use information across internal systems.

These models are pre-trained, tested at scale, and designed for ease of implementation, while most also boast proven performance. Buying also offers access to continuous innovation as vendors can push out updates and improvements faster than most in-house teams can manage.



Yet, limitations are becoming more apparent as enterprise use cases mature. Pre-trained models are typically designed for the average user, rather than the edge cases or proprietary requirements of specific industries or highly regulated environments.

Financial services offer a clear example. Banks and investment firms often need GenAI systems to work with sensitive, proprietary data, apply strict compliance rules and produce outputs that are auditable and explainable. Off-the-shelf models, optimised for broad applicability, can struggle to meet these demands when workflows deviate from standard patterns or require deep domain context.

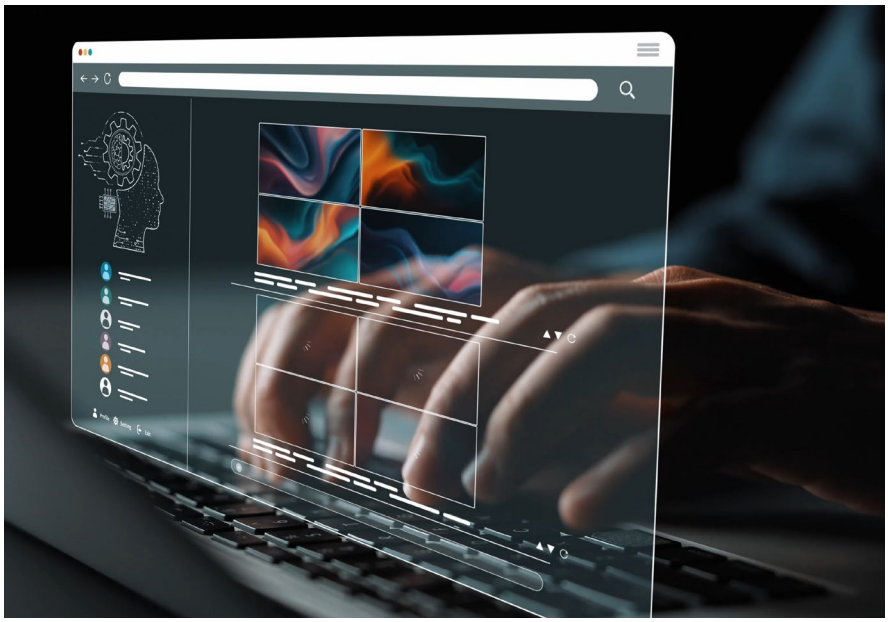
In these scenarios, generic tools may fall short of supporting end-to-end operations smoothly, pushing organisations to consider building or customising GenAI solutions.

There is also the issue of data privacy and vendor lock in. Many GenAI models operate as black boxes, requiring data to be sent off-premises, which introduces concerns around security and compliance. The more important GenAI becomes to your operations, the more exposed you could be to licensing costs, dependency and widespread security risks.

When building in-house makes sense

By contrast, building your own GenAI solution helps guarantee it will have the nuanced, custom functionalities and features your company needs. A custom-built model can be tailored to your data, domain and workflows, integrating with existing systems and giving engineering teams the ability to iterate and improve the model over time. This can offer a competitive advantage while ensuring the privacy of your data.

But the costs are steep. Building an LLM model in house ranges from at least $1 to $2 million in the first year, with additional costs for maintenance, storage and updates making the overall cost close to the multi-million mark annually. And as AI capabilities continue to evolve rapidly, the requirement for regular updates, retraining and optimisation is only set to increase. While a proprietary GenAI model might be a game-changer, it is inevitably a massive investment, not to mention the longer time to market and the risk of failure. It also requires the relevant team of skilled members to get the job done. Part of the reason building in-house is so expensive is the cost of the specialised talent needed to make these models work at scale. The high-end GPUs used to train and run large models are scarce and costly, while the AI engineers, machine learning specialists and data scientists capable of delivering production-grade systems command premium salaries.



These pressures are being compounded by a widening technical skills gap across engineering. Demand for experienced AI and machine learning talent continues to outstrip supply, making teams difficult to hire and even harder to retain. According to a data analysis by LinkedIn, the average time to hire an engineer is now 49 days, longer than in many other professions, including finance, IT and healthcare, slowing progress and adding further cost to in-house GenAI initiatives.

Moving from AI pilots to long-term value

As generative AI cements itself within business strategy, the build-versus-buy dilemma becomes less about the technology and more about prioritisation. Neither method is universally correct, but the wrong choice can slow progress. The businesses that will succeed are those who assess their data maturity and risk tolerance alongside clear ROI goals and long-term value.

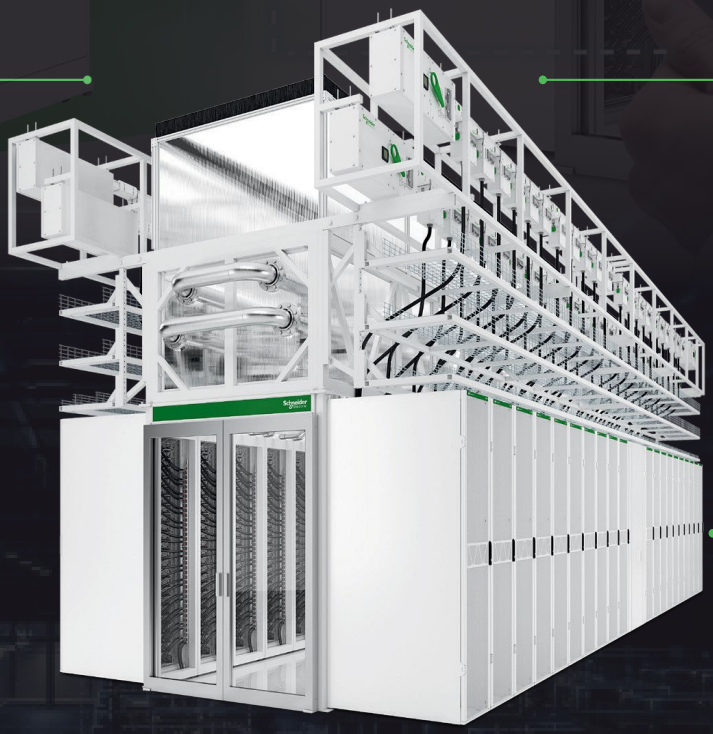
As generative AI cements itself within business strategy, the build-versus-buy dilemma becomes less about the technology and more about prioritisation. Neither method is universally correct, but the wrong choice can slow progress. The businesses that will succeed are those who assess their data maturity and risk tolerance alongside clear ROI goals and long-term value

Speed Meets Certainty: The Fastest Path to AI-Ready Infrastructure

Accelerate AI Deployments with Prefabricated & Pre-integratedPodand Rack Solutions

EcoStruxure™ Pod and Rack Solutions from Schneider Electric™ deliver a faster, lower-risk method to deploy AI and accelerated compute infrastructure at scale. How? By shifting complexity off-site.

Our solutions arrive fully engineered, factory-assembled, and compute-ready, compressing deployment timelines from months to days, enabling rapid, predictable, and globally scalable AI capacity without comprimising reliability.



Prefabricated for Speed
Factory-built and tested pods and racks reduce onsite work and accelerate time-to-compute.

Deploy in days...
Not weeks.
Not months.

Global Expertise, Delivered Anywhere
Engineering and service experts to support fast, high-density rollouts across 100+ countries.

AI-Ready Power & Cooling
Engineered for the next generation of liquid-cooled AI architectures, with integrated technical water and high-density power pathways built to evolve as workloads grow.

Scalable Pod Architecture
Flexible by design, allowing pod infrastructures to scale from initial deployments to multi-megawatt buildouts with consistent quality across global sites.

Pre-integrated Racks
EIA, ORV3, and NVIDIA MGX racks ship integration-ready for rapid deployment and reliable performance.

Discover the IT Pod

Schneider Electric™