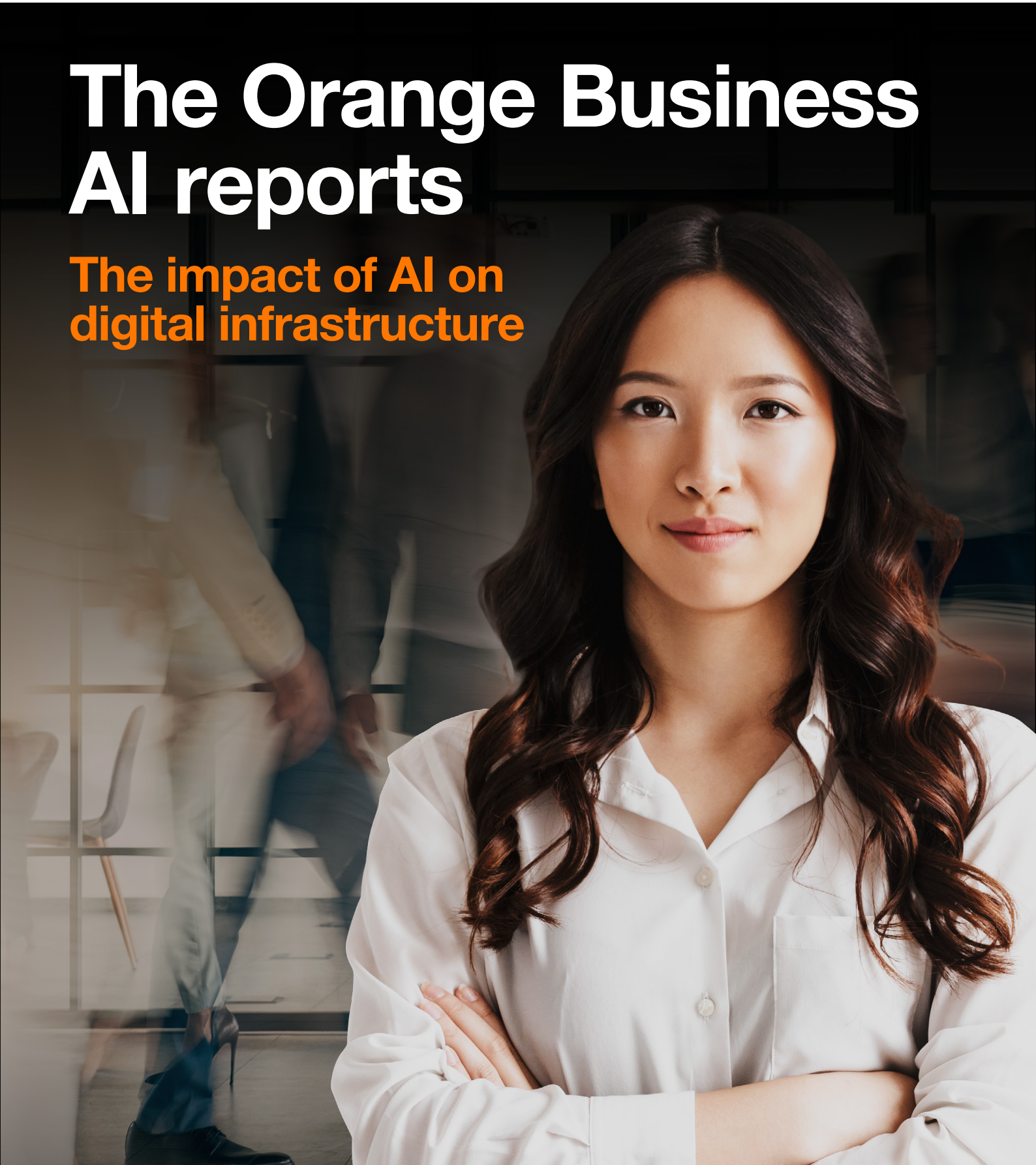
Business

Artificial Intelligence.
Real Wisdom.

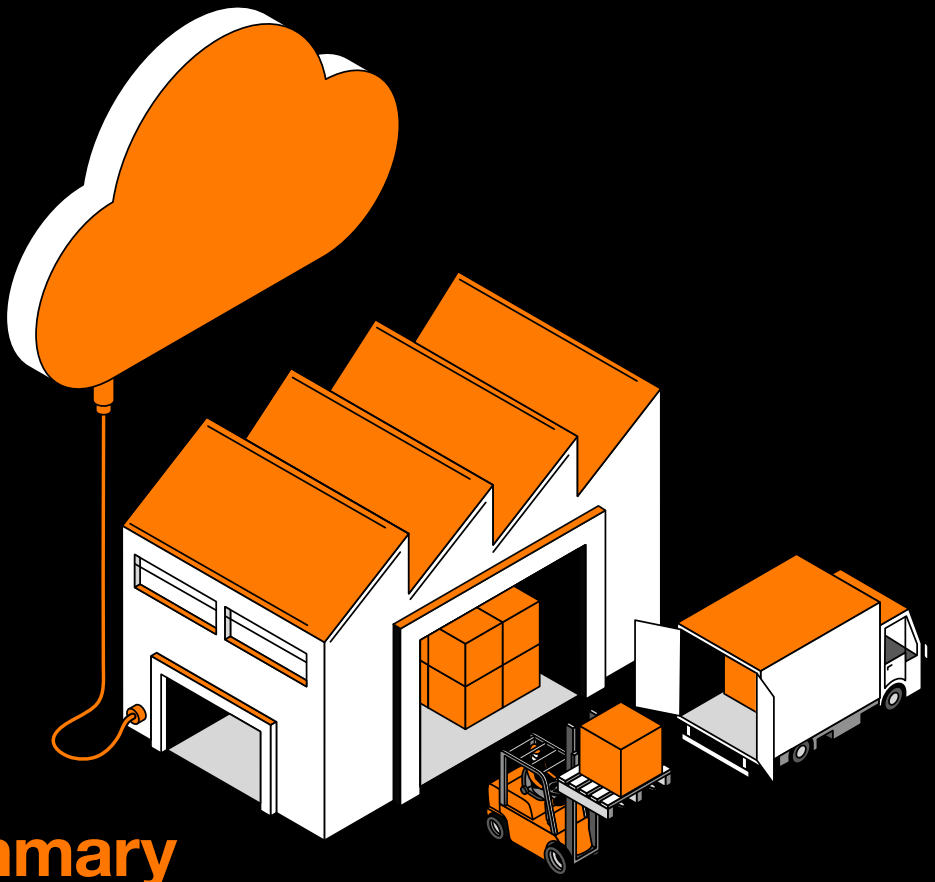
Lily's COO thinks their digital infrastructure is good to go with GenAI. But Lily knows where the gaps are – and how to fill them. Learn what she and other Orange Business customers know.

The Orange Business AI reports

The impact of AI on digital infrastructure



Business



Executive summary

Virtually every business executive aspires to leadership – and so many of them have their eyes on generative artificial intelligence (GenAI), with 85% of organizations actively using the technology in at least one business function.[i] John F. Kennedy once wrote that leadership and learning are indispensable to each other. So, there is huge value in learning what your industry peers think about their ability to embrace this disruptive new technology.

As listening to others is crucial to the accumulation of wisdom, we commissioned GlobalData to survey 400 companies worldwide to explore how companies were operationalizing GenAI and to better understand its impact on digital infrastructure. These results make clear that there is a gigantic difference between a Proof of Concept (PoC) and a full-blown enterprise-wide AI service.

Operationalising AI

Drawing on the services of a hyperscaler to spin up a PoC is relatively trivial. However, there are huge difficulties in scaling these PoCs to an enterprise-wide service, a process we describe as operationalization. A potential storm is brewing as a result of the gap between expectation and execution.

When it comes to operationalizing GenAI projects, our survey shows that AI leaders are clear on the challenges they face. They identified cloud, security, infrastructure, and data management as the critical underpinnings of any production-grade AI service, yet the survey demonstrates that organizations are struggling in each of these areas. A huge proportion – 96% – of enterprises said they needed to re-evaluate their cybersecurity strategy due to GenAI. 70% said the same thing about their cloud strategy. Less than half of organizations had or will have the IT infrastructure they need to support GenAI deployments. Finally, just over half (51%) listed the complexity associated with data management as a key pain point.

These are serious headwinds that will buffet organizations as they attempt to operationalize their AI services. However, as the proverb says, to be forewarned is to be forearmed, so we hope this helps to prepare you for your AI journey ahead.

Deployment and preparedness

The vast majority of those interviewed are already exploring GenAI (or plan to do so in the next 12 months) but notably few have deployed GenAI at scale. However, there are significant concerns around core parts of GenAI strategy – particularly data management and governance, legacy integration and vendor lock-in. In more than half of all cases, IT is responsible for rollout, with operations and the C-Suite (apart from the CEO) also taking direct responsibility.

The clear message is that more resources are required. The majority are investing in:

- Specialized personnel – a challenging task in an overheated market in which AI skills are in short supply
- Computing resources – which accounts for what is described as 'skyrocketing' cloud costs as organizations look to scale their AI projects

This lack of resources is also hindering the ability to create a unified observability solution, which is also hindering AI implementation.

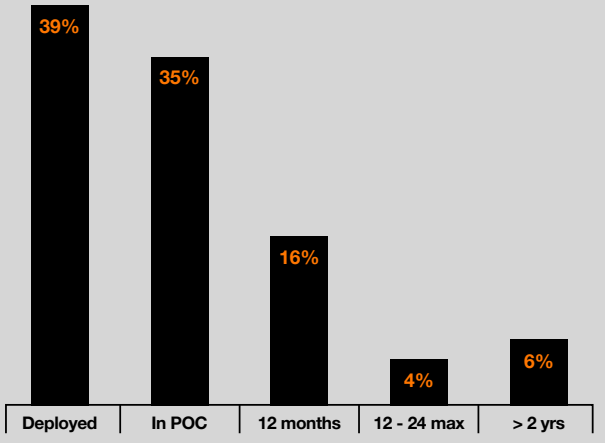
Few are undertaking their GenAI journey alone, with a large minority (42%) turning to third parties to get the help they need: cloud/app platform vendors, communication service providers, and IT service providers were the top three partners that enterprises turn to for third-party support.

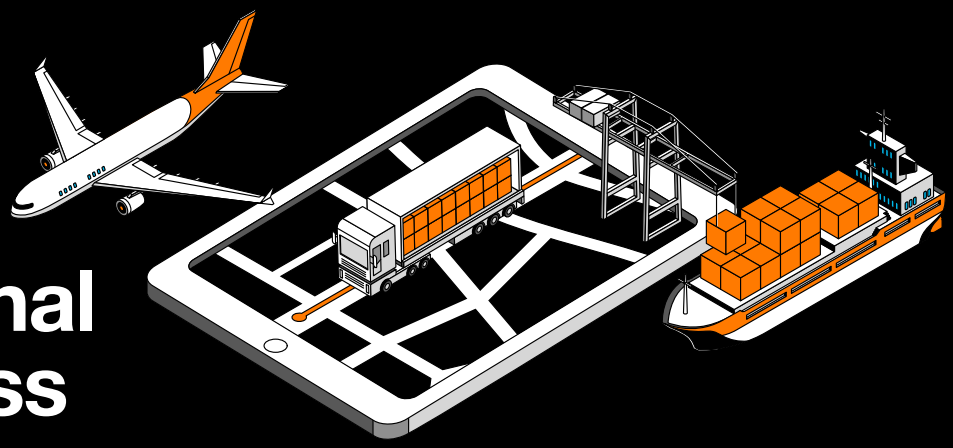
Deployment

There has been robust growth in both general AI and GenAI markets. In the next two years, 64% of European companies plan to prioritize spending on AI platforms, and 66% will prioritize spending on GenAI platforms. However, while many companies have launched their first project, a large portion are still in trials or PoC with GenAI .

GenAI is being evaluated across the entire business, with use cases ranging from writing code to customer support. .

Timeline for GenAI Deployment





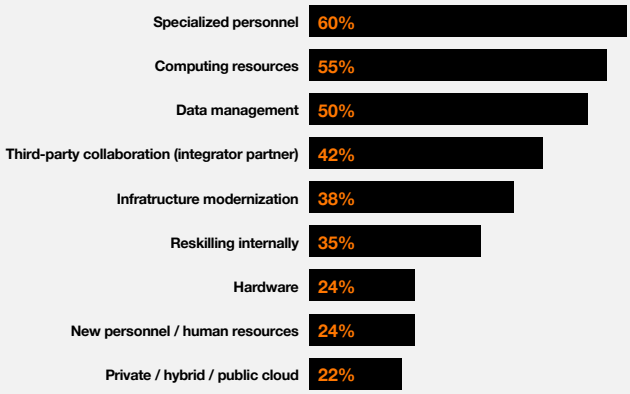
Organizational preparedness

Since the introduction of ChatGPT, enterprises have been trying to get to grips with the phenomenon. As they do so, CIOs and IT leaders express concern regarding the complexity of GenAI. Vendor lock-in is also a concern, since new solutions are coming to market at a rapid pace. In this evolving market, AI services must have portability built into the solution, so organizations can avoid being locked into their chosen foundational models.

Top GenAI implementation concerns

To overcome these issues and scale projects, enterprises are investing across the board to help scale their GenAI projects in areas such as talent, computing resources, and data management.

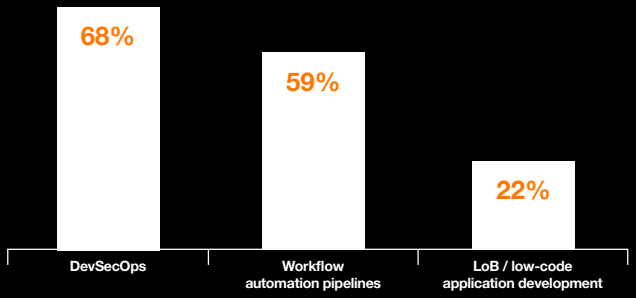
Investments planned to support scaling GenAI in the future



DevOps and observability

Enterprises feel their visibility of DevOps is not sufficiently unified in technology terms (e.g. security) and operationally across relevant teams. This hampers collaboration and reduces the efficiency and effectiveness of deployments. Once again, a lack of internal expertise is a pain point, this time affecting both workflow automation and application deployment and observability.

Areas needing improved colaboration

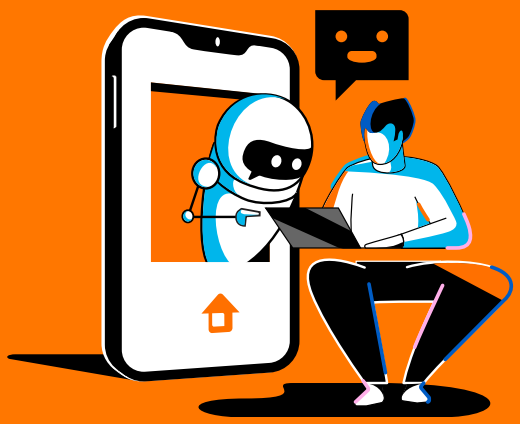


Recommendations

Your GenAI deployments will be judged on how they perform, how much they cost, their carbon footprint and, ultimately, how suitable they are for the tasks you need them to undertake. Selecting the most economical and effective models should be based on comprehensive evaluation, one that avoids hype and brand names. It may be that small language models for specific use cases are more suitable than general large language models.

It may also be that the success of the model will depend on your ability to acquire and tailor high-quality training data. This will involve having a robust data integration strategy that encompasses data cleaning, transformation, and validation processes, along with associated governance.

Training models require significant resources. To manage costs while ensuring reliability, consider:



- Implementing energy-efficient hardware and optimizing algorithms to reduce computational requirements through eco-design
- Selecting the appropriate AI model that best fits the requirements to prevent overconsumption
- Using on-demand hardware capabilities (GPU) to meet ad hoc requirements during training and inference
- Deploying local compute power where your business data sits, to support the integration of retrieval mechanisms
- Optimizing computational resources to effectively manage the complexities of integrating retrieval and generation processes

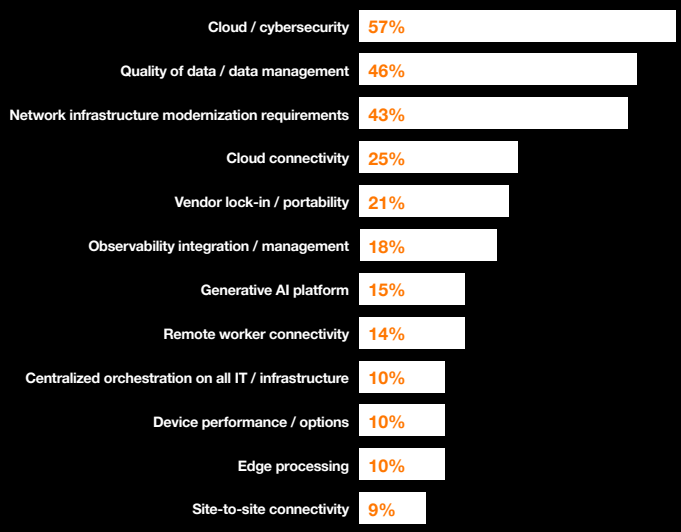
Infrastructure/Cloud

As noted, respondents are re-evaluating their security and cloud strategies to support their GenAI services and modernizing their network infrastructure for the same reason.

Additionally, 73% said they would need to upgrade their storage to support retrieval-augmented generation (RAG). Cloud / cybersecurity and quality of data / data management were the top two concerns relating to IT infrastructure.

When implementing GenAI, organizations are worried about security, data quality and management, network infrastructure and cloud connectivity infrastructure. However, even those with the right infrastructure in place need to ensure that they have the appropriate systems to allow them to scale in the future.

Top concerns related to IT infrastructure



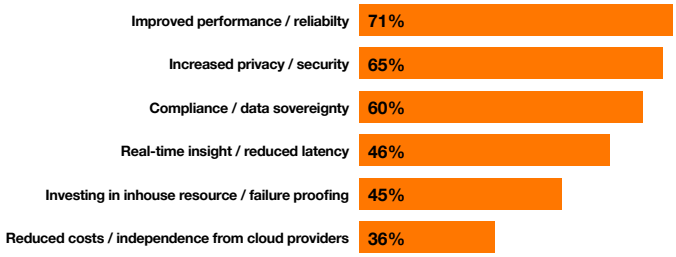
Connectivity strategy

Over 80% of organizations, particularly those using AI and GenAI applications to monitor safety or quality control, require real-time or near-real-time data analysis. Because GenAI draws on data that comes from all parts of a company network, businesses require reliable, secure, and fast connectivity – not only to share captured data but to enable its analysis in the right locations.

As organizations mature in their AI journey, they can benefit from moving away from cloud-based solutions to AI on-premises or at the edge of the network. Local AI processing can help meet the latency requirements of real-time applications, assuage security and regulatory concerns, and minimize the need to move sensitive data across geographic boundaries.

The survey found that over 65% of enterprises have a strong need to run GenAI on-premise or at the edge and that they were seeking to do this to improve performance (70%), increase privacy (65%), and ensure regulatory compliance (60%).

Benefits expected when running GenAI on-premises or at the edge via SLMs



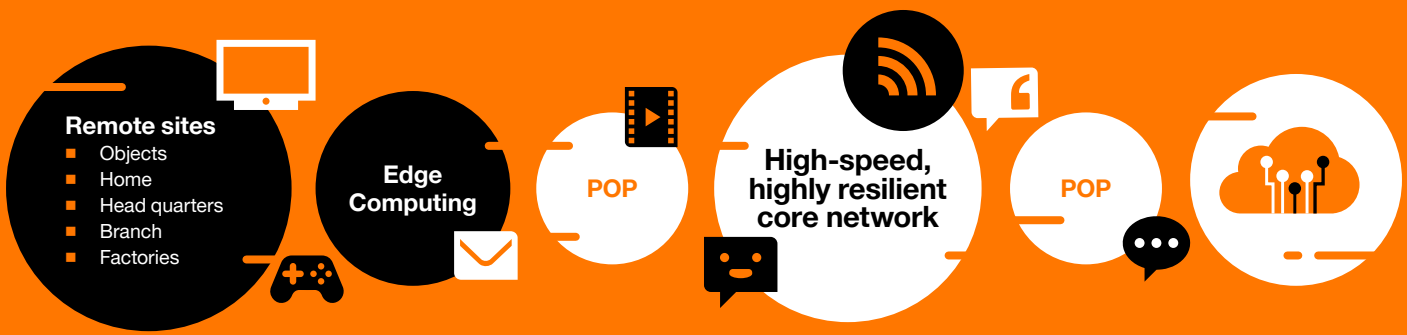
Recommendations - Infrastructure

The data that GenAI needs is generated everywhere and anywhere, and enterprises are increasingly focused on analyzing this data in decentralized locations. This demands high-speed, low-latency network infrastructure to facilitate rapid data transfer between edge locations, regional sites, central facilities and the cloud.

You could incorporate edge computing solutions that process data locally, thereby reducing latency and enabling real-time analysis. You could also consider approaches such as retrieval-augmented generation (RAG) – which enables large language models (LLMs) to access and incorporate information from external data sources in real time but without the need to retrain the entire model – when developing your network strategy.

Of course, the network needs to be secure, reliable, and scalable so you should implement robust security protocols – including encryption and secure access controls at the edge – to safeguard data in motion. Designing the network with redundancy to minimize downtime will help to ensure reliability, while allocating bandwidth and ensuring minimum service levels will help performance and scaling.

To further boost flexibility, scalability, and redundancy, consider multi-cloud strategies, which will also help prevent vendor lock-in.



Cloud strategy

Many organizations start their AI journey, especially their GenAI journey, with cloud-based AI solutions, which require robust cloud connectivity. The cloud is therefore critical to enabling AI at scale. To support multi-model AI environments, cloud implementations must be efficient, secure, properly provisioned, and cost-effective.

As noted, many organizations are struggling to align their cloud strategy with proposed new GenAI deployments – ensuring legacy integration, managing security, privacy and data sovereignty and maintaining infrastructure performance top of mind. Many organizations are re-evaluating their cloud strategies for these reasons and to address the skyrocketing costs of cloud services resulting from GenAI and LLM performance demands.

Data Management Pain Points

Storage

The demand for storage is growing as enterprises turn to RAG to fine-tune LLM results and also use more unstructured data. Seventy-three percent agreed that they would have to upgrade storage to take advantage of RAG, which requires the use of vector databases and embeddings. Additionally, 65% agreed that they would do the same because of GenAI's ability to more easily uncover insights in unstructured sources such as audio and images. The end result is the same – a need to upgrade storage, with the attendant costs.

Recommendations - Cloud

Most organizations have adopted a multi-cloud strategy that increases complexity, security risks, and operational challenges. These problems will only be compounded by GenAI deployments so all enterprises should rethink their cloud strategies.

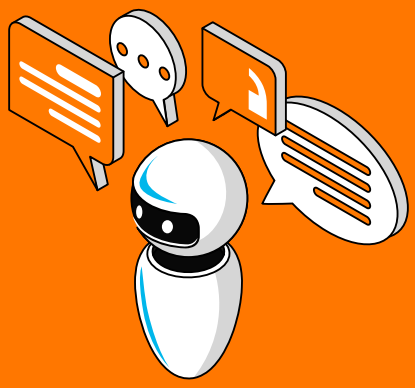
As with networks, the key focus is on security, reliability, and scalability. Potential actions include:

- Optimising GPU usage to manage high costs and ensure availability by securing access to specialized hardware in advance
- Implementing robust multi-cloud security, access control measures, and strong cybersecurity protocols with multi-cloud security management
- Utilizing scalable storage solutions that provide rapid access and maintain data integrity, incorporating regular integrity checks and backups, tailored to your private cloud, edge, or hyperscaler
- Adopting scalable architectures and dynamic resource allocation to sustain performance and efficiency during large-scale deployments

For this new strategy to be successful, you should create a network highway for moving applications across clouds. Employ an end-to-end design, ensuring deployment and unified operations for managing networks across multiple cloud environments and inside clouds.

To manage costs, consider transitioning network management, control, and data connectivity to an external cloud infrastructure.

Compliance and governance are key themes with an impact on the cloud domain. Implement rigorous data-cleaning processes and bias mitigation techniques to ensure reliable model output and establish data and document governance. Ensure compliance with data sovereignty laws, potentially deploying AI with multiple data locations.

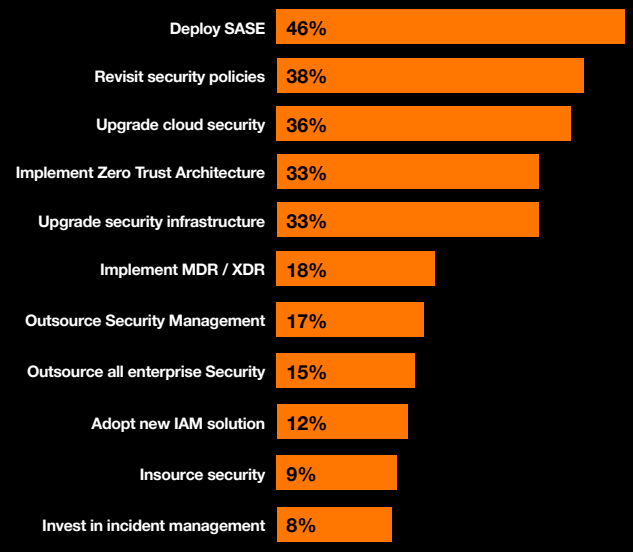


Data management and governance

Data quality is not a 'nice-to-have' but a 'must-have'. It was given as the most important reason for satisfaction with a GenAI project (51%) and was the third most quoted reason for dissatisfaction with a poorly implemented project (38%). So, data quality is the biggest single determinant of the success – or otherwise – of a GenAI project. This is not a can that can be kicked down the road: the experience of Orange Business's customers is that it is at least 20% more expensive to remediate data quality issues once the AI model has been built than it is to ensure clean data from the start.

Data governance – particularly around intellectual property protection and privacy – is both a significant public concern and an organizational challenge for customers. The introduction of GenAI has highlighted the importance of a robust data management and governance strategy, one that can handle the new complexities and oversight resulting from AI implementation.

Data management pain points

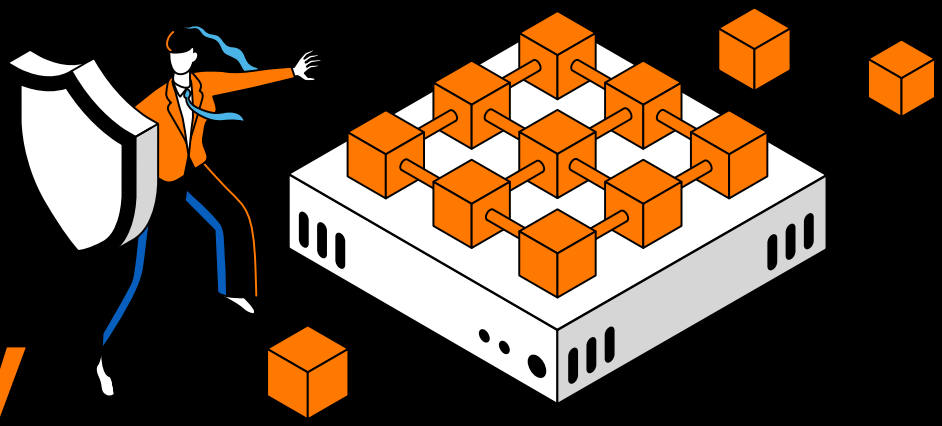


Recommendations

End-to-end visibility is necessary to ensure data access while also implementing the highest levels of security. Data governance is therefore critical. For efficient AI, having clear data ownership and processing guidelines will support compliance with relevant regulations. Integrating data strategies into your business plans helps uphold data quality, accessibility, and security.

This guidance also applies to unstructured data (such as documents), which will be crucial if a model is to be fine-tuned and RAG is to be successfully performed. You should implement access control but apply document governance that sets out criteria for document quality. (We may even see the creation of roles such as document engineer or scientist.)

Enforcing policies around data sensitivity and data lifecycle will reduce the risk of GenAI producing inaccurate results or violating data privacy. Overall, transformation into a data-driven organization requires the adoption of a data-centric culture: this includes educating employees about how to use GenAI as well as when and where to make use of it in their work practices.

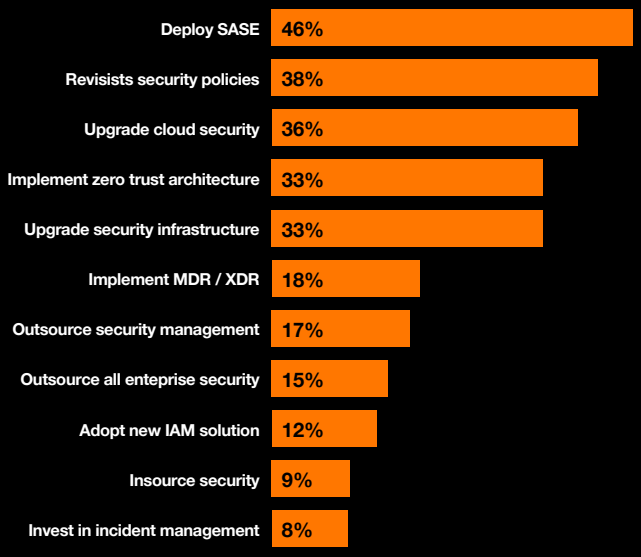


Security

Ensuring the security of AI data as it traverses the corporate network is paramount and this is not something that can be 'bolted on', it must be 'built-in'. Securing your data and your infrastructure in silos and then combining these solutions into an AI service will not deliver the security you need. And it is always less expensive to create strong security from the outset than it is to fix a breach.

Unsurprisingly, SASE (Secure Access Service Edge) and Zero Trust architectures, which enhance AI security by applying strict access controls and monitoring user activity across the network, regardless of location or device used, top the list of strategies deployed by respondents as they look to upgrade their security strategies.

While the vast majority of organizations surveyed are re-evaluating their cybersecurity strategy as a result of their use or planned use of GenAI, those using open-source solutions are especially concerned. A lack of expertise, restricted budgets, and data management are all presenting challenges.



Recommendations

Part of GenAI's appeal is the speed at which it can act, but that also means that any problems move much faster and can, through enterprises' interconnected nature, spread much further. As part of your cybersecurity strategy reassessment, you therefore need to establish secure access controls for data and GenAI deployments to protect sensitive information. Robust security postures and stringent governance procedures to restrict unauthorized use are also critical if data leaks and other threats are to be minimized.

These need to be deployed in a way that does not restrict the potential advantages of GenAI. To strike that balance requires:

- A Zero Trust architecture with streamlined technology and consistent user policies
- Centrally managed security policies for the entire network, simplifying enforcement, updates, and configurations, thereby maintaining a uniform security posture throughout your organization
- A secure, distributed, and flexible network using a Secure Access Service Edge (SASE) solution that encompasses on-demand cloud connectivity and Security Service Edge (SSE). SSE is a subset of the SASE framework and is defined by Gartner as 'a collection of cloud-based security services that protect access to applications, cloud services, and the web'
- Using SASE's cloud-native approach to scale rapidly and globally

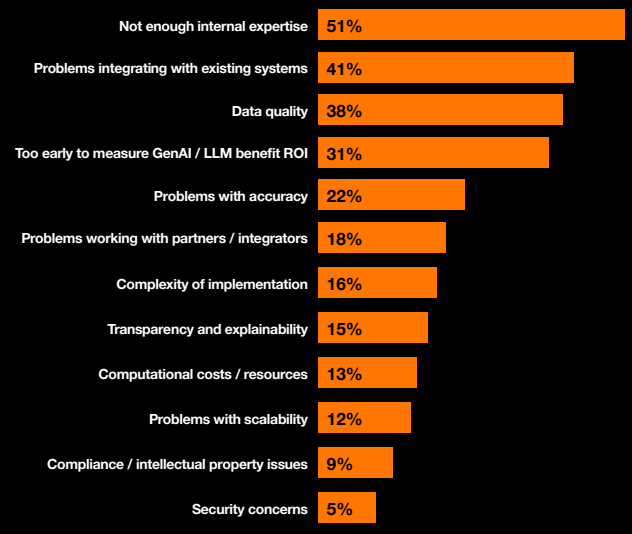
In addition, implement continuous monitoring and rapid response protocols to address potential security breaches. Rely on expertise and cyber defense solutions to prevent data corruption and mitigate unexpected behaviors or biases infecting models.

Value Delivery

GenAI deployment might be a strategic investment, but value for money is a key indicator of success in the eyes of most businesses. A majority of respondents felt their projects were delivering against that goal, but only 21% strongly believed their AI investments were generating value for money. However, while more than one-third (37%) are struggling to see what they are getting from their tech investments, respondents are optimistic that they will be able to effectively measure deployment ROI in the future.

More than half of enterprises are satisfied with their GenAI projects, but failures are common. Many of these result from a lack of internal expertise or an inability to integrate new AI-based services with existing IT environments. Embracing a robust data management strategy is key to successful outcomes, as is implementing a well-orchestrated cloud strategy that incorporates robust security.

Reasons for Dissatisfaction



Recommendations

Central to securing adequate return on investment in GenAI is ensuring that its solutions are used to the extent of their capabilities. In this respect, GenAI is no different from any other technology investment: success depends on securing high user adoption rates.

Developing specific training programs to enhance internal skill sets is therefore critical, as it will allow users to effectively leverage GenAI tools. So too is establishing whether existing devices are fit for purpose; nothing will hamper adoption more than end-user computing that is not compatible with GenAI solutions.

Cost remains a concern. One way to improve cost-effectiveness is to migrate away from subscription-based models for procuring cloud connectivity. Additionally, implementing technologies that facilitate AI model orchestration across a variety of cloud platforms can aid in cost monitoring and interoperability.

What comes next

It is important to remember that where you are on your AI journey has a major influence on your best next steps. Based on the results of the survey and our analysis of the market, Orange Business has defined five levels of maturity.

Based on all the above, Orange Business has developed an AI audit: this will generate a personalized report that will tell you where you stand in comparison with your industry peers. It will help you determine your specific maturity level and provide tailored recommendations based on your specific use cases. By considering these impacts, your organization can understand how to operationalize your GenAI PoCs and turn an experiment into an enterprise-wide service.

About the survey

Orange Business commissioned GlobalData to survey 400 companies with at least 250 employees based in Europe, North America, and Asia Pacific in 2024. Respondents came from manufacturing, retail, warehousing/ logistics, automotive, financial services and insurance, energy and logistics, government public sector (including healthcare), and mining/construction industries.

[1] Economist Impact report "Unlocking enterprise AI: opportunities and strategies"