AI Impact on Wide Area Networks

Cisco Report 2026

Executive summary 4

Key Findings 4

Takeaways 7

Introduction 8

Data analysis and findings 9

- Data volume growth 9
- Data flow length 11
- Data flow rate 12
- Traffic asymmetry 13
- Network latency impact 14
- Agentic AI traffic patterns 15

Looking ahead 18

- Enterprise usage of agentic AI and network impact: 2026-2035 18
- Consumer Usage of AI and Agentic AI and network impact: 2026-2035 21

About the report

The AI Impact on Wide Area Networks report represents the key findings and takeaways from a research initiative led by Cisco Systems, Inc. The report combines real-world traffic analysis (using the assurance tool Cisco Crosswork Assurance User Experience), third-party industry data, and Cisco controlled lab tests of AI agents. It explores how AI-driven traffic differs from traditional non-AI web traffic, with a focus on the impact of agentic AI traffic and the implications for network design, capacity planning, and assurance.

What sets the report apart is it's based on real-world traffic data including an early lens on agentic AI traffic (currently small but growing fast) that lets us start to see and measure a new class of AI network traffic and understand the implications. The report includes:

- Direct measurement of live AI inference traffic across service provider networks
- Empirical tests of AI traffic characteristics to train models that identify and track AI flows with precision
- A repeatable measurement framework and baseline to track AI traffic evolution and forecasting on an annual basis

Executive summary

The rapid adoption of artificial intelligence (AI) —particularly **AI inference** and **agentic AI systems**—is beginning to reshape network traffic patterns in measurable and structurally important ways. By 2027, 80% of executives believe their company's competitive survival will depend on agentic AI,[1] and consumer usage of AI is already prevalent and accelerating. This is driving a fundamental shift in how traffic is generated, distributed, and experienced, with implications for service providers and enterprises that manage large-scale networks.

Key Findings

The future of connectivity is no longer just about faster downloads—it's about the vital "spinal cord" of intelligence between agents and AI models—a critical dependency that's forging new paths of communications that impacts how we manage networks. While the industry has spent decades optimizing the network for human-paced, bursty video streams, the rise of agentic AI is changing network traffic profiles and behavior. By 2035, one-quarter of network traffic is projected to be AI inference (Source: Cisco model). These flows don't behave like the web. They live longer, demand more upstream capacity, and operate at software speed, not human speed. The connectivity between agent logic and AI models effectively becomes the agent's "spinal cord"—a critical dependency whereby any network degradation directly impairs agent functionality.

[1] Cisco report The Race to Agentic AI: Why Infrastructure will make or break workforce integration https://www.cisco.com/c/dam/en/us/solutions/collateral/artificial-intelligence/race-to-agentic-ai-report.pdf

Token consumption data

▲ **10x**

year-over-year growth

AI inference traffic is small today—but is growing extremely fast.

While AI inference traffic remains negligible compared to dominant categories like video streaming, observed growth rates are exceptional. **Token-consumption data shows nearly 10x year-over-year growth**, while in some service provider measurements we are seeing around 4x growth in just eight months. Sustained growth at these rates means AI traffic will become a meaningful component of overall network traffic by 2035.

AI inference flows

▲ **2x**

longer

Flow rate for web vs. AI traffic

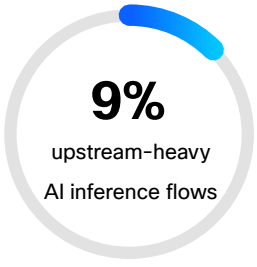
▲ **10x**

larger

AI inference traffic flows behave differently at the transport level vs. typical web transactions.

- AI inference flows last approximately **2x longer**, mainly due to token-by-token content generation.
- Median flow rates of **regular web transactions are 10x larger than AI inference traffic flows** which have smoother sustained throughput, rather than bursty delivery.

These characteristics will require **flow-aware network and security systems**, including firewalls, intrusion detection, and Deep Packet Inspection (DPI), to handle greater scale, distribution, and state efficiency over time.



Traffic symmetry status quo is changing.

AI inference—and especially agentic AI—introduces significantly more upstream traffic due to large, context-rich prompts.

- Approximately **9% of AI inference flows carry more upstream than downstream traffic**, versus about 0.5% for typical web traffic.

This trend is expected to intensify as agents maintain and transmit growing contextual state, with direct implications for **access planning and upstream capacity modeling**.

#s ~ 5 seconds for AI inference latency vs. ~ 20–50 milliseconds for network latency

Network latency is not the dominant bottleneck—yet.

End-to-end AI inference latency is driven primarily by model processing, typically ranging from hundreds of milliseconds to several seconds. Network latency is currently small enough and does not yet justify large-scale architectural shifts toward edge inference purely for latency reasons. Inference hardware improvements will lead to network latency becoming more relevant, and we are already seeing proof points of that. Measuring and assuring the AI inference experience is becoming critical for perceived service quality.

AI agents

▲ **450%**

more total traffic

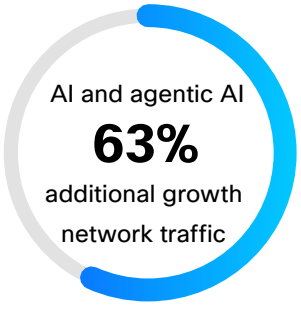
70%

of that traffic is AI inference

Agentic AI fundamentally amplifies network usage.

AI agents act as **network "power users,"** operating at software speed rather than human speed. Empirical testing shows:

- Up to **450% more total traffic** is generated per task when performed by an agent.
- Approximately **70% of that traffic is AI inference**, making inference paths mission-critical.
- The connectivity between agent logic and AI models effectively becomes the agent's **"spinal cord"**—a critical dependency whereby any network degradation directly impairs agent functionality.



AI traffic impact by 2035*

Enterprise networks:

Without agentic AI, enterprise traffic is projected to grow by about 2.5x over the next decade. With agentic AI adoption, traffic growth could increase by about **9x**, driven by autonomous task execution and inference-heavy workflows.

Consumer networks:

Consumer AI adoption is accelerating and will reach near-universal usage by the mid-2030s. AI and agentic AI are projected to increase consumer-driven network traffic by about **6.6x**, representing about 63% additional growth compared to non-AI scenarios—making this the dominant driver of overall internet traffic expansion. AI inference traffic is projected to be 25% of total network traffic by 2035.

*Source: Cisco model

Takeaways

- AI and agentic AI will not just increase traffic volume—they will **change traffic shape**, symmetry, duration, and criticality.
- AI inference paths will become **strategic network assets**, requiring high levels of resilience, observability, and differentiated treatment, for example, Quality of Service (QoS) and path security.
- Traditional planning assumptions (burstiness, downlink dominance, human-paced interactions) will need to adapt to **the new reality of AI and agentic AI traffic** that lasts longer, demands more upstream capacity, and operates at software speed.
- Continuous measurement of **AI-specific Key Performance Indicators (KPIs)** and **adaptive network architectures** will be essential to sustain expectations for performance and user experience.

This report marks the first step in ongoing research on the impact of AI on network traffic. As AI adoption deepens across enterprises and consumers, future research work will refine projections, validate emerging patterns, and guide operators in evolving their networks for an AI-driven world.

Introduction

Adoption of AI and AI-based applications such as AI agents is growing at a very fast pace. In a Cisco report with Omdia, 80% of executives view agentic AI as critical to company survival by 2027[2]. Other industry reports confirm similar growth projections for both enterprises and consumer AI applications. In this context, service providers and organizations running large-scale networks need to evaluate how the proliferation of AI is going to impact networks in terms of traffic growth, patterns, or other characteristics.

This report is the first in an annual series where we will share key findings and implications to help network leaders make decisions on the evolution of their network infrastructure. Cisco has initiated research on AI traffic patterns and how they differ from non-AI web-based transactions, with a specific lens on agentic AI traffic and applications. The data presented in the report is based on real traffic data analysis from Cisco Crosswork Assurance User Experience, Cisco empirical tests, and third-party reports.

[2] Cisco report The Race to Agentic AI: Why Infrastructure Will Make or Break Workforce Transformation. https://www.cisco.com/c/dam/en/us/solutions/collateral/artificial-intelligence/race-to-agentic-ai-report.pdf

Data analysis and findings

Data volume growth

The first questions network operators are trying to answer are “How much AI traffic will exist?” and “How will this traffic impact our network?”

AI traffic, in this case, refers to AI inference traffic which flows across the WAN between the device requesting model inference and the data center or location where the model is hosted and the inference happens. Inside data centers, terabytes of traffic flows are happening during AI model training, but that traffic is not crossing the WAN, except for large data transfers for AI training purposes.

The actual volume of AI inference traffic is still negligible compared with other types of traffic, such as video streaming, and will likely remain very low in relative terms in the near term. However, there are very significant growth rates:

- Token consumption growth of 10x year-over-year in OpenRouter AI from December 2024 to December 2025.)[3]

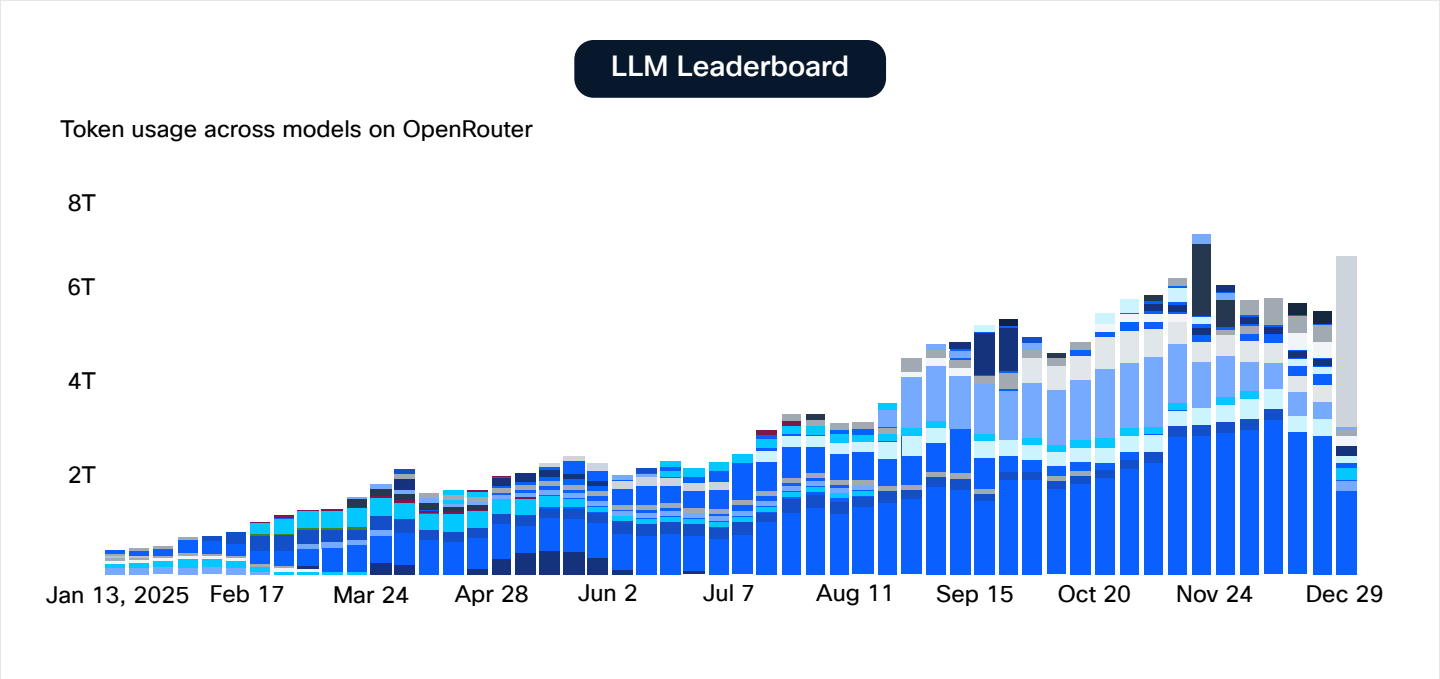


Figure 1. OpenRouter token consumption CY2026

Source: Openrouter.ai

[3] OpenRouter, LLM Leaderboard. https://openrouter.ai/rankings#leaderboard

Traffic analysis with data from two service providers shows consistent traffic growth for AI inference over a similar period: **4x traffic growth over a period of eight months** of traffic monitoring.

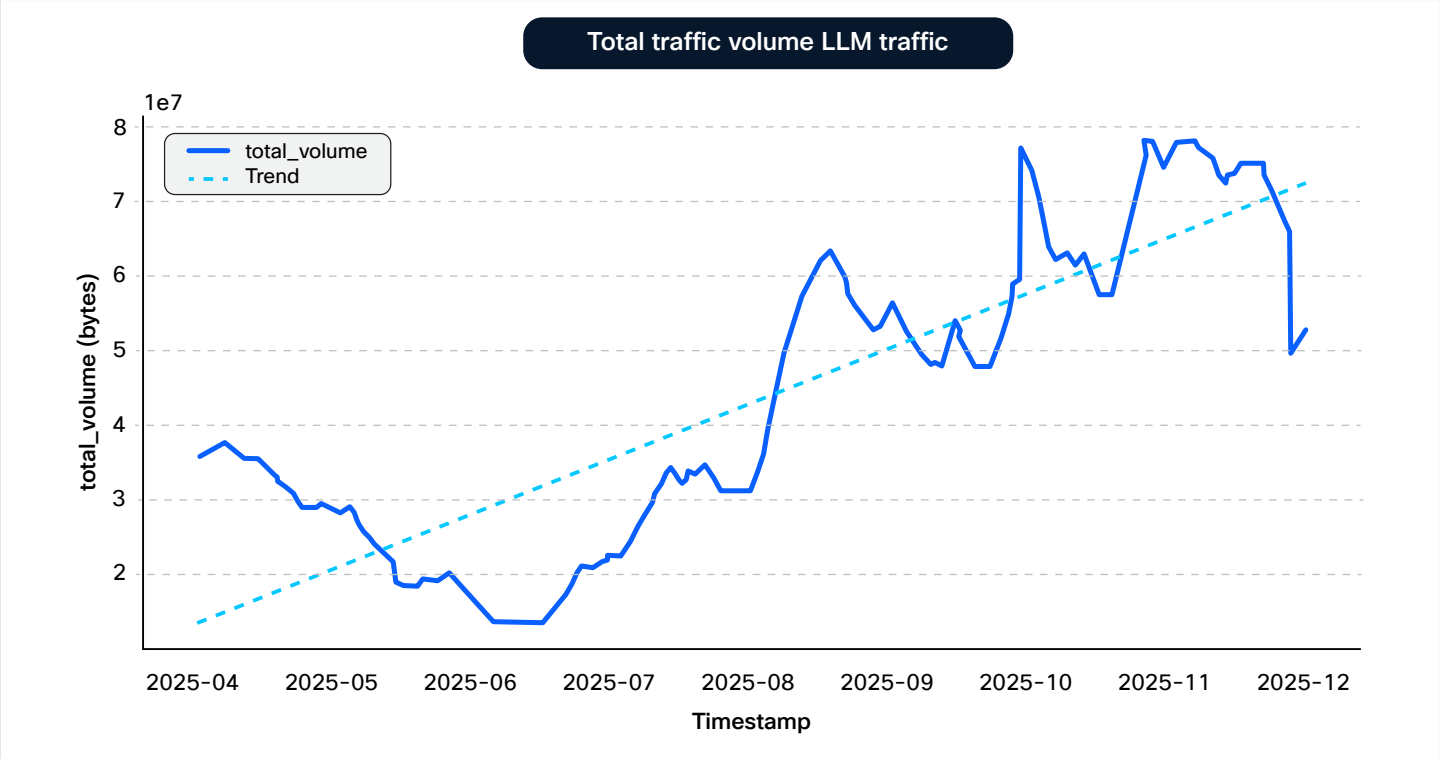


Figure 2. AI inference traffic volume evolution

Source: Cisco data

AI inference services are using both Transmission Control Protocol (TCP)- and Quick UDP Internet Connections (QUIC)-based transport. The split in terms of number of flows is almost 50%, while in terms of data volume QUIC represents 57%.

Implications: While AI inference traffic is still low in absolute terms, sustained high growth rates will result into a relevant portion of all traffic over time. Clearly, the adoption of AI by applications and users is driving fast growth. The adoption of QUIC as transport for AI inference presents encryption challenges for DPI systems to detect traffic at the application layer.

Data flow length

AI inference flows present different characteristics compared to non-AI web traffic. While not dramatic, these differences may impact capacity planning. Analysis of data flow length shows differences between AI inference flows versus regular web transactions.

Statistically, AI inference flows last 2x longer than regular web transactions.

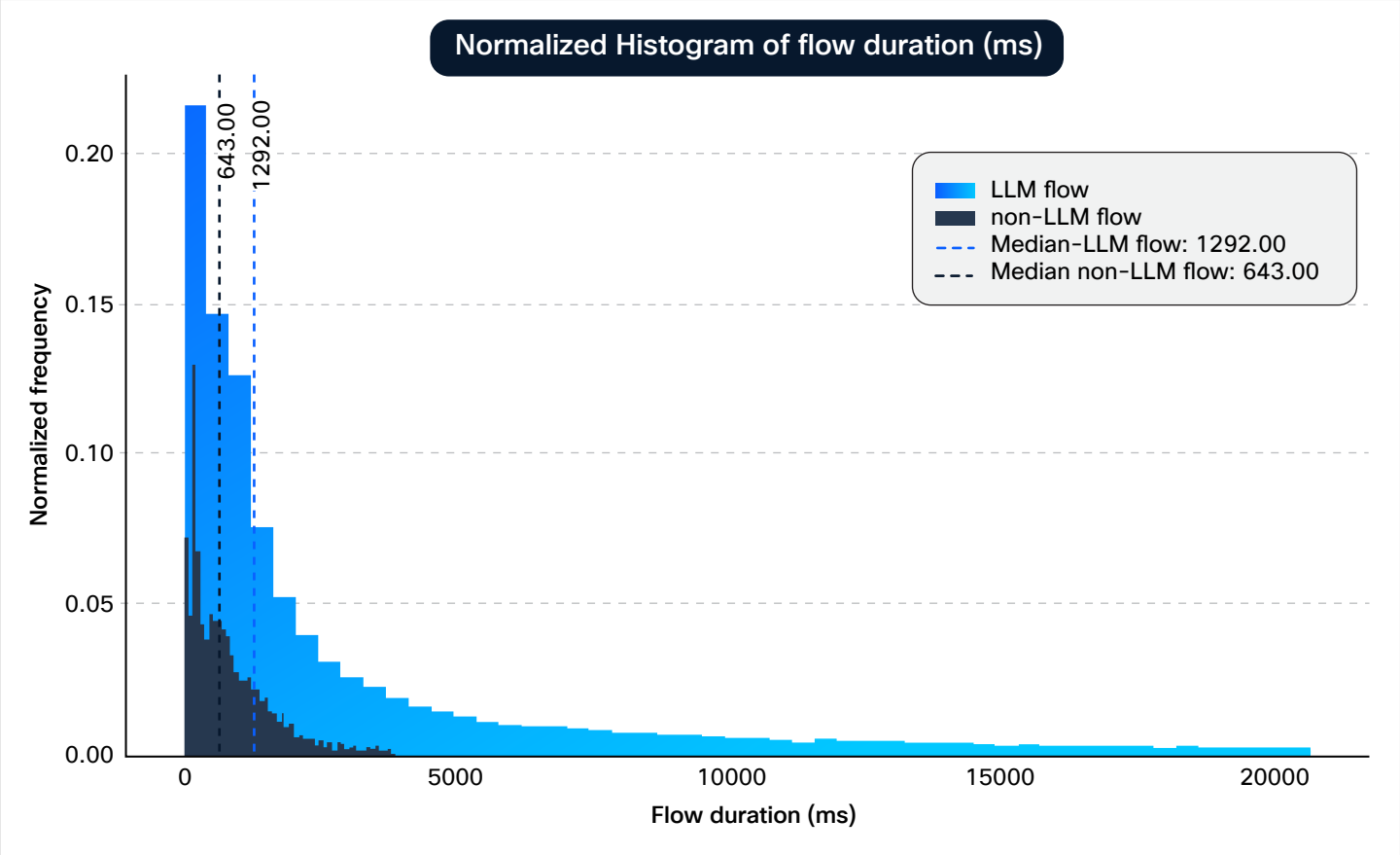


Figure 3. Flow duration distribution of AI inference flows vs. non-AI web transactions

Source: Cisco data

The main driver is the way AI inference traffic generates content, one token at a time, resulting in longer and lower rate flows compared to other flows.

Implications: For "flow-aware" network systems that must keep state for flows in tables, the proliferation of AI inference flows that last longer means growing flow tables will need to be effectively planned. Over time, security and flow-aware network systems are likely to become more distributed to cope with forwarding state growth, similar to the approach used by Cisco Hypershield™.

Data flow rate

Related directly to the flow length, the data flow rate for AI inference flows shows a different pattern compared to regular web transactions. **Median flow rate is 10x larger for regular web transactions compared to AI Inference flows**.

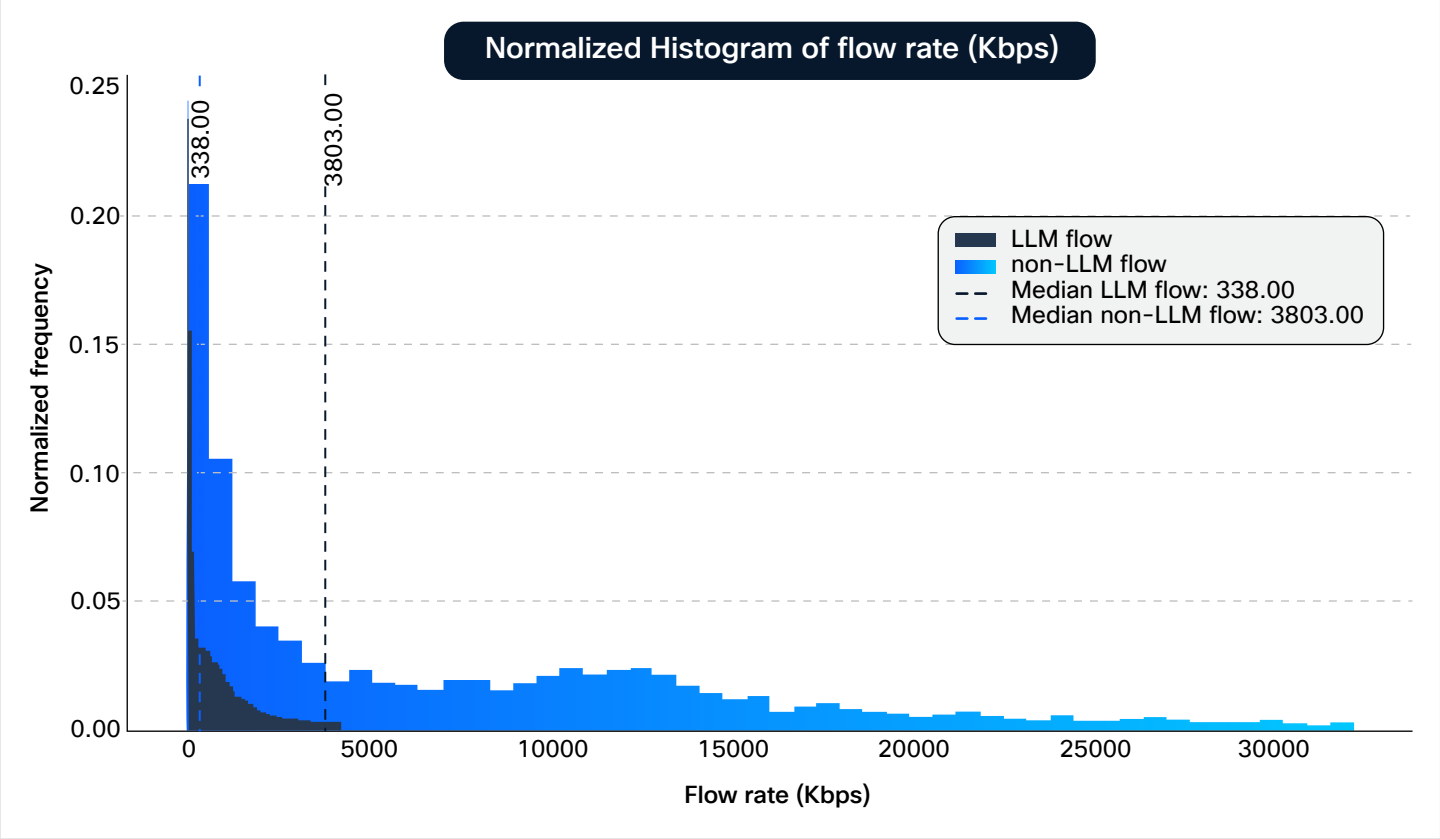


Figure 4. Flow rate distribution of AI inference flows vs. non-AI web transactions

Source: Cisco data

The main reason for this difference is the process that generates the data in AI inference. A data flow that occurs **one token at a time** is like having a “shaper” applied to the data flow. Regular web traffic flows peak much higher, as the content can be retrieved from wherever it is stored and delivered to the user.

Implications: Different median average-traffic and peak-to-average rates may require different QoS settings in the network to manage AI inference traffic versus non-AI inference web transactions.

Traffic asymmetry

Potential changes in traffic asymmetry are expected with AI inference traffic. Network traffic asymmetry varies depending on access type (mobile vs. wireline networks) and services. Mobile network traffic tends to be more upstream heavy due to social networking, with more content sent upstream compared to wireline networks.

Analysis of AI inference flows versus non-AI web transactions shows clear differences in traffic symmetry. As Figure 5 shows, in AI inference traffic, **9% of flows have more upstream traffic than downstream traffic** compared to other HTTP transactions, where this occurs in only about 0.5% of the flows.

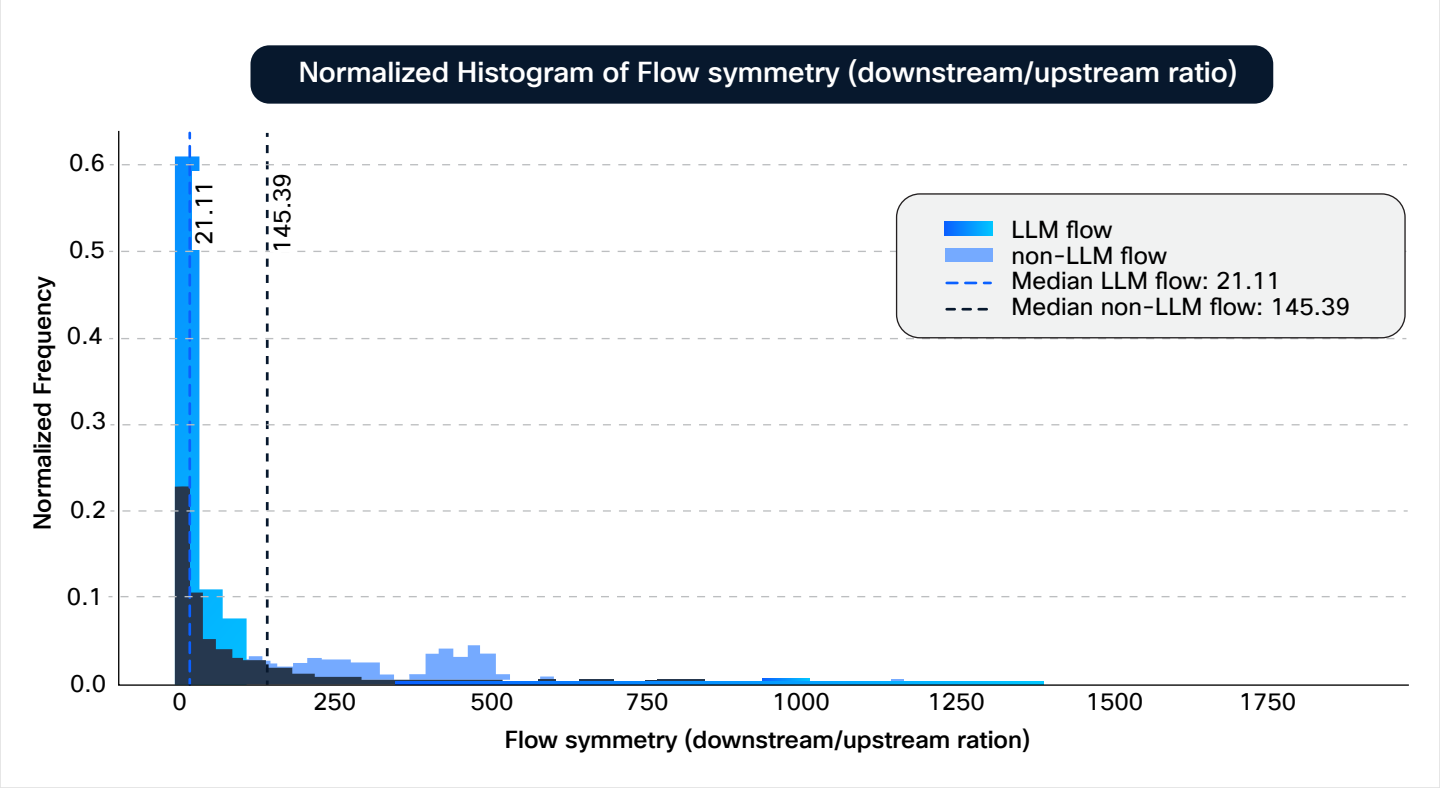


Figure 5. Traffic flow symmetry of AI inference flows vs. non-AI web transactions

Source: Cisco data

We foresee this trend further **consolidating and growing**. The reason for this difference is the fact that AI inference requests often capture a growing amount of context in the prompt in order to produce a relevant response. As adoption of agentic AI grows, more significant changes to network traffic symmetry patterns can be expected. Agents will capture continuous context as tasks are executed, and this state needs to be inserted into the prompts.

Implications: Particularly for radio capacity planning, traffic symmetry assumptions are a relevant factor. As adoption of AI inference and agentic AI grows, it will be important to track the evolution of traffic symmetry, as it will continue decreasing over time.

Network latency impact

Large Language Model (LLM) inference requests tend to have significantly higher latency than typical web API calls, and the response times can be more variable. Traditional web application REST APIs often strive for subsecond or even sub-100-millisecond (ms) response times. For instance, **a good API response time is typically below 100 ms** in many web scenarios. By contrast, even short LLM queries incur response times of hundreds of milliseconds just to begin producing an output, and full responses often take **seconds**. LLM latency can be modeled as follows:

Latency = Time To First Token (TTFT) + Time Per Output Token (TPOT) × Number of output tokens

TTFT is the initial overhead to process the prompt, and TPOT is the generation time per token. This means latency grows with output length. For example, OpenAI GPT-5.2[4] might take on the order of 9 ms per token, and Anthropic Claude Opus 4.5 about 20 ms per token. A **100-token GPT-5.2 answer could take about 1 second**, and a maximum-length 600-token answer could approach 5.4 seconds. There is wide variability among LLMs, model providers, implementations, and the hardware used for inference.

Implications: AI inference latency varies widely from a few hundred milliseconds to multiple seconds. Network latency is still at this point a very small component of the total end-to-end inference latency. Inference hardware is, however, evolving and becoming much faster. As inference accelerates, the network latency component plays a more relevant role on the end-to-end perceived experience. Network latency will become a key factor for inference distribution, combined with others like scale, data sovereignty, and security. Additionally Service Providers will need to monitor the actual AI inference latency that the customer is experiencing, as it's a key factor in perceived user experience. With Cisco® Crosswork Assurance, it is possible to measure key AI inference KPIs across different network endpoints, along with network and service KPIs.

[4] LLM Leaderboard—Comparison of over 100 AI models from OpenAI, Google, DeepSeek & Others. https://artificialanalysis.ai/leaderboards/models

Agentic AI traffic patterns

In addition to AI Inference flows, agentic AI as an AI application deserves a special focus due to its transformational impact and the implications for networks.

AI agents are software processes that use an **AI model as a brain** to make decisions. Then, depending on the type of agent, multiple tools are used to complete specific tasks or goals.

There are three main sources or destinations of data in an agent (see Figure 6):

- The agent itself (where the agent code runs)
- The AI model (where AI inference happens)
- The tools and/or sources of data for the agent

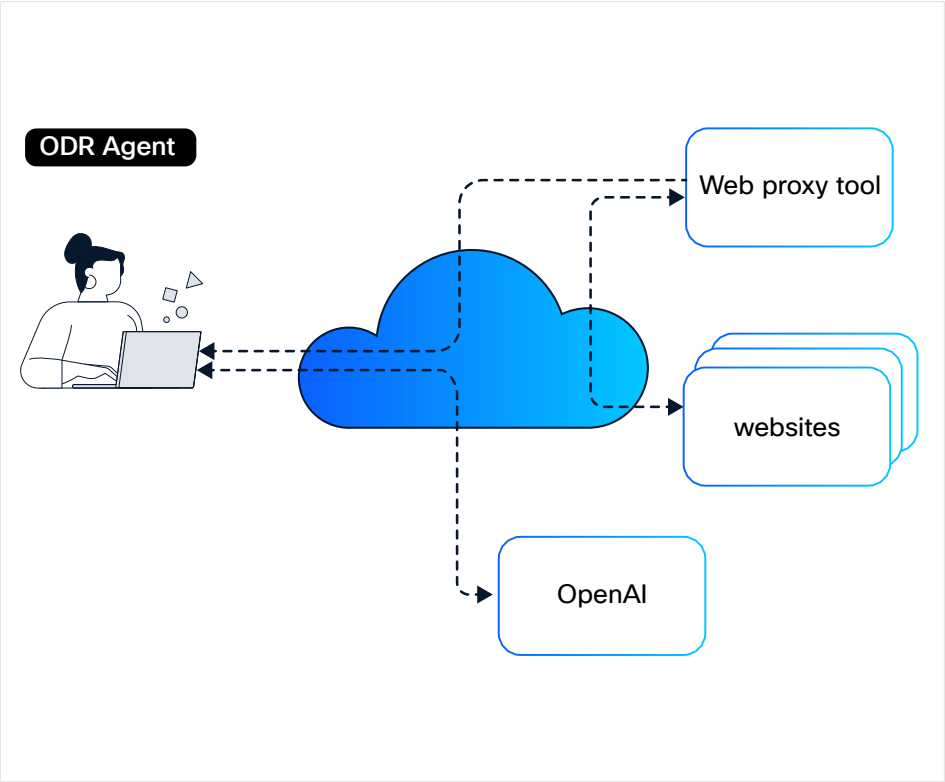


Figure 6. AI Agent testing environment

Source: Cisco

For example, we are using an Open Deep Research agent from Langchain5 (other agents will behave similarly). Let's analyze what happens when a given task is requested to the agent (see Figure 7).

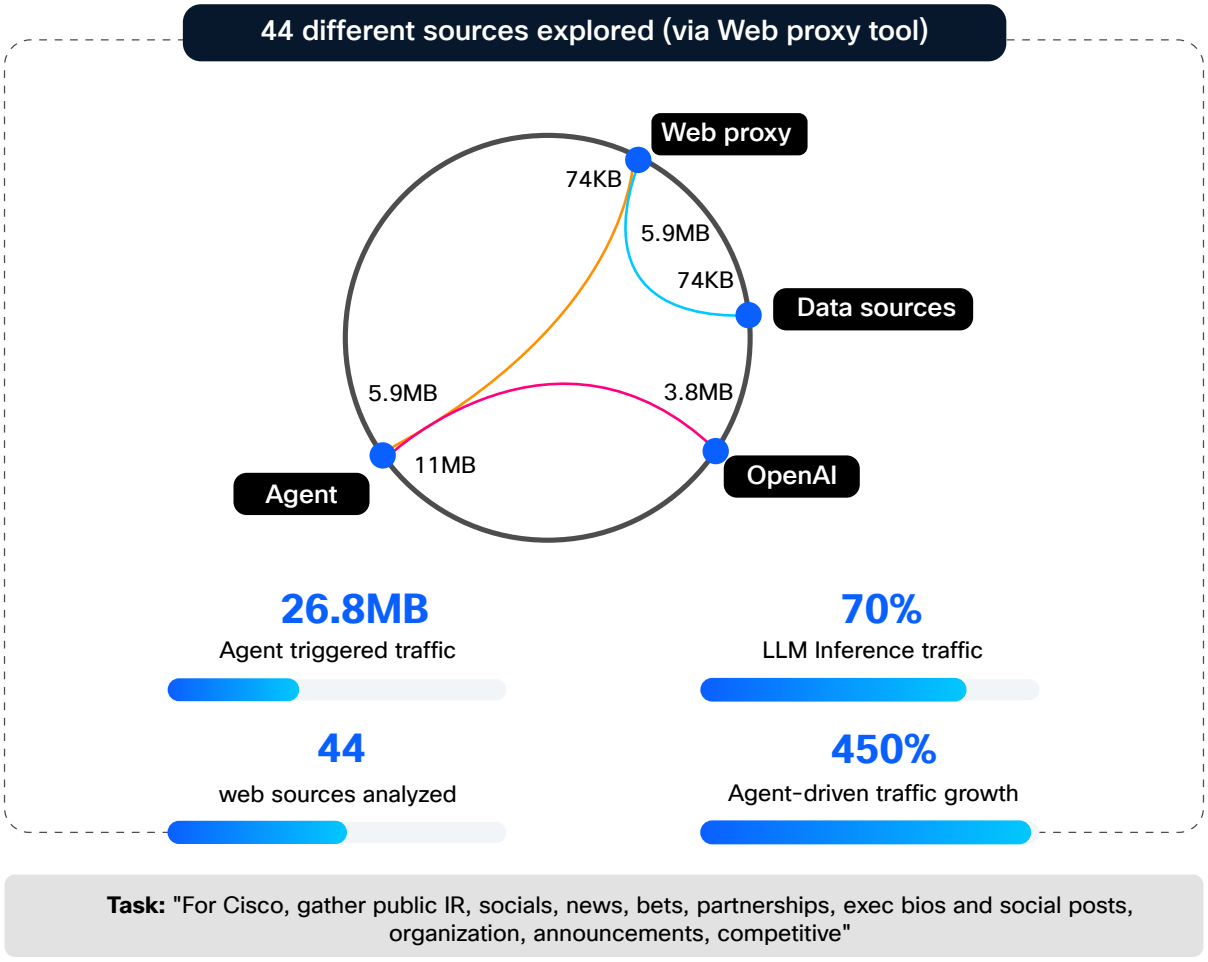


Figure 7. AI agent traffic flows

Source: Cisco

A request to an agent generates a significant amount of network traffic between the different components of the agent. Overall, the network traffic generated or triggered by the agent is much larger than the network traffic generated if the same task was performed manually by a human. In our example there is a **450% increase in traffic** associated with using an agent, and notably, **70% of that new incremental traffic is AI inference**.

The following are key considerations related to AI agent behavior and its interaction with and dependency on the network:

- AI agents act as network **power users**. Agents consume cloud and network resources at **software speed**, compared to when these tasks are performed manually at **human speed**. This results in a significant increase in network traffic.
- AI inference traffic is not only a significant portion, but the critical portion of traffic that is necessary for the agent to operate. The network path between the agent code and the AI model can be considered the agent's **'spinal cord'**. Any failure or constraint in that communications path will significantly impact the agent's ability to operate.

[5] Open Deep Research, https://github.com/langchain-ai/open_deep_research/tree/main

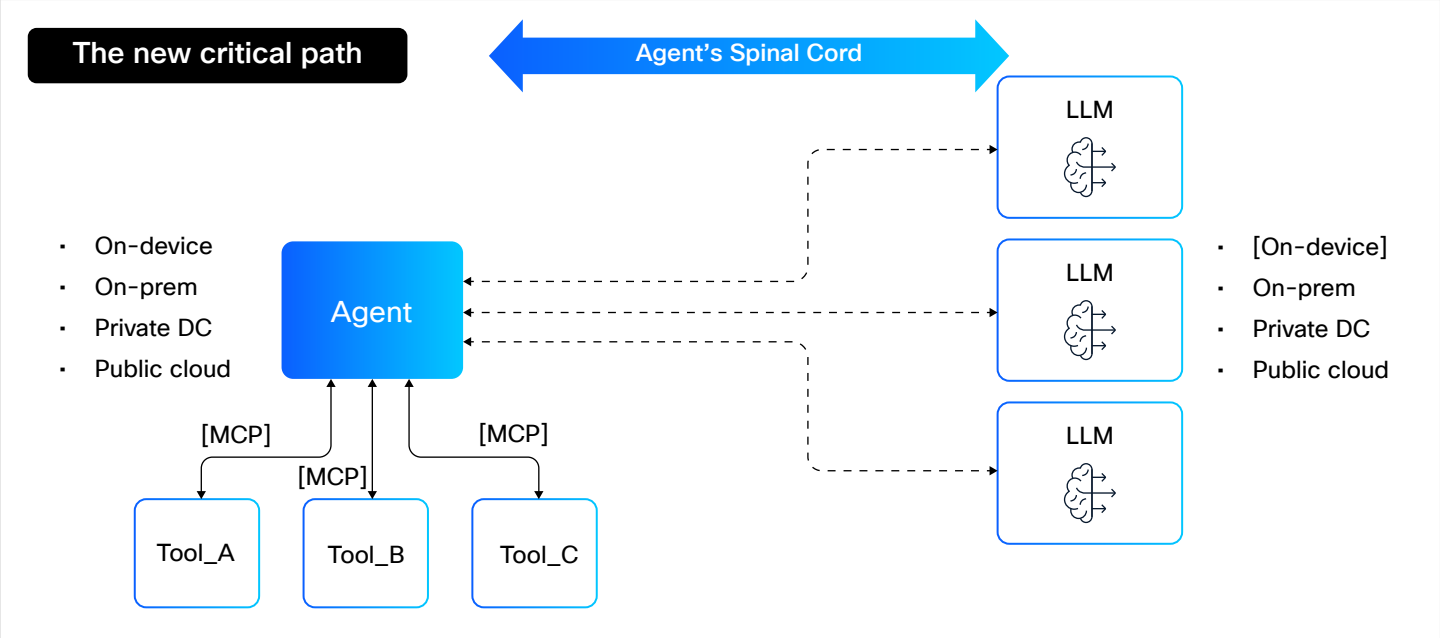
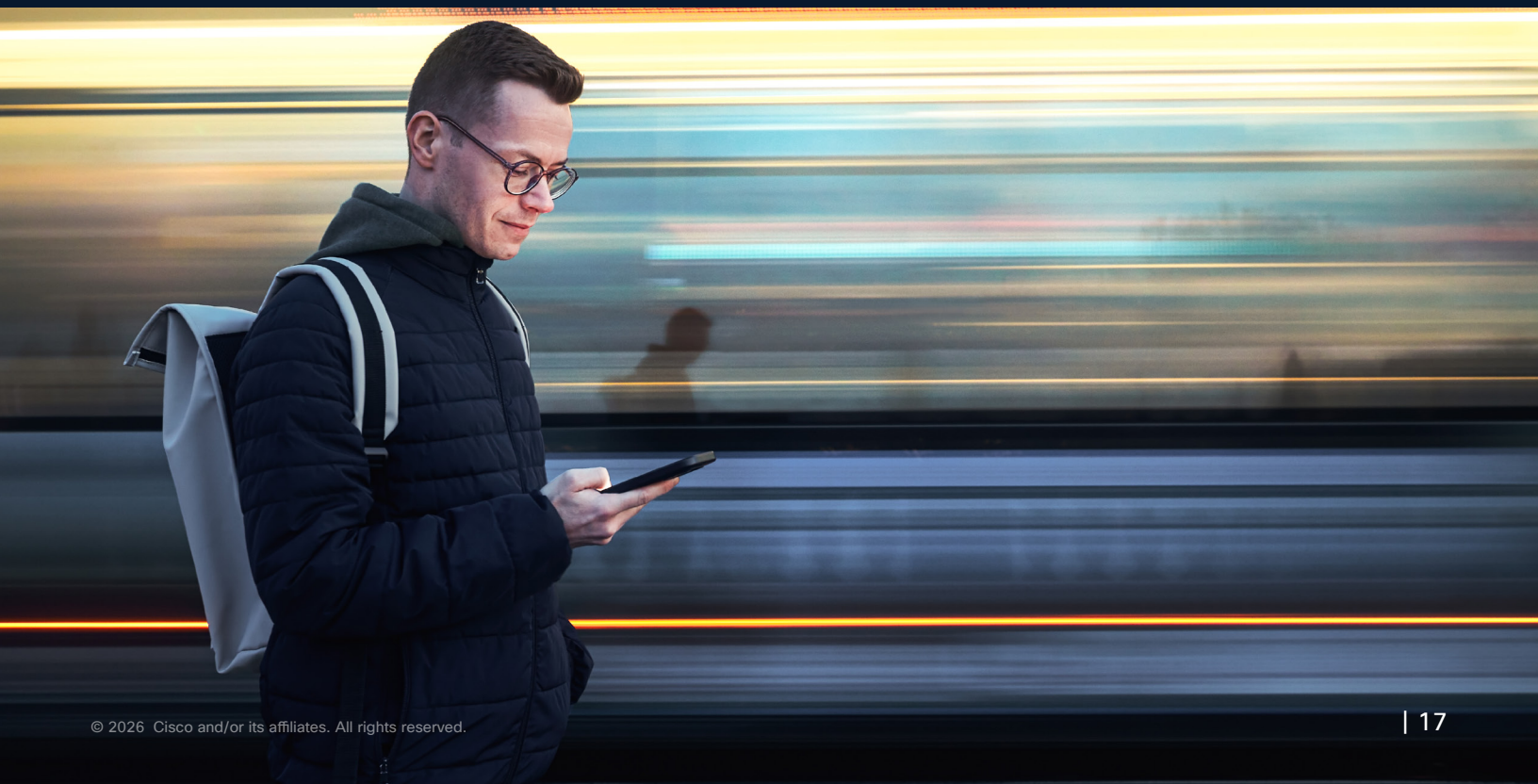


Figure 8. AI Agent traffic pattern and new critical path

Source: Cisco

Implications: As agentic AI adoption grows in both enterprise and consumer applications, network traffic will grow accordingly, primarily driven by the adoption of agents in user tasks. In that process, AI inference traffic will grow significantly as it is a large percentage of AI agent-driven traffic. Resiliency at the network level for communication paths is critical for AI inference across the agent 'spinal cord' between the agent code and the model.



Looking ahead

This report explores the behavioral differences and special characteristics of AI inference traffic and agentic AI to understand the impact the traffic can have on the network.

Our next step is to envision what the future of network traffic may look like as consumer and enterprise adoption increases for these technologies.

This report provides projections through 2035. While long-term modeling involves inherent uncertainty, the objective is to support service providers and network operators with extended planning horizons. As additional data becomes available and models are refined, updated projections will be issued in future reports.

Enterprise usage of agentic AI and network impact: 2026–2035

Agentic AI, with autonomous, action-taking AI agents, is only at the early stages of enterprise usage, but momentum is rapidly building. Analysts note that adoption in 2024–2025 was nascent (fewer than 5% of enterprise applications had AI agents in 2025[6]), yet organizations across industries are already planning for widespread implementation. Key projections for the mid-2020s include:

By 2026

40%

of enterprise applications will include integrated, task-specific AI agents

- **By 2026:** Gartner forecasts that **40% of enterprise applications will include integrated, task-specific AI agents**, up from less than 5% in 2025. This reflects a rapid shift from simple chatbots toward more capable agents embedded in software tools (for example, AI-driven assistants in Customer Relationship Management [CRM], Enterprise Resource Planning [ERP], and IT service management applications).

By 2027

80%

believe their company's survival will depend on agentic AI.

- **By 2027:** Surveys show adoption accelerating year-over-year. A Cisco Report, shows that by 2027, 80% of executives believe their company's competitive survival will depend on agentic AI.[7] In IBM's 2025 global executive survey, 24% of business leaders said they already have AI agents taking independent action in their operations—and **67% expect to have AI agents autonomously making decisions in workflows by 2027.**[8] Even in regulated domains, executives anticipate that agents will automate significant work (for example, an average of **29% say they will automate risk and compliance tasks by 2027** in leading firms).

---

[6] Gartner press release, Gartner Predicts 40% of Enterprise Apps Will Feature Task-Specific AI Agents by 2026, Up from Less Than 5% in 2025.

[7] Source: Cisco report The Race to Agentic AI: Why Infrastructure will make or break workforce integration.

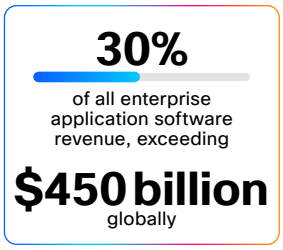
[8] IBM, Agentic AI's Strategic Ascent.

These early projections underscore that **agentic AI usage is moving from experimental to strategic.** Organizations are laying the groundwork through investments in generative AI platforms and automation, which will pave the way for deploying autonomous agents.[9] Cost reduction and productivity gains are key motivators for early adopters (39% cite these as primary drivers). By 2027, many enterprises will have progressed from planning and pilots into active use of AI agents in day-to-day operations.

The use of AI agents will become commonplace in business processes. Gartner analysts predict that by 2028, **"one-third of all user interactions with software will shift from traditional apps to agentic AI-driven interfaces,"**[10] illustrating how pervasive these agents will be in everyday workflows. In customer operations, Gartner expects that **80% of routine customer service issues will be handled autonomously by AI agents by 2029**.[11]

Looking further ahead, industry experts anticipate **continued acceleration of agentic AI adoption into the 2030s**, with these AI "co-workers" becoming ubiquitous across business domains.

Key forecasts and expectations for the mid-2030s include:



- **Enterprise Software Revolution:** Gartner projects that by 2035, agentic AI will drive approximately **30% of all enterprise application software revenue**, exceeding **$450 billion** globally (up from just 2% of software revenue in 2025).[12] In essence, nearly one-third of the enterprise software market might be attributable to AI agent capabilities by 2035 – a radical shift in a decade. At that stage, most enterprise software is expected to have AI agents deeply embedded, and new software business models will revolve around autonomous functionality.
- **New Normal of Work:** By 2035, the enterprise environment will likely feature humans and AI agents working side by side in most business processes. Gartner describes a democratized future state where employees routinely create or customize AI agents on-demand, and autonomous agents orchestrate complex workflows across departments.

[9] Omdia, New Omdia Analysis Shows Agentic AI Outpacing Growth Rates of Traditional Generative AI, September 2025.

[10] Gartner press release

[11] Gartner press release

[12] Gartner press release

By the mid-2030s **agentic AI is no longer viewed as an emerging technology, but rather as a standard component of enterprise operations**. Companies slow to embrace autonomous AI may find themselves at a competitive disadvantage. Overall, the consensus is that **the 2020s are the adoption decade for agentic AI, and the 2030s will be the era of its maturation and full integration** into the fabric of business.

Based on the data captured from industry reports, the following captures a projection model for agentic AI penetration in enterprise tasks from 2026 to 2035.

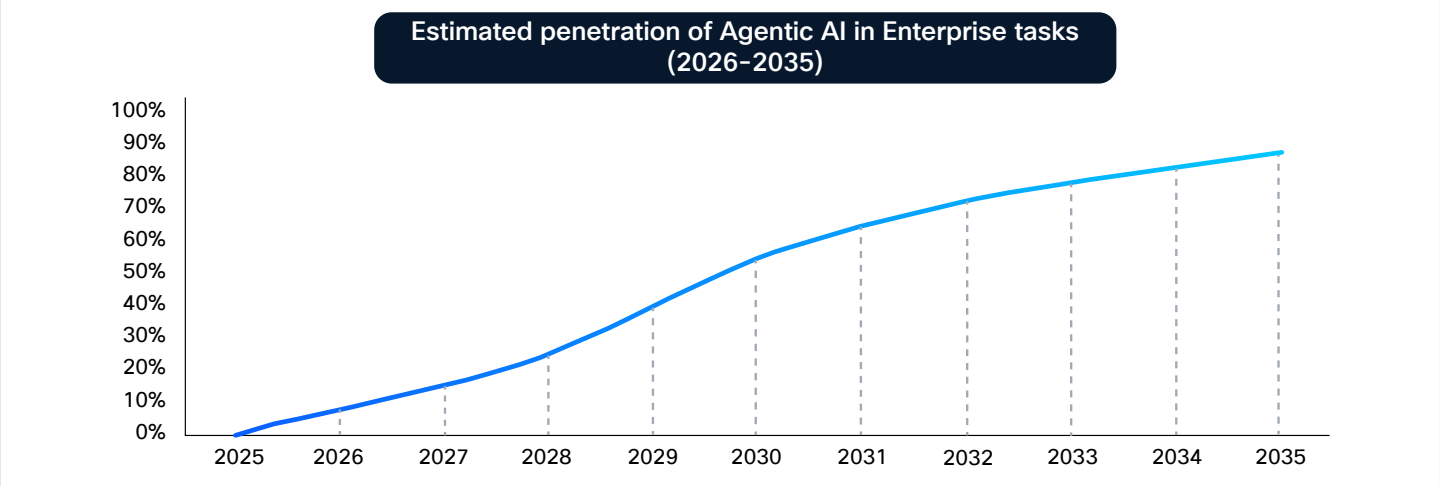


Figure 9. Projected adoption of agentic AI in enterprise tasks

Source: Cisco model

Considering the network implications of using AI agents to automate tasks and the forecasted agentic AI penetration, Figure 10 shows the overall impact on network traffic driven by enterprise services and agentic AI adoption from 2026-2035.

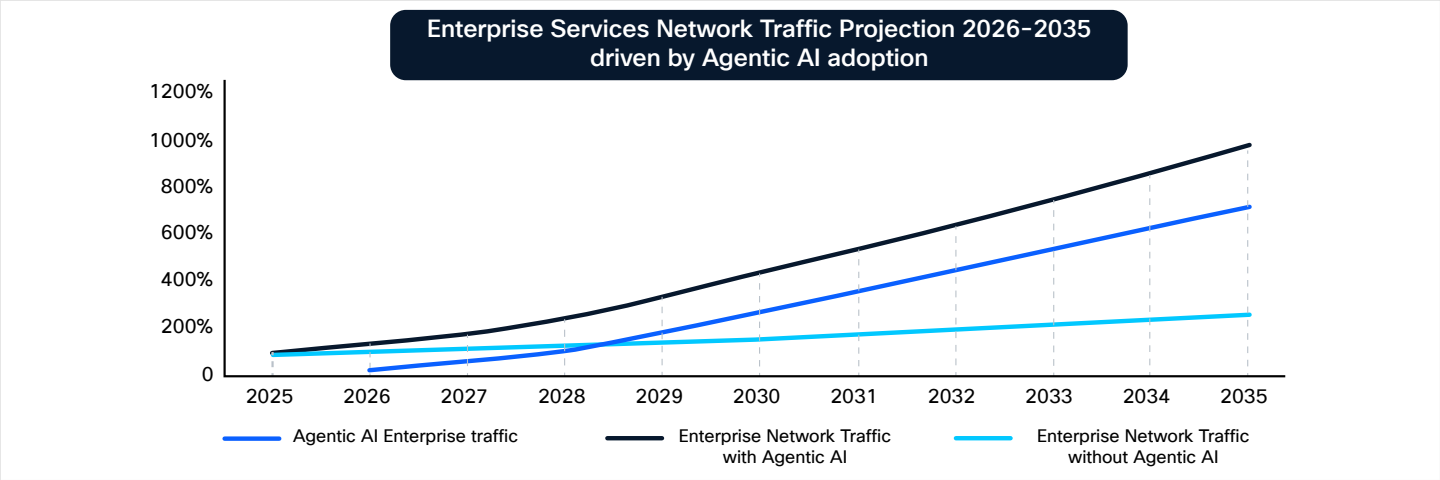


Figure 10. Projected impact of agentic AI in enterprise services network traffic (% increase)

Source: Cisco model

Without agentic AI, enterprise traffic would be expected to grow by about 250% over the period 2026-2035. However, **agentic AI will bring significant disruption and impact to enterprise network usage, driving 9x traffic growth.**

Consumer Usage of AI and Agentic AI and network impact: 2026–2035

AI is rapidly becoming embedded in everyday consumer life and is poised for an even more dramatic surge in adoption. Key drivers include **generative AI chatbots and assistants** (such as ChatGPT)**, AI in gaming and entertainment, AI-powered image and video generation, smart apps for shopping and banking,** and the rise of **augmented reality (AR)** experiences that rely on cloud AI inference. Agentic AI with autonomous AI agents that can act on a user's behalf is also emerging in personal applications.

61%

of adults in the U.S. had used AI in the past six months

Consumer adoption of AI reached a tipping point in the mid-2020s. Surveys in mid-2025 found that **61% of adults in the U.S. had used AI in the past six months,** with about **19% using AI daily.**[13] Scaled globally, this equates to roughly **1.7 to 1.8 billion people** worldwide who have accessed AI tools, including **500 to 600 million daily users**. This explosive growth has come just a few years after the advent of mainstream generative AI (ChatGPT launched in late 2022), indicating an unprecedented pace of consumer technology adoption.

46%

of all consumer transactions in the U.S. by 2030

By 2030, many analysts project that AI will be as common in daily life as smartphones are today. Gartner suggests that AI will be deeply integrated into daily activities; for instance, by 2028 a large share of business purchases may be handled by AI agents,[14] and the consumer realm is on a similar trajectory. Cognizant and Oxford Economics forecast that **AI-driven purchasing will account for about 46% of all consumer transactions in the U.S. by 2030**.[15] In other words, nearly half of consumer spending could be influenced or executed by AI (with similar figures of 46% in Germany and 55% in Australia).

By the early **2030s**, we can expect most consumers to

regularly rely on AI

By the early 2030s, we can expect **most consumers to regularly rely on AI**—either consciously (such as asking a chatbot or using an AI assistant) or behind the scenes (AI-powered features in apps, appliances, and vehicles). As AI becomes more trusted and integrated, occasional users today are likely to become daily users by 2030. Menlo Ventures analysts expect consumer AI adoption to ultimately **catch up to enterprise AI** in both usage rates and economic value, given the broader reach into the general population.

By **2035**, projections tend to converge on the assumption that AI will be nearly ubiquitous for anyone using connected technology. Although precise penetration percentages for 2035 are scarce, it's reasonable to expect **global consumer AI usage to approach close to 90% to 100% of internet users by the mid-2030s,** effectively meaning **billions of people interacting with AI daily**. AI will likely become a default layer in devices and services—much like broadband or electricity—largely invisible, yet indispensable.

---

[13] Menlo Ventures, 2025: The State of Consumer AI.

[14] Networkworld, AI's Dark Side Shows in Gartner's Top Predictions for IT Orgs, October 2025.

[15] Cognizant, Cognizant Study Shows Consumers Who Embrace AI Could Drive $4.4 Trillion in Spending Over Five Years, January 2025.

By 2030

23%

of Americans had made a purchase using AI in the past month

A major trend in this period is the rise of **agentic AI**, where autonomous AI agents can proactively perform tasks for users. These range from **personal shopping bots** that find and buy products for you to AI assistants that manage schedules, finances, or home devices with minimal oversight. Industry futurists predict that by 2030, personal AI assistants will act as consumer proxies—screening offers, negotiating deals, and automating decisions. Morgan Stanley reports that as of late 2025, roughly **23% of Americans had made a purchase using AI in the past month**.[16] The report expects **"agentic shoppers" to account for 10% to 20% of all e-commerce spending in the U.S. by 2030.**

By 2030

$3 to $5 trillion

of global retail revenue could be orchestrated by AI agents

Globally, the impact of agentic consumer AI could be enormous. McKinsey research forecasts that by 2030, **$3 to $5 trillion of global retail revenue could be orchestrated by AI agents** that anticipate needs and execute purchases autonomously.[17] This would be a seismic shift in how consumers find and buy products—a change on the order of the original e-commerce revolution but potentially occurring at a much faster rate. Agentic AI isn't limited to shopping; similar autonomous assistance includes personal finance (AI bots optimizing investments or budgets), travel (AI agents planning trips end to end), and even health (AI assistants scheduling doctor visits or managing medications).

By 2030

a majority of consumers worldwide will likely use AI assistants and tools as routinely as they use the internet

If even a fraction of that agent-mediated paradigm extends to consumers, by 2030, a typical individual might routinely rely on an AI agent to handle a host of daily tasks, from grocery replenishment to filtering incoming information. In short, **agentic AI adoption**, while nascent today, is slated to become a mainstream aspect of consumer life in the 2026–2035 timeframe, essentially giving people a "digital personal assistant" that is continuously working on their behalf.

The device landscape for consumer AI is also broadening. The smartphone remains king in sheer numbers, but AI is permeating our homes, wearables, cars, and emerging AR interfaces. In 2026–2035, this multidevice proliferation ensures that AI will be accessible almost anywhere, anytime. The net effect is that adoption rates will climb not just because of personal preference, but because the infrastructure and devices around us will be inherently AI-enabled.

Between 2026 and 2035, AI is set to transition from a novel technology to a universal utility in the consumer world. Both general AI usage and agentic AI adoption are projected to grow exponentially, reaching billions of people and touching all demographics globally. By 2030, **a majority of consumers worldwide will likely use AI assistants and tools as routinely as they use the internet**—with younger generations leading the way and older generations steadily following. Agentic AI, in the form of personal autonomous agents, is expected to handle an increasing share of our digital and physical tasks, from shopping to scheduling, heralding a new era of convenience (and raising new questions around security, trust, and control as well as bias in models).

[16] Morgan Stanley, Here Come the Shopping Bots, December 2025. https://www.morganstanley.com/insights/articles/agentic-commerce-market-impact-outlook

[17] QuantumBlack, AI by Mckinsey, The Agentic Commerce Opportunity: How AI Agents Are Ushering in a New Era for Consumers and Merchants, October 2025.

Physical AI will have a major impact on network infrastructure as robotics technology advances. Broader commercial adoption across industrial and consumer segments will increase reliance on AI inference. Many systems will combine embedded, real-time inference for immediate actions with cloud-based reasoning capabilities. For this analysis, robots can be viewed as physical agents equipped with tangible tools rather than purely digital ones. From a network standpoint, we expect the aggregate impact to be comparable to that of digital agents. This is a critical area to monitor closely in future reports.

Crucially, these adoption trends are not happening in isolation; they coincide with advancements in devices, such as AR wearables, and deeper integration of AI into every app and service. Therefore, consumer AI penetration can be visualized not as a simple linear increase, but as an accelerating curve approaching saturation. Reports and forecasts consistently point to **high double-digit growth rates in AI usage and market size through the 2020s.**[18] If those trends hold, by 2035 we will live in a world where **nearly everyone, across all age groups and regions, regularly relies on AI**—Whether through direct interactions or invisible assistance embedded in their environment. In effect, AI will be ambient: an ever-present copilot for the average consumer.

Based on the data from industry reports, Figure 11 shows a projected model for AI and agentic AI adoption in consumer tasks from 2026-2035.

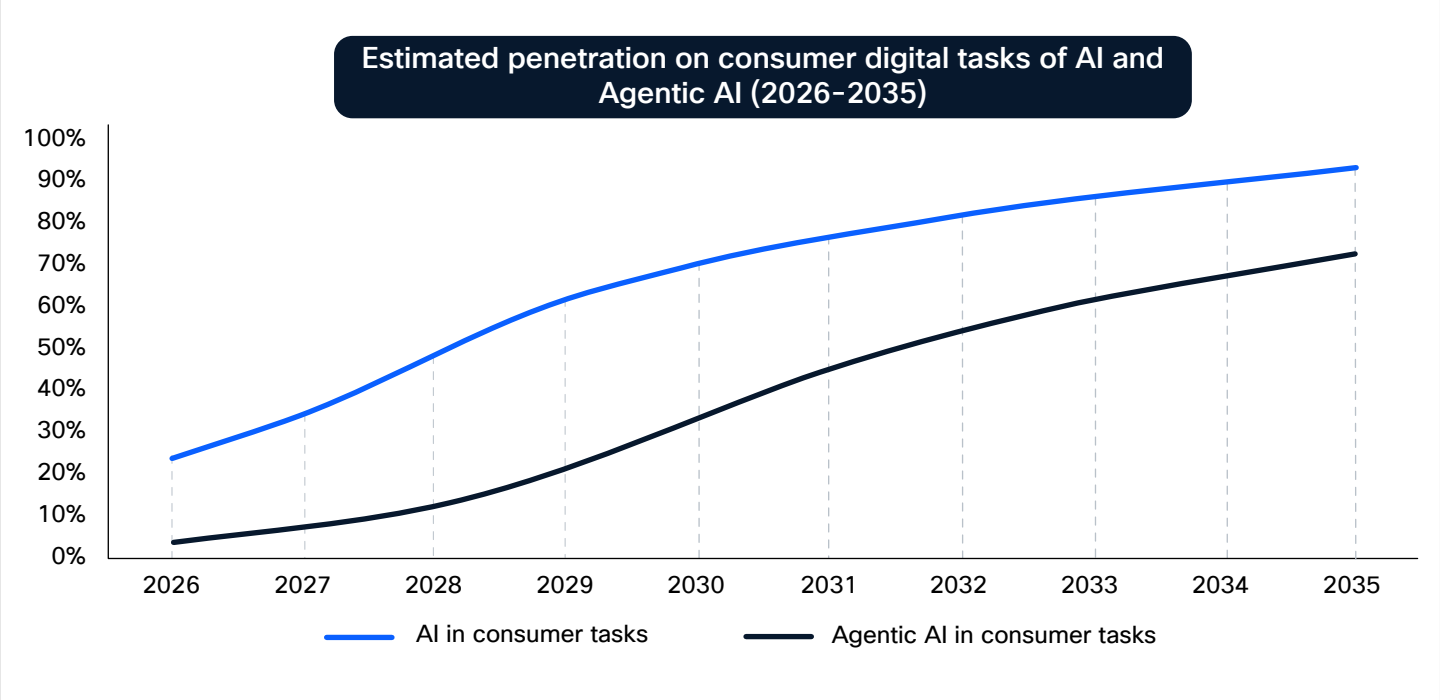


Figure 11. Projected adoption of AI and agentic AI in consumer digital tasks

Source: Cisco Model

[18] GlobeNewswire, Voice Assistant Application Market Report 2025-2035: Industry Revenues to Grow from $8.1 Billion in 2025 to $153.5 Billion by 2035, December 2025.

Not all consumer tasks or digital services have an equal impact on the network, and not all of them will be equally impacted by AI or agentic AI. The model in Figure 12 shows the aggregate impact on overall network traffic driven by the adoption of AI and agentic AI over the period 2026–2035.

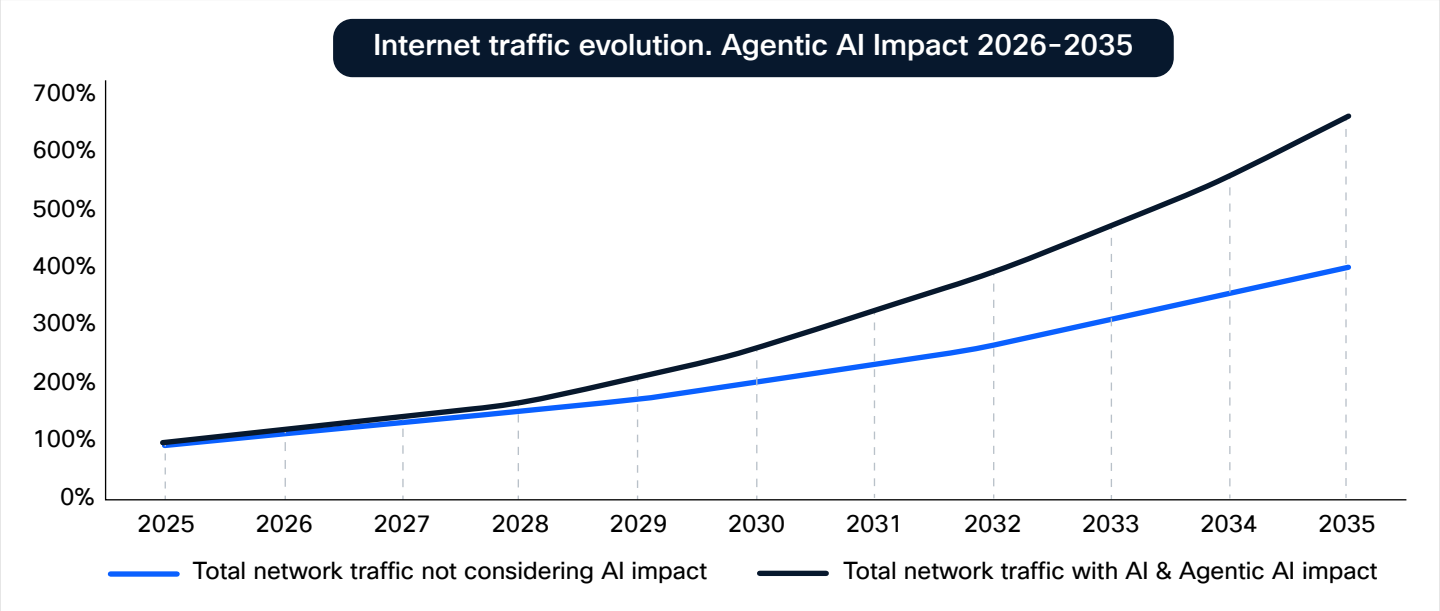


Figure 12. Projected impact of AI and agentic AI in internet network traffic

Source: Cisco model

The implications are significant, as consumer traffic is the primary contributor to overall internet and service provider network traffic. Without the impact of AI or agentic AI, network traffic is expected to grow 4x over the period 2025–2035, driven by the adoption of digital services. However, with the adoption of AI and agentic AI across many of those digital services, **traffic is projected to grow 6.6x from today's levels, or a 63% additional increment compared with the non-AI-impacted projection.** A relevant portion of that traffic will be **AI inference, which is expected to account for 25% of total traffic by 2035.**

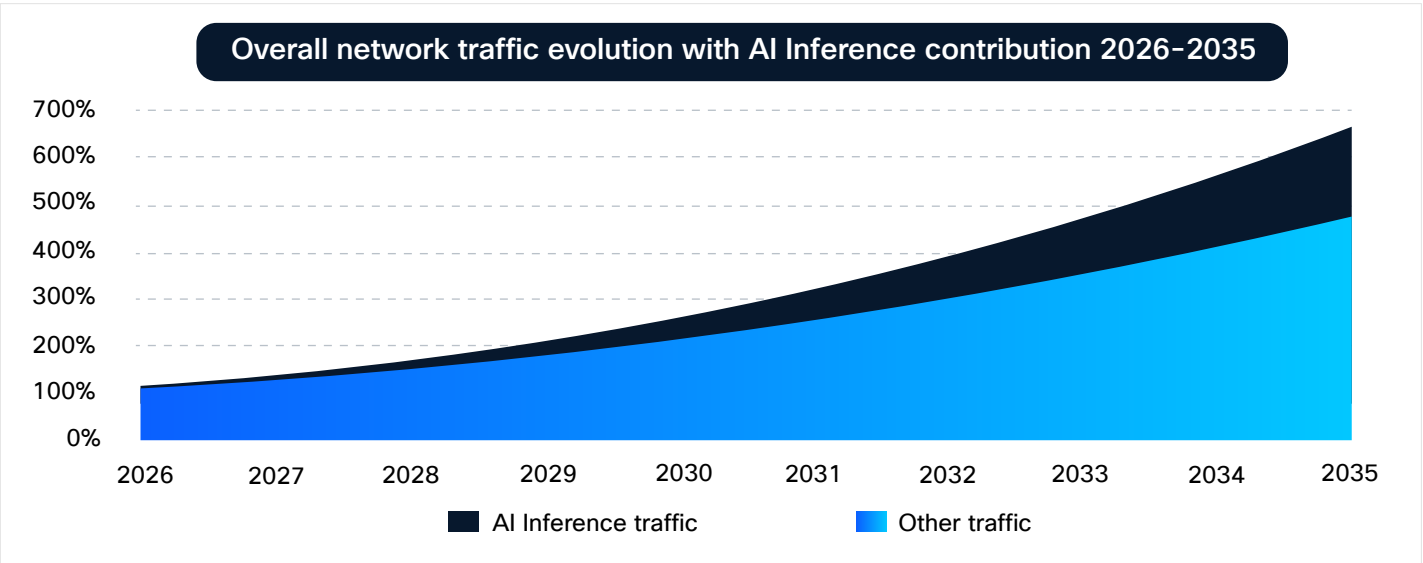


Figure 13. Projected network traffic growth, AI inference traffic versus other traffic

Source: Cisco model

Total **network traffic will experience significant growth** over the period 2029–2032, when adoption of agentic AI will experience a more pronounced increase, with a compound annual growth rate **(CAGR) of around 25% in AI inference traffic**.

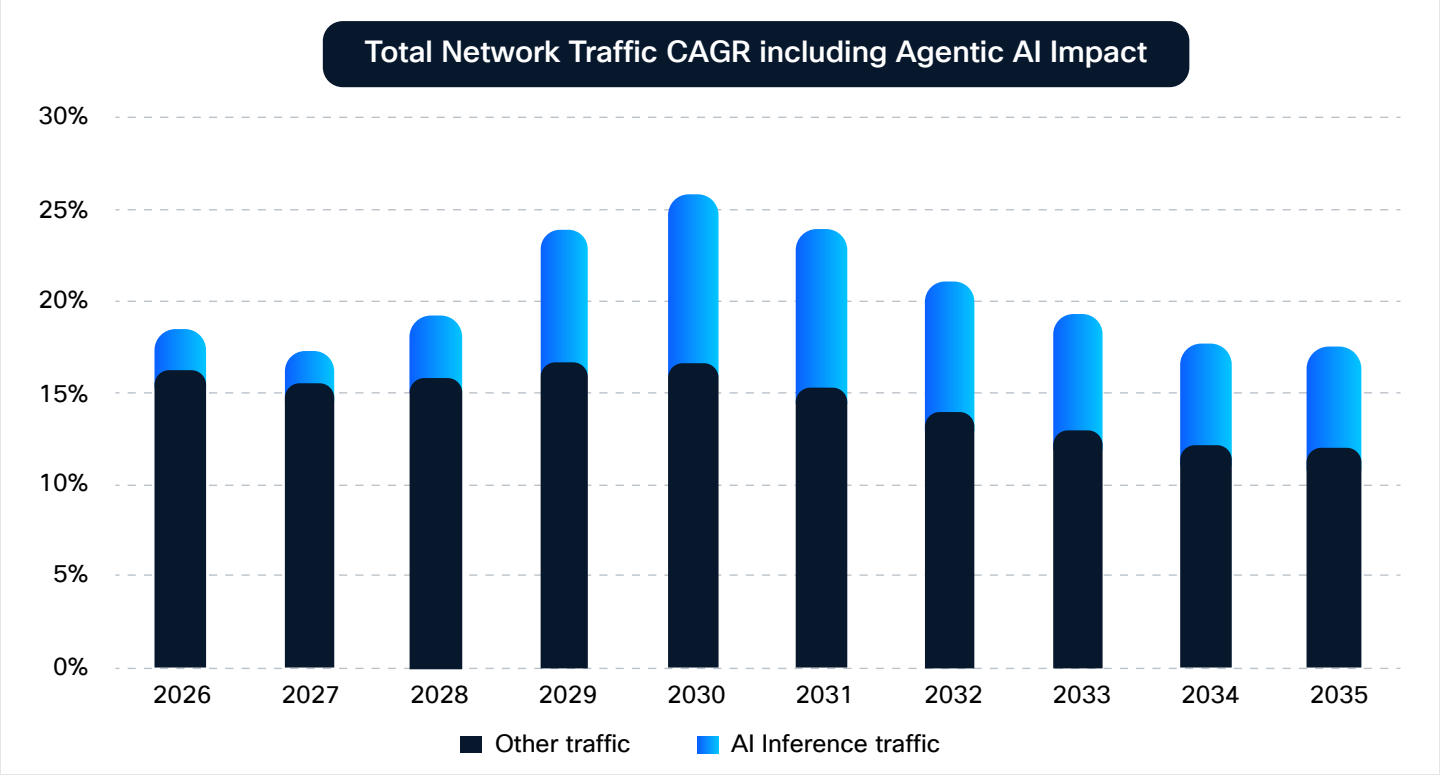


Figure 14. CAGR for both AI inference traffic and other traffic over the target period

Source: Cisco modeling

As adoption of AI and agentic AI continues in both enterprise and consumer digital services, **the transformation ahead is significant in terms of network traffic volume, with new implications for traffic symmetry, flow duration, and resiliency for critical paths.**

Further reading: Cisco report 2026 The Race to Agentic AI: Why Infrastructure Will Make or Break Workforce Transformation https://www.cisco.com/c/dam/en/us/solutions/collateral/artificial-intelligence/race-to-agentic-ai-report.pdf.

This research has been conducted in close collaboration with Opanga Networks.

Report Authors:
Johan Gustawsson (Cisco)
Javier Antich (Cisco)
Waris Sagheer (Cisco)

Contributors:
Ben Hadorn (Opanga Networks)
John Burnette (Opanga Networks)
Ryan Johnston (Opanga Networks)

C99-5828901-00 05/26