How to scale your AI

Our experts share strategies for success they see in real-world deployments.

verizon business

The buzz around artificial intelligence (AI) has grown deafening.

We're told it has the power to transform businesses in a multitude of ways. But what does that really mean? And how can we bring it to life? Businesses around the world are already exploring how to use it to drive innovation, enhance efficiency and boost productivity. In many ways, we're only starting to understand what AI is capable of.

But how do we get the most out of AI, and really put it to work? The problem comes with how some businesses are approaching it. AI can be infused in many parts of your business. The potential feels game changing in some applications. But are you able to adequately manage, support and secure it? And how do you take it from proof of concept to full-scale global deployment?

Many businesses have already successfully deployed AI products and services. Companies of all sizes, in all industries, have had successes and failures. While some AI initiatives have worked on a smaller scale, when the organization tries to build them up, they fall apart. And of course, large language models (LLM) aren't the whole story. Many companies have been working with machine learning (ML) for years.

"For every 10 AI projects, maybe eight of them get canned because they're not delivering," says Colin Wilson, Enterprise Architect at Verizon Business, "but two are monumentally successful and making a really big difference."

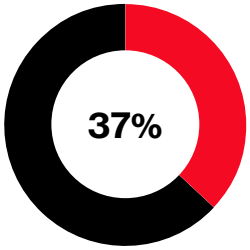
To be successful with AI at scale, you need to lay the foundations correctly, avoid bottlenecks with digital infrastructural upgrades, plan for new ways of working and anticipate the demands of scalability on data, connectivity and security.

> AI has the potential to be more transformative than electricity or fire.[1]
>
> **Sundar Pichai**
> CEO, Google

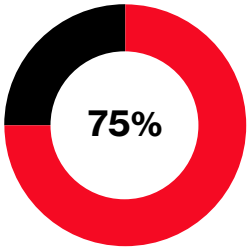
1. https://time.com/partner-article/7279245/15-quotes-on-the-future-of-ai/

What's the state of AI?
So, how exactly are businesses using AI? How are they scaling it throughout their operations? What challenges are they facing? And what lessons have they learned along the way?

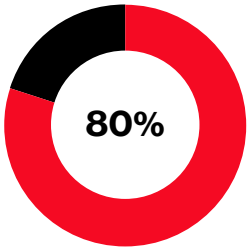
To learn more, we commissioned a study in partnership with S&P Global, interviewing IT decision-makers who are all involved in managing or implementing infrastructure for IT workloads for AI. The report is published in three parts, considering three standout themes: AI network infrastructure, AI security and AI best practices. Key findings from the S&P Global study show that:



of respondents think their infrastructure limitations are preventing their organization from keeping up with AI advancements.



believe they'll need to make changes to their infrastructure.



More than 80% think they'll need to update their security approach.

Make sure to check out the full reports for more essential insights.

What's the problem?
Overall, it seems that businesses are increasingly investing in AI to improve efficiency and innovation. But they're being held back from experiencing the full benefits because of limited infrastructure, bandwidth, latency and security. Compliance is another key concern, as businesses navigate new risks and regulations. Naturally, they're cautious about the cost of implementation—and getting a good return on investment. But, according to the S&P Global study, the resounding theme from those that have deployed AI so far is that they wish they'd laid stronger foundations at the start.

There is, of course, potential to overcome these limitations, helping you deploy AI seamlessly and effectively. But to do that, there are several things you'll need to consider. At Verizon Business, we've been exploring the impact and opportunity of AI for some time. And we've been working with customers to get their IT infrastructures ready to roll out their own AI initiatives.

Here, we outline key themes, insights from those that have already deployed AI and the latest thinking from our own experts on strategies for AI deployments.

1. Reliable connectivity is essential.

AI requires data—serious amounts of it. The problem comes when you need to move that data. First, you need to move it to where it's processed. Then, you move it to where it will be used. That's a lot of movement, all of which has to happen very quickly. And that requires a powerful network infrastructure, with secure and robust connectivity, built to support AI workloads. A slow network can mean inaccuracy and potential failure of the AI application.

"The top tech companies like Nvidia, Google and Meta are all developing their own custom-made AI chips, to sit in their data centers, or in the cloud, because they need to process information the fastest they possibly can," says David Bailey, Global Solutions Executive, Verizon Business. "That's all well and good. But if you have an unreliable network in the middle, it's irrelevant having that fast processor at the top."

Colin Wilson provides a possible use case. A company creates an AI model that can accurately view and monitor the inventory of every shelf in a store. Not only that, but it can also track how customers move through the store, what they look at and buy, and then analyze all that data to build up patterns of potential future sales. Retailers are going to want that model. And that's where the challenge comes in. With so many people wanting it, the company first has to hyperscale its graphics processing units (GPU) and capacity to meet demand. Then, of course, the customers need to be able to get their data to the company's AI model for it to be processed.

> The growing demand for AI ... will increase the need for substantial investments in network upgrades.
>
> **Sanjiv Gossain,**
> Vice President & Head of EMEA, Verizon Business

“If the data being processed comes from video feeds in the retail spaces,” says Colin Wilson, “you’re dealing with vast amounts that need to be pushed to the model running in the cloud. So, now, you’re suddenly looking at provisioning tens of gigabits of data in a matter of days.”

This means that for many AI applications, machines will be processing data from live video feeds, which creates huge amounts of data and takes up a lot of bandwidth. Everything needs to be backed up by a fast, robust, powerful network that can consistently handle this extreme flow of data. The network also needs to be flexible enough to scale up quickly when the flow of data increases. If not, it could fall at the first hurdle.

The kind of connectivity is also very important, says David Bailey. We live in a world where data is nearly everywhere, all traveling at the same time across networks. To run AI projects, you need a provider that gives you high priority access to reliable connectivity. You can’t afford to be lumped in on a public network, competing with the vast amounts of data produced by other sources. “You’ve got AI for consumers (wearables, mobile phones etc.) all creating more data,” he says. “Businesses on a lower priority, public cloud internet connection are competing with kids coming home from school producing yet more data, and their network performance will start to suffer.”

If you don’t have control over performance, capacity for scalability, prioritization and segmentation, you have to compete with this vast proliferation of data, and your AI performance will likely suffer. As Sanjiv Gossain, Vice President & Head of EMEA, Verizon Business, says: “The growing demand for AI applications, particularly those that require real-time data processing and high-bandwidth communication (such as autonomous vehicles, smart factories, ports and retail stores), will place a significant strain on existing network infrastructure. This will increase the need for substantial investments in network upgrades and expansion to accommodate the increased data traffic.”

Re-evaluating and potentially redesigning your network to get it fit for implementing and scaling AI projects could involve upgrading hardware, implementing new technologies or optimizing your network architecture. Verizon Business offers a combination of tailored network solutions, as well as consultative and managed services that can help you address these complex challenges.

2. You need to properly manage your data.

For AI to be effective, the data it processes needs to be high quality. Without regular updates of this quality data, AI has nothing to learn from and simply can't perform how you want it to. Or worse, feed it poor quality data and you'll get poor results. This, naturally, creates implications for your network. To be able to constantly update your AI with the data you want it to learn from, your network must be fast and powerful enough to manage it quickly.

"You have to keep updating your data to get the latest and greatest output," says Marc Mombourquette, Senior Product Marketing Leader, Verizon Business. "The AI is only as good as the data that you put into it."

In many cases, he says, projects fall apart when businesses try and move project data to the cloud. This is because the businesses fail to consider the time, effort and cost of moving their data from an internal network onto an external one.

To feed the data demands of your AI, you need a robust plan for how to manage and move data—and a network that enables it. "Enterprises need to think about how they move data back and forth," says Marc Mombourquette. "And they need to think about improving their network to give them the ability to move data worldwide in an instant."

> AI is only as good as the data that you put into it.
>
> **Marc Mombourquette,**
> Senior Product Marketing Leader, Verizon Business

It also stands to reason that the more AI deployments you have, the better your data strategy needs to be—with the ability to shift different data sets to different AI models. To help manage the movement of data in and out of models, architecture now often uses network segmentation, creating secure pipes to ensure secure and seamless flow of data.

"We've seen a shift to enterprises running multiple AI workloads or models and they're trying to connect different data sets to them," says Colin Wilson. "They've got this spider's web of data in one place, users in another and workloads or large language models running in a third location. This is all happening at scale and driving network demand."

From the S&P Global report:

"Once you start automating processes and talk to the LLM more often, your latencies, your hardware, your GPUs, everything will increase. I think in the matter of the next few years, the majority of AI will be agentic, which means more chatting, more going back and forth between the model and the workflow."

Senior generative AI data scientist, financial services, >100,000 employees, US

3. Fast-growing AI needs a rapidly scalable network.

Some AI projects will need to grow faster than others, accelerating and scaling rapidly. As this happens, you'll need more bandwidth to enable the sudden increase in traffic, as the AI takes in and analyzes more and more data. But if you can't rapidly increase your bandwidth, your AI won't be able to process data fast enough, and your application will cease to work.

"A company might build a prototype," says Marc Mombourquette. "They've got a couple of GPUs, some storage and they've fixed their network to give them the streaming bandwidth needed internally. Everything seems great. They transfer it to the cloud and assume it's going to work the same. But it doesn't, because they didn't take into account the Herculean effort of moving hundreds of terabytes out to the cloud on a regular basis."

The answer lies in a robust, dynamic, scalable network. It enables you to increase your bandwidth to handle big surges in data when you need to and keep your AI running optimally. And when you don't need the extra capacity, you can dial it back down.

The ability to quickly increase bandwidth also helps when it comes to exchanging workloads and data sets globally. "I might have 20 sets of data scattered around the world," says Colin Wilson. "I've got 20 different AI models that all need to see some of those data sets. And when the AI model produces an output, I've got to pass the output from model one as an input to model two and get these AI workloads talking to one another. It's a latency and capacity nightmare for a network."

Scalable networks can also help you deal with spiky data traffic. AI workloads often involve irregular traffic patterns, which means the network needs to be able to handle sudden surges—or spikes in data volume. If your bandwidth is only capable of handling regular traffic and not these spikes, then you'll experience increased latency and the AI won't function properly.

> Trying to scale AI rapidly is what breaks a lot of companies.

Marc Mombourquette,
Senior Product Marketing Leader, Verizon Business

You can increase your scalability by implementing burstable and flexible fiber optic connections through your network. It gives you the extra capacity to turn your bandwidth up and down when needed, so you can handle those sudden spikes.

To help manage the increasing bandwidth demands you'll be dealing with, Verizon Business offers high bandwidth solutions. This includes 400 Gbps options, particularly for private network connectivity between hubs—with a focus on minimizing latency, which is especially critical for real-time AI applications.

For cloud workloads our on-ramp services will scale to meet your needs. And being a truly global operator, we can optimize the path your traffic takes to reach cloud vendors around the world.

With dynamic network management, we provide bandwidth on-demand capabilities, which allows you to increase or decrease bandwidth as needed. In the near future, bandwidth adjustments will be automated based on AI-identified requirements. They're particularly useful for AI model training and data transit. We also support private IP backbones with high-capacity fiber connections that can carry hundreds of gigabytes or terabytes per second. This gives you the ability and confidence to handle vast AI data volumes very quickly, and deal with data spikiness without dropping out.

From the S&P Global report:

"Optimizing data transfer—specifically improving bandwidth and latency—is a key area to consider. In AI/ML workloads, large volumes of data are transferred to the cloud and bandwidth can easily become a bottleneck."

Director of technology, professional services, 1-5,000 employees, Japan

4. Latency can be make or break for AI applications.

For many AI use cases—voice applications, customer service chatbots, or anything involving vehicles, transport, medicine or public safety—low latency is critical. In these applications AI needs very low latency to be able to process data and work correctly, otherwise it could fail. Imagine an AI-powered autonomous vehicle in a logistics hub suddenly unable to receive or transmit data or commands because the network is sluggish. Disaster.

> Time-critical AI apps simply won't work without low latency.
>
> **Duncan Kenwright,**
> Managing Director Global Solutions, APAC, Verizon Business

Applications like voice platforms, call centers, autonomous systems, interactive experiences and safety applications are particularly sensitive to latency. They need fast data processing to work properly. If they fail or experience delays, customers simply may not tolerate it. That can lead to poor customer engagement, damaged brand reputation and loss of business.

"If you have an AI voice platform," says Duncan Kenwright, Managing Director Global Solutions, APAC, Verizon Business, "then you can't have latency. If customers are speaking to a machine and its responses aren't quick enough, they'll just give up."

To ensure you have the necessary low latency to keep AI applications functioning, you should consider edge computing, which changes where the data is processed. By processing data much closer to its source—rather than in a remote cloud data center—you can process it in real time while reducing bandwidth usage and minimizing latency. All of which means your AI can work faster with less risk of dropping out.

"In health and safety or traffic control, seconds really count," says Colin Wilson. "You've got to run at least some of that workload at the edge very near to where the situation is."

Of course, not everything needs to be processed at the edge. Data that isn't needed for real-time decision-making, such as historical sales, inventory records, logs and customer data, can be processed in batches in data centers. That's why many companies operate a hybrid approach, with the AI that handles more generic tasks running in the cloud, and edge computing used to handle more sensitive, low-latency apps. Marc Mombourquette says Verizon Business networks are intelligent and programmable, as "customers want to control where and when AI workloads are running. And to achieve that, we are building new capabilities to give users more insight and control via new tools. Essentially, we are allowing them to program their network resources to optimize their operations as they see fit."

From the S&P Global report:

"Real time sales data analytics; the S&OP [sales and operations planning] and warehouse AI projects are scheduled to be completed in six to 12 months. The experiences with them will inform our future cloud on-prem strategy … AI edge computing will most certainly be a consideration once we move into production optimization."

Head of cybersecurity & land infrastructure, machinery manufacturing, 1-5,000 employees, Germany

5. As your AI develops, so does your cyber risk.

As your AI applications grow and develop, so does the level of data required to power them. It means you'll have lots more data that you not only need to manage, but also need to protect from cybercriminals, who are growing—both in sophistication and the number of attacks they perpetrate each year. It's no surprise, then, that 80% of AI leaders interviewed for the S&P Global report believe they'll need to make moderate to significant changes to their cybersecurity to keep up with their AI plans.

> The more you have, the more you need to protect.
>
> **Thomas Raschke,**
> Senior Product Marketing Manager, Verizon Business

And that's not the only problem, because the cybercriminals are also harnessing the power of AI to make their own attacks more effective. AI enables bad actors to create more sophisticated attacks, like identity-based scams and AI-written phishing emails.

So while you need to protect the increasing volumes of your own AI data, you also need to protect yourself from AI-enhanced cyber threats.

"In one example," says Toni Horne, Director Solution Architecture Asia, Verizon Business, "a UK-based CFO, who would often direct a member of his staff to make transactions over video calls, was deep-faked. The staff member received an email and suspected it was phishing.

But then he received a meeting invite (which was usual). He attended the meeting with half a dozen colleagues—who also turned out to be deep fakes. Unwittingly, the staff member processed the request to transfer $25 million to a fraudulent account."[2]

This shows how effective AI can be in the wrong hands, and the kinds of attacks you need to protect against. And it highlights how essential security is—at so many levels and in so many ways. "There's so many layers to security when you start thinking about AI," says Colin Wilson. "Is my data sovereignty intact? Is my data being compromised in transit and sent somewhere it shouldn't? Is my data at rest accessible to external threat actors? Then you've got the AI model itself. How is that secured?"

[2] Found in the global AI Incident database. https://incidentdatabase.ai/cite/634/

So, where do you start?
There are several elements to consider, according to the S&P study:

- **Data security:** Ensure you have the right security measures to protect the data used in your AI applications. This should include everything from data poisoning (targeting and compromising the data sets used to train your AI) to deepfake identities, phishing scams and others.
- **Model security:** Guard against potential vulnerabilities of the AI model itself.
- **Data privacy:** Protect sensitive data used by AI systems.
- **Protection of AI learnings:** Safeguard your AI models and algorithms from intellectual property theft or manipulation.
- **Legal implications:** Manage any legal and liability questions related to the outputs and decisions generated by your AI systems.

The amount that needs to be protected, and the ways in which to protect it, can be complex and costly. Key findings from the report show that as AI expands the surface area for attack, organizations should consider using zero-trust frameworks as well as integrating advanced tools for application programming interface (API) security, enhanced authentication and real-time threat detection. And the resounding advice from both the report and Verizon Business experts is that security can't be an afterthought. It needs to be one of the principal concerns, built in right from the start. As one of the AI leaders surveyed for the report says: "Think about security first; build the proper control early."

You don't need to build or manage this alone. Verizon Business experts can help you implement the tools and solutions that bolster your cybersecurity to give you a greater focus on, and protection from, data breaches, ransomware attacks, insider threats or accidental data exposure.

Thomas Raschke, Senior Product Marketing Manager, Verizon Business, highlights key approaches to security. Network security organizations should implement robust firewalls, deploy intrusion detection and prevention systems, and introduce network segmentation to isolate AI workloads, thereby limiting the impact of potential attacks. With data security he recommends employing encryption, access controls and data loss prevention (DLP) measures to protect sensitive data used in AI models. For AI-specific security challenges, such as adversarial attacks, data poisoning and model theft, techniques such as adversarial training, input validation and model watermarking are advised.

Thomas Raschke recommends employing a cybersecurity framework, such as the one developed by the National Institute of Standards and Technology (NIST). It's a widely used set of guidelines that can help you manage threats and reduce cyber risks. It gives you a structured approach to managing cybersecurity incidents built around five key principles:

- **Identify**
- **Protect**
- **Detect**
- **Respond**
- **Recover**

It's ideal for assessing and mitigating AI threats, and something that we at Verizon Business use.

6. Secure AI requires a new focus on data governance.

Securing your data and your AI is just one step. You also must make sure you have robust ways to treat data feeding in and out of AI systems and good governance around the status, validation and legality of AI information.

> It takes years to build trust and an instance to destroy it.
>
> **Thomas Raschke**
> Senior Product Marketing Manager, Verizon Business

You'd have to be living in a cave not to have read about the various controversies and conversations around AI, plus the myriad legal and ethical concerns of where the data comes from and what it's used for. From questions around author and artists' rights, to AI hallucinations, algorithmic bias and issues of data privacy, there's lots to consider. And much to be wary about. You need to make sure that any AI project complies strictly with industry regulations and ethical guidelines.

Historically the challenge has been managed with DLP, but this now needs to be applied to the AI domain.

From the S&P Global report:

"The biggest challenge is securing sensitive data across hybrid environments while maintaining compliance with regulatory body requirements across multiple countries and regions."

Head of architecture/cybersecurity, banking, Singapore

As Thomas Raschke says, "DLP can also play a crucial role in bolstering AI security by addressing the unique challenges that arise with the integration of AI into various systems and processes. AI systems heavily rely on vast amounts of data, often including sensitive information, making them potential targets for data breaches and misuse."

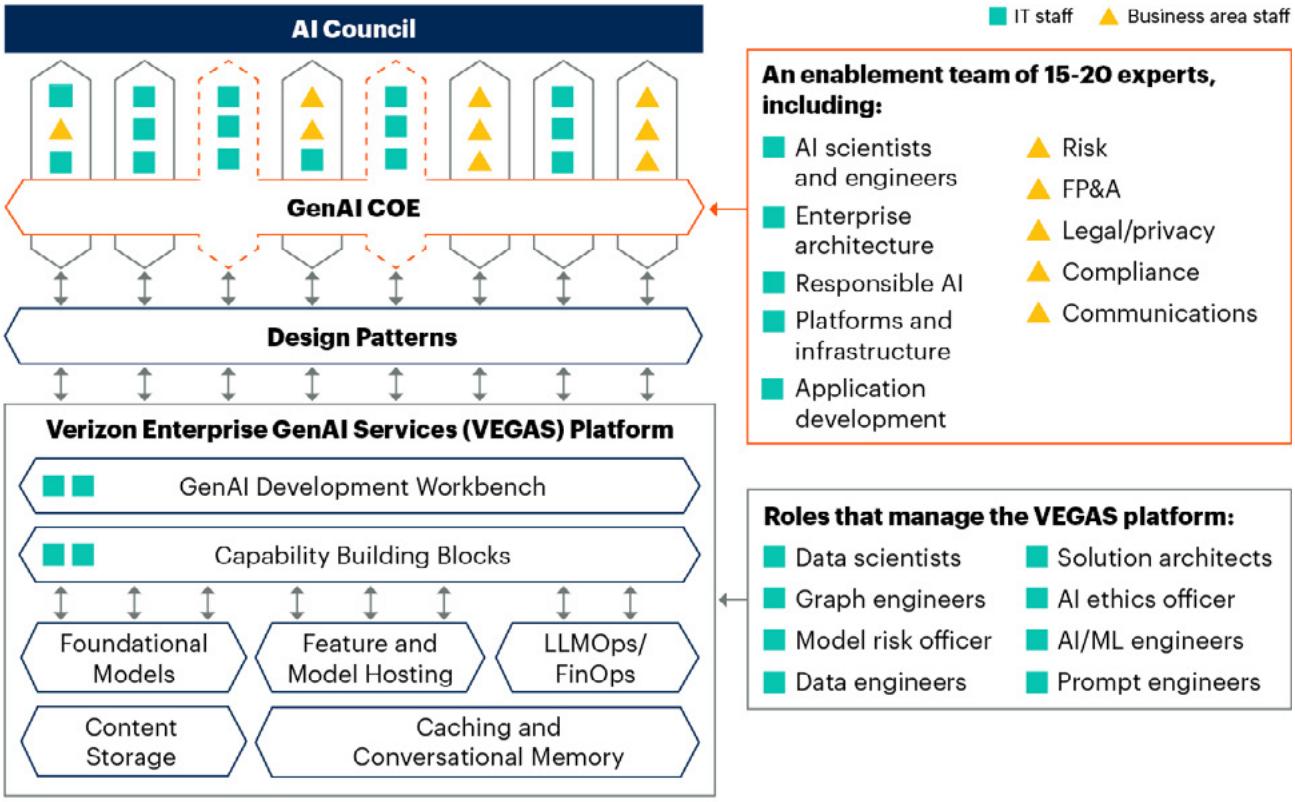
In short, your data must be accurate. Any AI output may also need to be verified and not misleading. You need to safeguard processes to try to ensure data is not misused in any way. And you must abide by all data privacy laws, like the European Union's General Data Protection Regulation (GDPR).

"Right now," says Colin Wilson, "the 'doing something' in AI usually happens with a person looking at the AI model output and saying, 'Yep, that's correct. I'm going to do something.' As we get more confidence in those AI models, the person steps out of the loop and the AI model says, 'Here's an anomaly. We need to do X.' The 'doing' will happen automatically. In my opinion, there are several security, legal and commercial considerations around all of that which need to be thought through. What happens if my AI model gets it wrong?"

All of this highlights the importance of good governance. The scale of Verizon Business and our desire to innovate placed us at the forefront of these challenges. We have developed our own Responsible AI framework. Our journey to successfully scale distributed genAI efforts also included setting up genAI center of excellence (COE), an enablement team of 15 to 20 cross-functional experts. This team brings to life the enterprise-wide strategy and direction from our AI council (a cross-functional leadership team).

The Gartner® report 'Case Study: Enable and Scale GenAI Experiments With Verizon's Platform Strategy' is designed to help CIOs learn to enable and effectively scale distributed genAI initiatives.

Verizon's delivery model for GenAI



Source: Adapted from Verizon

816569_C

7. Private networks can enable better control of data for AI.

As you come to scale your AI, the volumes of data will inevitably grow. At this point, consideration must be given to local area network (LAN) technologies that will support low latency, high capacity, secure connectivity requirements. Will gathering data from sensors and devices via public or private connections be a viable option? While spectrum slicing may work for public utilities and transport, it's likely to be expensive and complex at a site level.

At the site level, especially campus environments, factories, distribution centers and larger retail stores that are creating huge volumes of data to feed AI models, private networks are likely to offer an advantage.

In that case, you might want to consider using a private network, which our experts say gives you better control, helping you secure sensitive data while still getting the speed and scale of data transfer that you require. "Private 5G is different," says Colin Wilson. "I can build my private network as dense as I like. I control what connects and I can run my workloads within that campus environment to process AI data before I then push it out to the cloud."

The S&P Global research report also shows that many businesses are adopting a hybrid model to get the security and control required, while still keeping costs down. You can run lighter workloads in the cloud. And you can manage the more sensitive, critical elements in a private network with edge compute where you have more control. This could give you greater peace of mind that your data remains secure.

As one AI leader tells us in the report: "We take a hybrid approach: AI training is carried out in the cloud, while real-time inference at the point of service is performed on-premises. This allows us to balance cost and service-level requirements effectively ... for processing sensitive data—such as confidential or personal information that cannot be sent externally—we often rely on on-prem infrastructure to ensure data security."

> Bigger enterprises don't like having their data leave their own four walls.
>
> **Robert M Leitner**
> Associate Director, Product Marketing, Verizon Business

8. Managed services can reduce the headache.

Building AI solutions is no easy task. They involve multiple technologies, including compute, storage, networking and security. Each one of these can be complex, so integrating them together can be extremely difficult. Plus, you need to know how to leverage your data for AI. Many businesses lack the skills and expertise to do this in-house. That's where a managed service can be invaluable. We can integrate the network with the existing digital ecosystems providing the foundational infrastructure for AI strategies.

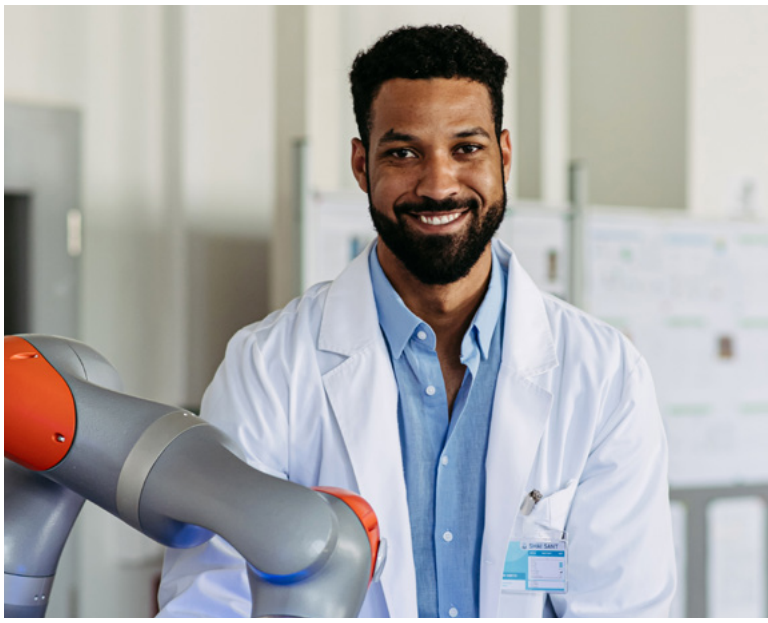
Managed services can help with:

- **Architecture:** Designing your infrastructure from the ground up to provide the security, capacity and flexibility you need.
- **Scalability:** Access expertise in network infrastructure, data management and security to help take AI deployments beyond the sandbox phase.
- **Digital integration:** We connect with your IT service management (ITSM) toolsets to provide simplified visibility and management of your network services in your single pane of glass.
- **Enhance visibility:** End-to-end visibility of the network, allowing for better control and optimization; receive proactive insights of network performance and anomalies.
- **Cost:** Optimize expenses with a more cost-effective option than building and maintaining your own in-house network infrastructure.
- **Security and compliance:** Help enhance data security and ensure compliance with expertise from our global, certified security professionals.

With a managed service, you can save yourself the time, cost and the headache of setting up a robust, dynamic, scalable network to support and power your AI ambitions. Managed services are flexible to customers' differing needs, helping them to determine a model that works for them and offering support while leaving customers with the guiding control. Qualified experts can help you take your AI application from sandbox to global deployment.

> Bring in outside experts to help put it all together.
>
> **Marc Mombourquette**
> Senior Product Marketing Leader, Verizon Business

Naturally, the way you need to use a managed network service might differ from how other businesses use them. But by integrating toolsets into your ITSM environment, you can benefit from a holistic view of your network performance and how it affects application performance. It can help you detect and fix faults quicker than if you were managing things yourself. And with AI monitoring your network performance, developments underway are aiming at getting automated upgrades based on predicted needs.



Summary: Making AI work for your business.

Whatever your AI ambitions may be, and whether you're starting out or you've already built an application and want to take it to the next level, you need the right network to make it a success. Before you deploy AI at scale, you need to assess your traffic patterns and make sure you're set up to handle the demands of AI. Ask yourself the following questions:

- Are you able to access, use and move your data quickly and effectively?
- Is your network built to handle large volumes of traffic?
- Is it highly resilient and built to offer the low latency that AI demands?
- Can it scale up or down to handle spikes?
- Can you guarantee that your data is fully secured?

At Verizon Business, we've been exploring the possibilities of AI for some time. We know what it takes to make it work. And we've been helping businesses of all sizes prepare and right-size their network to experiment with AI, then take it to full-scale deployment.

"AI presents a completely different challenge," says Robert M Leitner, Associate Director, Product Marketing, Verizon Business. "Enterprises have to rethink their entire network in order to execute on an AI strategy. It's not a hodgepodge of stitching together a bunch of networks that they already had."

At Verizon Business, we can deliver a scalable, high-bandwidth, low-latency, secure network to deliver data from the edge to wherever the AI compute is being carried out—in a corporate data center, in a public hyperscaler AI instance, or a third-party service provider. Whether the AI workloads are across a multi-cloud environment, deployed on our customers' premises or at the edge of the networks, we can provide the full suite of offerings to help enable your AI vision and strategy.

Find out more about our network and security solutions at verizon.com/business/en-gb/resources/deploying-ai-at-scale.