Capvion Insights

Strategic perspectives on capital, growth and transformation

The Network Was Built for Humans.
AI Is About to Change That.

Inference, Agents and Sovereignty: The Rise of the AI WAN

As AI starts talking to AI, the wide area network becomes part of the AI architecture, and one of the places where AI policy is enforced

Sander Barens | September 23, 2026
with associates

Capvion Insights provides strategic perspectives on value creation, transformation, and leadership in technology and telecommunications businesses.

Each publication examines how strategy, capital, and execution interact to shape enterprise value in complex markets.

Drawing on extensive board-level, advisory, and operational experience, Capvion and its associates combine strategic thinking with real-world execution across numerous M&A transactions, platform integrations, and growth transformations in the technology and telecom sectors.

Our insights are designed to support founders, executives, and investors in navigating complex growth, restructuring, and transaction environments.

Each publication reflects Capvion's commitment to clarity, independence, and practical relevance.

Executive Insight

For most of the internet era, the wide area network has been built on one assumption: there is a human at the end of the request. AI is breaking that assumption. Inference is spreading across clouds, data centres and sovereign environments, and agentic AI means a single prompt can trigger hundreds of machine-to-machine transactions that nobody ever sees.

That changes what enterprises need from the WAN. Some AI traffic runs comfortably over the public internet. Other workloads need the dependable, measurable performance that Carrier Ethernet services such as E-Line and E-LAN were designed to deliver. And all of it needs the network to act as a control point, deciding which agent may use which model, with which data, in which jurisdiction.

We call this the **AI WAN**: a network that connects people, agents, models, data and compute, and applies the right connectivity characteristics to each interaction. Organisations that design for it now will scale AI with fewer surprises in performance, cost and control.

Key Takeaways

- **Inference is where AI gets used.**
  Training is concentrated; inference is spreading across clouds, data centres and sovereign environments.
- **The WAN is now part of the AI architecture.**
  When models, data and compute sit in different places, connectivity decides how well they work together.
- **AI is starting to talk to AI.**
  Agentic AI turns one human request into chains of machine-to-machine traffic that no human ever sees.
- **Ethernet is back.**
  Deterministic Carrier Ethernet services such as E-Line and E-LAN are becoming strategically relevant again for distributed AI workloads that need predictable latency, jitter, packet loss and availability.
- **Classify the workload, not the category.**
  AI traffic is not homogeneous; some belongs on the public internet, some on direct cloud connectivity, some on deterministic connections.
- **The WAN becomes a control point.**
  Identity, Zero Trust and observability must extend from users and devices to the agents themselves.
- **Sovereignty is about control, not just location.**
  Residency matters, but so do models, chips, licences and the ability to keep operating when the rules change.
- **AI will consume AI.**
  Automated traffic already exceeds human web traffic, and AI agents will accelerate the shift toward machine-to-machine interactions.

Context

For most of the internet era, networks have been built around a simple assumption: somewhere, there is a human at the end of the request.

We open an application, join a video call, download a file or ask an AI assistant a question. The network connects us to the service we want. For much of the last decade the direction of travel has been clear: applications moved to the cloud, connectivity moved toward the internet, and software became very good at absorbing whatever variability sat in between.

AI is beginning to challenge that model. Not simply because it needs more bandwidth, but because the nature of the traffic itself is changing. AI is moving from learning to doing, from centralised training to distributed inference, and from humans talking to machines toward machines talking to machines.

That changes what we need from the WAN.

Analysis

From Training to Inference: The WAN Joins the AI Architecture

Much of the AI infrastructure race so far has focused on training: enormous GPU clusters, data centres, power and the high-performance networks that connect thousands of processors.

But once a model is trained, inference is where AI actually gets used. Every question answered, document analysed, image generated or decision supported requires inference.

The distinction matters. Training tends to be concentrated, while inference increasingly happens everywhere. An enterprise application may call a model in one cloud, retrieve corporate data from another environment and use specialised compute somewhere else. Governments may require certain workloads to stay within sovereign infrastructure, while less sensitive workloads use commercial AI platforms.

AI infrastructure is becoming distributed, and distributed infrastructure needs to be connected. That makes the WAN part of the AI architecture, not just the plumbing underneath it.

When AI Starts Talking to AI

Today, most of us still experience AI as a conversation: we ask, AI answers. Agentic AI changes this, because one request can set off an entire chain of machine-to-machine activity.

Ask an enterprise agent to arrange a business trip and it may interact with your calendar, a travel platform, corporate policy, a payment service and several other agents before presenting anything back to you. A human initiates one request. The infrastructure underneath executes hundreds of transactions.

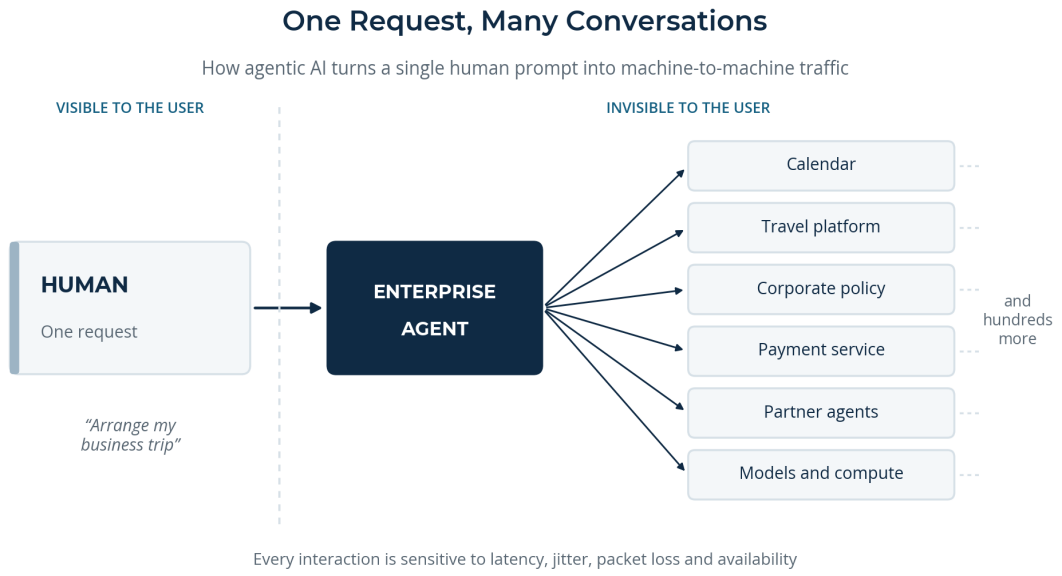

Figure 1. Agentic AI turns one human request into machine-to-machine traffic the user never sees.

Machines also interact differently from humans. A person barely notices a momentary dip in network performance, and applications can buffer, retry or adapt. But when models, agents, data and compute operate together as one distributed process, latency, jitter, packet loss and availability directly shape both performance and economics. A small delay repeated across hundreds of machine-to-machine interactions is no longer small.

> *"The AI network will increasingly carry conversations that no human ever sees."*

Ethernet Is Back

There is an irony here. For years, enterprise networking has moved toward internet connectivity, cloud access and software overlays. That evolution will continue, and the public internet will remain perfectly adequate for many AI workloads.

But AI is also reminding us why enterprises have historically paid for deterministic connectivity. Ethernet never went away, but deterministic Carrier Ethernet services such as **E-Line and E-LAN** are becoming strategically relevant again for certain distributed AI workloads, because they can provide what those workloads need: dependable connectivity, predictable performance and measurable service levels.

Not All AI Traffic Is Equal

None of this means every AI workload needs deterministic Ethernet. The more important point is that enterprises need to understand the requirements of each workload, rather than treating AI traffic as one homogeneous category.

Some inference can comfortably cross the public internet. Other workloads need guarantees on latency, jitter, packet loss and availability. Some need direct cloud connectivity. Others need deterministic connections between data centres, AI clouds, sovereign environments and enterprise locations.

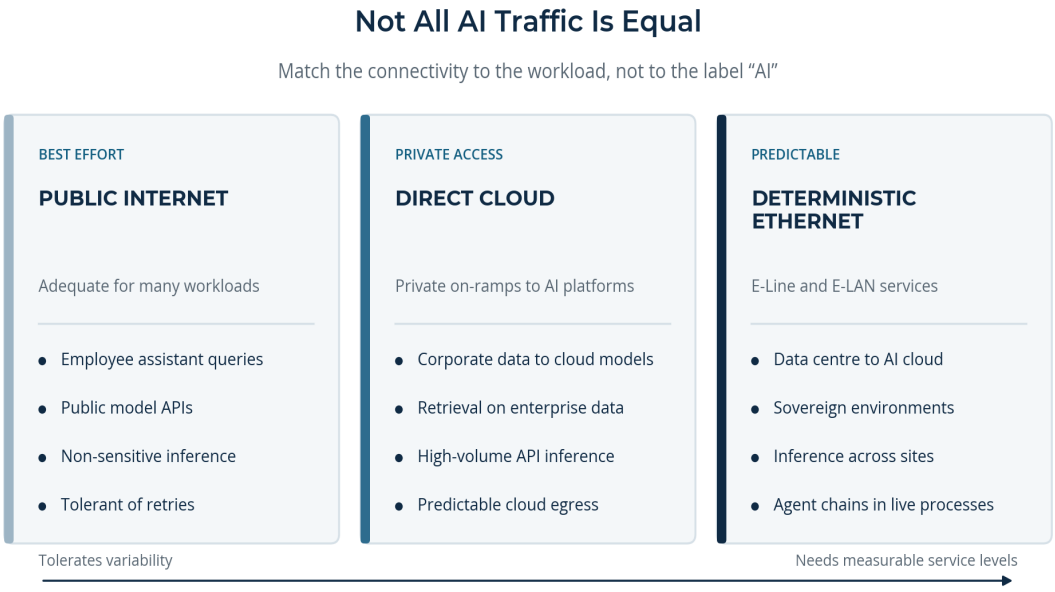

Figure 2. An illustrative mapping of AI workloads to connectivity, from best effort to deterministic.

The next-generation WAN will therefore not simply connect branches to clouds. It will connect **people, agents, models, data and compute**, and select the right connectivity characteristics for each interaction. This is what we call the **AI WAN**.

The WAN Becomes a Control Point

Connecting everything creates a new problem. When networks mainly connected people to applications, security started with familiar questions. Who is the user? Have they authenticated? Are they allowed to access this application?

Now imagine thousands of AI agents communicating autonomously. An agent might access sensitive information, invoke another model, share data with an external service, initiate a transaction or talk to an agent operating in a different jurisdiction. The question becomes far more sophisticated:

Can this agent, acting on behalf of this organisation, use this model, with this data, in this location, for this purpose?

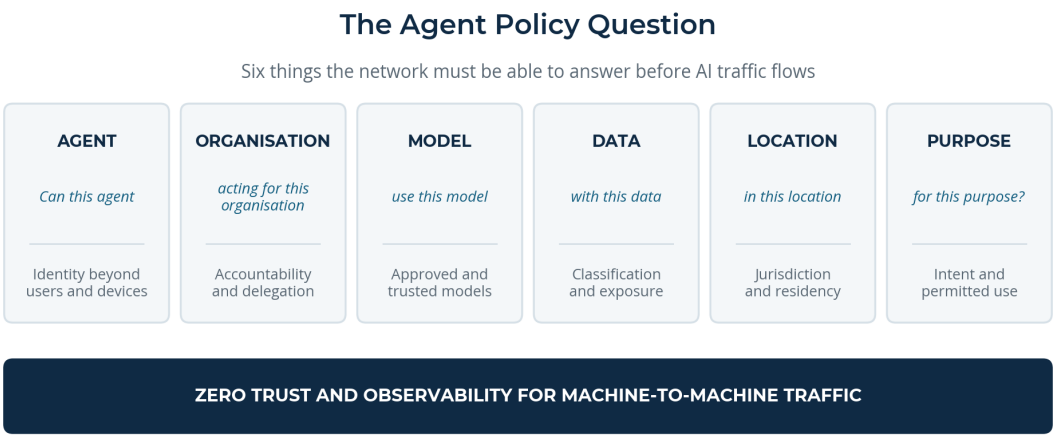

Figure 3. The policy question the AI WAN must be able to answer for every agent interaction.

Identity therefore has to extend beyond humans and devices to agents themselves. Zero Trust becomes relevant to machine-to-machine communication. Observability must tell organisations not only where traffic travelled, but what initiated it and what it was allowed to do.

The WAN is no longer simply transporting AI traffic. It becomes one of the places where AI policy is enforced.

Sovereignty Is Not Just Geography

This matters most for governments and regulated enterprises.

Sovereign AI is often discussed as a question of location: keep the data in the country, operate the infrastructure locally, run the model domestically. Those things matter, but geography alone does not create sovereignty. A locally operated AI environment may still depend on foreign chips, cloud platforms, foundation models, software licences or technologies whose continued availability is controlled elsewhere.

Real sovereignty has several dimensions: where the data resides, where the compute runs, who controls the models and, ultimately, whether an organisation or country can keep operating if an external provider, regulator or government changes the rules.

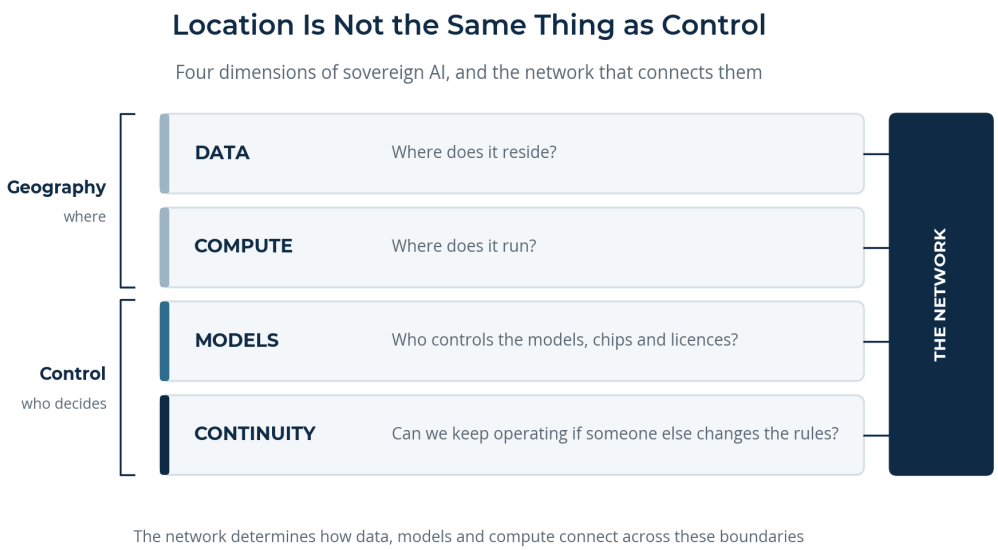

Figure 4. Geography covers where data and compute sit. Sovereignty also requires control over models and continuity.

Location is not the same thing as control.

The network matters here too, because it determines how data, models and compute connect across those boundaries. For a consumer asking an AI assistant a question, this is largely invisible. For an enterprise handling sensitive intellectual property, or a government operating critical infrastructure, it is fundamental. The same AI technology can create very different infrastructure requirements depending on who consumes it.

The Three C's, Revisited: When AI Consumes AI

In our August issue we introduced the **Three C's of AI Adoption**: Capacity, Clarity and Consumption. The AI WAN changes the question behind each of them.

	The question so far	The question in the AI WAN era
Capacity	Is there enough compute?	Can the right compute actually be reached, with the performance the workload needs?
Clarity	Which model or platform fits?	Which combination of model, infrastructure, network and governance fits each workload?
Consumption	How do people and organisations adopt AI?	What happens when the consumer of AI is no longer a human, but another AI?

Increasingly, **AI will consume AI**. In fact, the shift from human to machine-generated activity has already begun. According to Imperva, automated bots accounted for more than 53% of web

traffic in 2025, compared with 47% generated by humans. A year earlier, automated activity had overtaken human traffic for the first time in a decade.[1]

AI is not responsible for all of that traffic, but it is accelerating the shift: Imperva now identifies AI agents as a distinct category of automated activity, alongside traditional good and bad bots.

As agents increasingly interact with models, applications and other agents, machine-driven consumption could create orders of magnitude more interactions than humans typing prompts ever will.

> *"AI may be the newest workload in technology, but the network characteristics it needs are remarkably familiar. The enterprises that scale AI with confidence will be those that treat connectivity as a design decision, not an afterthought."*

Capvion Perspective

The boardroom conversation about AI infrastructure now needs to move beyond how much can be bought, to how everything connects. Three habits will separate the organisations that get this right.

Classify the Workload Before You Buy the Connectivity

Map each AI workload against its latency, jitter, packet loss, availability and sovereignty requirements before choosing between the public internet, direct cloud connectivity and deterministic Ethernet. Overbuying deterministic connectivity wastes money. Underbuying it quietly erodes performance and the business case.

Give Agents an Identity, Not Just a Network Path

Treat agents as first-class identities in security and network policy. Extend Zero Trust to machine-to-machine traffic, and invest in observability that shows what initiated each flow and what it was permitted to do. Designing this in is far easier than retrofitting it once agents are in production.

Design for Control, Not Just Location

Test every sovereignty claim across all four dimensions: data, compute, models and continuity. Ask what happens if a provider, regulator or government changes the rules, and make sure the network gives you options to reroute, rehost or switch. For operators and service providers, this is also where the opportunity sits: offering AI-grade connectivity, policy and sovereignty as a service, not just a pipe.

---

[1] Imperva, a Thales company, 2026 Bad Bot Report: Bad Bots in the Agentic Age (April 2026). The 2025 Bad Bot Report put automated traffic at 51% of web traffic in 2024, the first time in a decade it surpassed human activity.

Conclusion

We have spent the first phase of the AI infrastructure race asking how to build enough intelligence: more GPUs, more models, more data centres and more power. The next question is how we connect all of it.

As inference becomes distributed and agents start communicating with other agents, the WAN must do more than move packets between locations. It must connect the right workload to the right model and compute environment, with the right combination of performance, security, economics and sovereignty.

For years, the WAN quietly connected people to applications. AI is giving it a much bigger job.

The network of the future will connect intelligence to intelligence, and make sure we can trust what happens in between.

About Capvion

Capvion is a strategy and capital advisory firm that works with leadership teams, investors, and entrepreneurs on complex growth, transformation, and value creation challenges.

The firm combines strategic insight with hands-on operational and transaction experience across technology and telecommunications markets, advising companies and investors on corporate strategy, mergers and acquisitions, transaction readiness, and enterprise value optimisation.

Capvion operates as a senior-led advisory boutique, working closely with a network of experienced associates and industry specialists. This model allows the firm to bring in the right expertise at the right moment, combining strategic clarity with deep domain knowledge across transactions, integrations, and operational transformation.

Through its advisory work and publications such as Capvion Insights, the firm shares practical perspectives on how strategy, execution, and capital interact to shape enterprise outcomes.

For more information visit www.capvion.com or contact insights@capvion.com