expereo

Scale tomorrow for your global internet network.

AI & the bandwidth reckoning

www.expereo.com

AI has evolved beyond the story of innovation; it has graduated to an infrastructure reckoning.

Many global enterprises are seeing sustained traffic growth, capacity limits exposed, and networks built for a different era are breaking. Responding incrementally is creating business risks you can't ignore. When the network falls behind, AI initiatives stall, performance degrades, and costs escalate. Transformation slows when it should accelerate.

This guide gives you a structured path to respond before AI demand overwhelms the network. It follows a clear two-stage model:

1. Protect performance today
2. Engineer for scale tomorrow

To protect your network today, you need to follow three critical steps:

- **Identify where demand is rising**
- **Identify the chokepoints**
- **Add bandwidth to prevent degradation**

From there, you need to shift gears and **challenge assumptions,** because more bandwidth alone does not solve for latency, cost, or control.

The final stage is to engineer your network for what comes next:

Modernize the foundation to introduce intelligence and resilience.
Strategically engineer a network designed for sustained AI scale.

This is the shift from reactive upgrades to deliberate engineering:
Add capacity and chase demand.

And redesign your network to keep pace with change.

AI is accelerating. Your network must keep pace.

AI does more than increase traffic.

It **permanently changes** the volume, direction, and sensitivity of that traffic.
It **increases** sustained bandwidth consumption.
It **introduces** latency sensitivity into workflows that previously tolerated delay.
It **creates** burst patterns that overwhelm circuits designed around averages.

Your data already reflects this shift:

30-50%
Traffic growth within one to two years once AI scales

2-5x
Bandwidth increases in strategic areas

20-50%
Sustained baseline uplift following AI rollout

Yet many enterprises continue to forecast 10–20% growth based on historic SaaS and collaboration trends.

That gap manifests as congestion at egress points, oversubscribed cloud on-ramps, and inspection layers operating permanently at peak.

Bandwidth can't be treated like a background utility anymore, it needs to be a primary enabler of AI performance.

Proactive

Organizations that address bandwidth capacity proactively will protect stability and cost control.

vs

Reactive

Those that do not will be forced into reactive upgrades under pressure.

Understand the uplift

AI is rewriting network demand across all industries.

AI is the dominant bandwidth driver across industries right now. Cloud, real-time applications, and edge and IoT architectures amplify the impact.

Agriscience
Precision farming and autonomous operations depend on constant data movement for:

IoT sensor networks	Robotics and autonomous machinery	AI-driven yield modeling and pest forecasting

These environments are seeing 2–5x bandwidth increases in key zones to unlock productivity gains of up to 9%.

Banking
Digital finance is equally demanding of:

Real-time AI fraud detection	GenAI advisory platforms	Embedded finance and instant payments

Institutions expect 30–50% traffic uplift within two years.

The hidden surges CIOs miss
The biggest risk is not the obvious workload. It is the secondary effect.

East-west traffic often **doubles or triples** faster than forecast	Cloud egress can add **20–80%** to bills if not architected for locality	AI inference permanently lifts baseline network usage by **20–50%**

AI does not spike traffic. **It resets it.**

Identify the chokepoints

Where bandwidth breaks first.

Bandwidth pressure does not cause failures everywhere at once. It fails at critical chokepoints.

Common failure points

1. Internet egress
2. Branch access
3. Cloud on-ramps
4. Security inspection layers

These were designed for SaaS, video, and email. Not distributed AI workloads.

The symptoms

1. Video freezes and pixelates
2. VoIP becomes robotic
3. SaaS slows to a crawl
4. Backups overrun windows
5. Replication fails

This is not cosmetic.

The results

1. Monitoring delays
2. Backups fail
3. Replication lags
4. Security visibility narrows

Bandwidth constraints are not just UX problems.

Create the headroom

The enterprises that win with AI increase bandwidth early.

The immediate and necessary step is to create headroom by purchasing additional bandwidth.

AI workloads are burst-heavy and latency-sensitive. If capacity expansion lags production deployment, queuing, jitter, and packet loss become inevitable.

Practical actions include:

1. Expanding cloud on-ramp bandwidth
2. Upgrading circuits in AI-intensive sites
3. Reducing oversubscription at inspection points
4. Rebalancing branch versus core capacity

IT leaders should prioritize 95th percentile utilization over averages.

Averages routinely conceal burst-driven congestion and performance risk.

IT leaders should conduct 3–12 month trend reviews and forecast headroom before AI initiatives go live. Waiting for visible congestion is already too late.

Expanding bandwidth protects performance in the near term. It stabilizes production rollouts and reduces immediate operational exposure.

But capacity alone does not address the architectural shift AI introduces.

Challenge the assumptions

Bandwidth buys time. Architecture buys control.

The AI and bandwidth misconceptions that reveal a broader network reality:

Misconceptions	2026 reality
We'll buy more when we need it	That works for soft upgrades. New circuit lead times mean capacity often arrives after AI workloads have already outgrown it.
Average utilization is enough	Bursts as short as 5–15 minutes cause queuing, jitter, and packet loss. The 95th percentile is the value below which 95% of all bandwidth utilization measurements fall over a given period, making it a realistic way to represent sustained peak usage while ignoring rare, short-lived spikes.
Enough bandwidth equals low latency	Even 50ms of extra latency can damage AI inference, trading, or real-time user experience.
The cloud handles scaling	Egress and interconnect fees scale rapidly. Many enterprises see 20–100% increases in cloud costs.
Our tools will warn us	Legacy monitoring misses AI burst patterns. Pressure builds quietly until users complain.

Bandwidth expansion is a necessary and important first step.

But enterprises must also think strategically about future-proofing their overlay and underlay.

Modernize the foundation

Thinking ahead strategically: Modernize the overlay and underlay.

AI requires a shift from reactive capacity expansion to deliberate network engineering.

Overlay

Overlay optimization introduces control and intelligence:

- SD-WAN-driven performance management
- Dynamic routing across multiple paths
- Policy-based workload steering
- Latency-aware path selection for inference and real-time traffic

Underlay

Underlay modernization improves physical and carrier foundation:

- Smarter carrier selection based on performance
- Geographic route optimization
- Multi-carrier resilience
- Traffic engineering designed for distributed AI workloads

This is how enterprises prevent the recurring cycle of emergency upgrades and uncontrolled cloud cost expansion.

Bandwidth creates headroom. Architecture creates advantage.

Build the 5-year advantage for your global internet

A two stage strategy

The challenge is not if AI will impact your network, but how prepared you are when it does.

You need a dual approach.

1

Protect today

The first priority is protecting performance in the present with additional bandwidth.

That means adding headroom, addressing chokepoints, and scaling capacity in high-growth zones.

2

Engineer tomorrow

The second priority is engineering the network for the next five years.

That involves optimizing the overlay for intelligent routing, and architecting for sustained east-west and multi-cloud expansion and redesigning the underlay for resilience and performance consistency.

Without the second step, the first becomes recurring operational spend.

Expereo supports both speeds of transformation.

Expereo enables rapid bandwidth expansion to stabilize AI growth.

We provide intelligent overlay control and proactive monitoring aligned to 95th percentile performance realities as well as design and optimize global underlays across multiple carriers.

The objective is not to just increase capacity; we work with you consultatively to ensure your network scales in line with AI ambition, without sacrificing control, visibility, or financial predictability.

Talk to an expert

expereo