



ANM TECH DAY

AI for Networking, Networking for AI

Speaker Introduction

Quick Bio



Anthony Budetti

Solution Architect
Albuquerque. NM

- **Second year at ANM**
- **9 Years experience in IT**
- **Albuquerque native**
- **Specialize in route/switch/firewall/wireless**
- **Attended Cisco networking academy at CNM**
- **CISSP, CCNA, Azure Associate**



Today's Agenda

ACT I — NETWORKING FOR AI

- My AI Story
- AI Basics + Use Cases
- MCP & Agentic AI
- Why Networks Must Evolve
- Cisco Networking for AI
- Palo Alto for AI
- Broader Market

ACT II — AI FOR NETWORKING

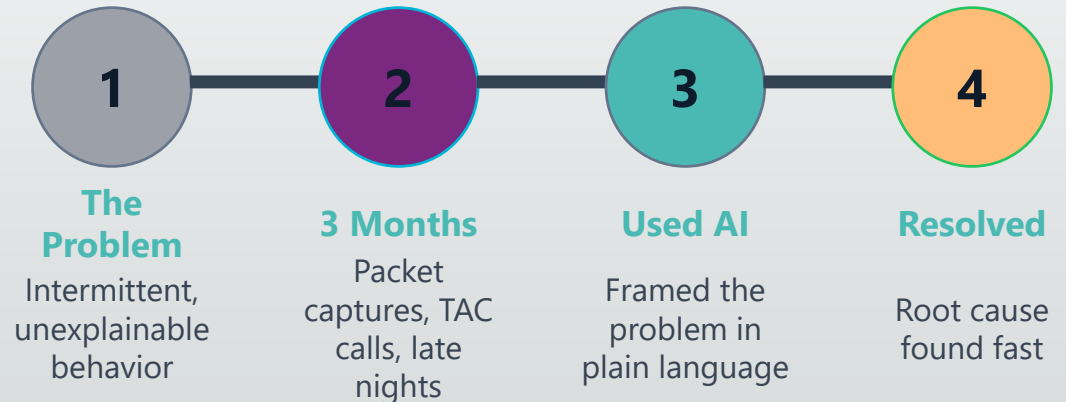
- LLMs + Agentic AI in Ops
- Cisco AI Canvas + Assistant
- Demo: AI Canvas
- Q&A

BEFORE WE DIVE IN...

A Story About a 3-Month Problem

Show of hands...

Who has worked a ticket for more than a **week?** **A month?** **Even longer?**



All of the Tools

None of the Answers

Application

App Dynamics

Redgate

Virtualization

Vcenter

Vrealize

Security

Arctic Wolf

Palo Alto

Storage

Pure Storage

Pure1 VM Analytics

Cloud

Azure Monitor

Log analytics

Networking

PRTG

Syslog



ChatGPT to the Rescue

One Poorly Written Prompt

ChatGPT ▾

I'm having issues with running my production SQL workloads in Vsphere 8. My vm's are not performing and I need some assistance.

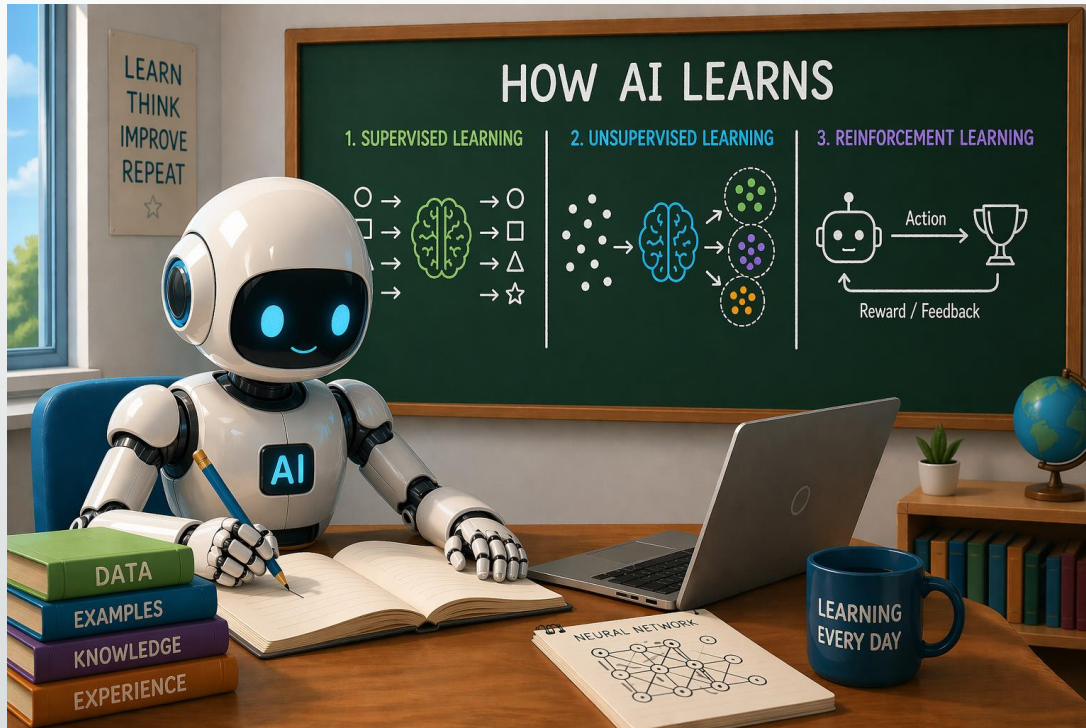
Running production SQL workloads on VMware vSphere 8 can involve a complex interplay of settings, configurations, and best practices to ensure optimal performance. Here are some common areas to investigate and potential solutions to improve VM performance for SQL workloads:

"That was the moment I stopped thinking of AI as a novelty and started treating it as a tool I can't afford to work without."



AI Fundamentals

How AI learns · Where it's used in networking



Supervised

Trained on labeled examples of right and wrong. Like a new engineer reviewing resolved tickets.

Anomaly detection, spam filtering

Unsupervised

No labels — AI finds its own patterns. Like handing someone thousands of logs saying "find what's weird."

Traffic clustering, behavioral analysis

Reinforcement

Learns by doing, reward/penalty style. Like a self-optimizing routing protocol rewarded for low latency.

WAN path optimization, QoS tuning



AI Fundamentals

How AI learns · Where it's used in networking

Machine Learning

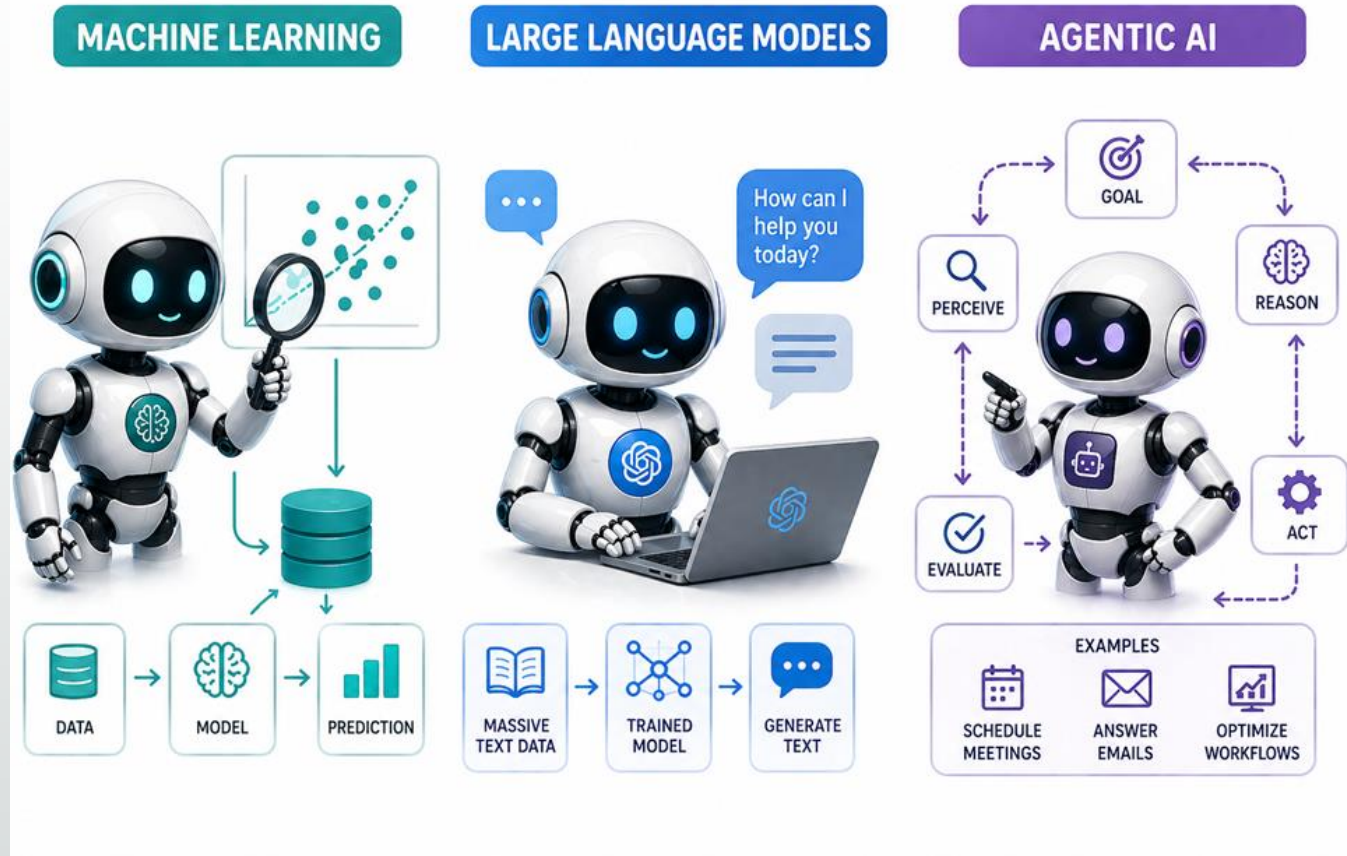
Pattern recognition at scale. Already inside your SIEM, SD-WAN controller, and firewall — whether you know it or not.

Large Language Models

The AI you talk to in plain English. Fine-tuned on vendor-specific data, it can reason about your exact network stack.

Agentic AI

Beyond answering questions — AI that acts. Correlates data, opens tickets, pushes config fixes, closes the loop.

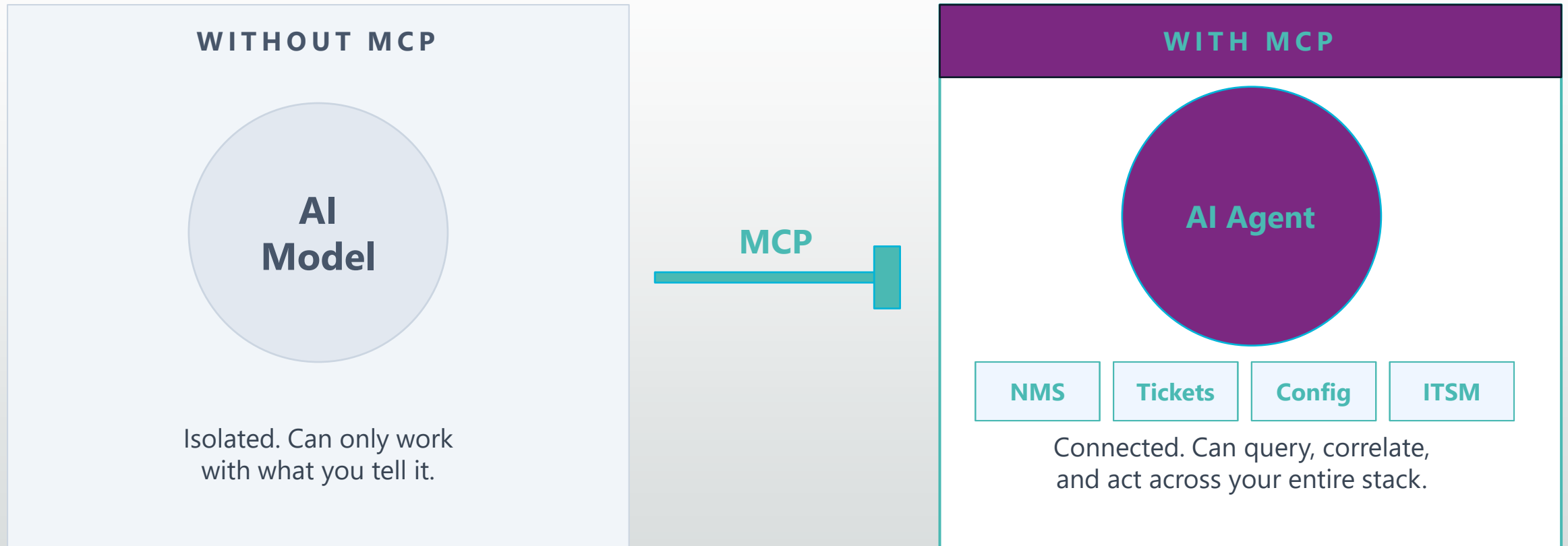


The evolution of AI: from pattern detection → language understanding → autonomous action



MCP & The Rise of Agentic AI

Model Context Protocol — the open standard that lets AI agents act, not just answer



*We're moving from AI that answers questions → AI that takes action.
That's not a small step — it's a fundamental shift in network operations.*

The Network is No Longer Just Plumbing

30–40%

GPU utilization lost
to network-induced stall

800G

Current AI fabric port
speed baseline

\$30K+

Per GPU — every idle
cycle is expensive

KEY PROTOCOLS & STANDARDS

RoCE v2

GPU-to-GPU memory reads over Ethernet without CPU involvement — dramatically lower latency for AI training

Ultra Ethernet Consortium

Industry initiative (Cisco, AMD, Intel, Meta, Microsoft) defining Ethernet standards purpose-built for AI

400G → 800G → 1.6T

The speed roadmap — 400G is already baseline for AI fabrics, 800G in production, 1.6T on the horizon

Cisco Networking for AI

End-to-end coverage: from GPU spine to campus edge — built for the AI era

ASIC PLATFORM

Silicon One

- 102.4 Tbps switching capacity
- Unified chip: hyperscale AI fabrics to edge routing
- G200/G202 variants in production today

AI DATA CENTER

Nexus 9000 Series

- Native RoCEv2 support
- 400G and 800G available
- Purpose-fit for GPU cluster interconnects
- Cloud-scale ASIC architecture with Silicon One

SMART CAMPUS

Catalyst 9350

- Embedded telemetry streaming built in
- ThousandEyes integration native
- UPoE+ and MgiG to support WiFi 7
- 1.6Tbps of stacked Bandwidth
- 4X100G Uplinks



Palo Alto Networks for AI

AI-native security for the AI era — from new attack surfaces to post-quantum threats

THREAT LANDSCAPE

New AI Attack Surface

Prompt injection, model poisoning, shadow AI usage, and data exfiltration through LLM APIs are threats that traditional perimeter security was never designed to catch.

NGFW PLATFORM

App-ID + Precision AI

Classify and control AI application traffic (ChatGPT, Copilot, custom LLM endpoints) with granularity no other vendor matches today. Inline ML — no cloud roundtrip required.

SASE ENFORCEMENT

Prisma AI Access (SASE)

Enforce AI governance policies at the edge. Control who can use which AI tools, with what data, from anywhere. The policy plane for enterprise AI adoption.

FORWARD-LOOKING

Post-Quantum Cryptography

"Harvest now, decrypt later" attacks are already happening — nation-states collecting encrypted traffic today to decrypt when quantum compute matures. Palo is integrating NIST-approved PQC algorithms now.

The Broader Market

Every major networking vendor is making AI a centerpiece of their roadmap right now

INTENT-BASED AI

Juniper

- Apstra platform: declare desired state, AI verifies and auto-corrects in a closed loop
- PTX series routers with Express Silicon targeting AI backbone interconnects
- HPE acquisition creates the largest networking entity outside of Cisco

AI-POWERED ACCESS

Aruba

- Aruba Central AIOps: anomaly detection, root cause analysis, and client experience scoring
- One of the most mature AI-for-networking platforms currently on the market

AI SECURITY AT SCALE

Fortinet

- FortiASIC custom processors with on-device ML inference — no cloud dependency
- Purpose-fit for OT, branch, and distributed environments needing local AI security
- Strong story where you cannot or will not send security telemetry offsite

The question isn't whether to engage with AI in networking — it's which platform to bet on and when.

Supercharging Your Network Operations Team

Reactive

Alert fires.
Team investigates.
Hours or days to resolve.

Where most teams are today

Proactive

AI detects patterns
before failures occur.
Issues surfaced proactively.

What AI gets you now

Autonomous

AI correlates, plans,
and executes fixes.
Human approves actions.

Where this is heading fast

AGENTIC AI USE CASES

- Incident correlation in under 60 sec — AI pulls logs, checks BGP states, reviews change history, and ranks probable causes
- Auto-remediation with approval gates — detects the issue, plans the fix, waits for engineer approval, then executes
- Capacity planning — 90 days of utilization data, saturation risk identification, procurement recommendations generated automatically

Deep Network Model

The most advanced networking LLM

Purpose-Built for Networking

Up to 20% more precise reasoning for troubleshooting, configuration, and automation.

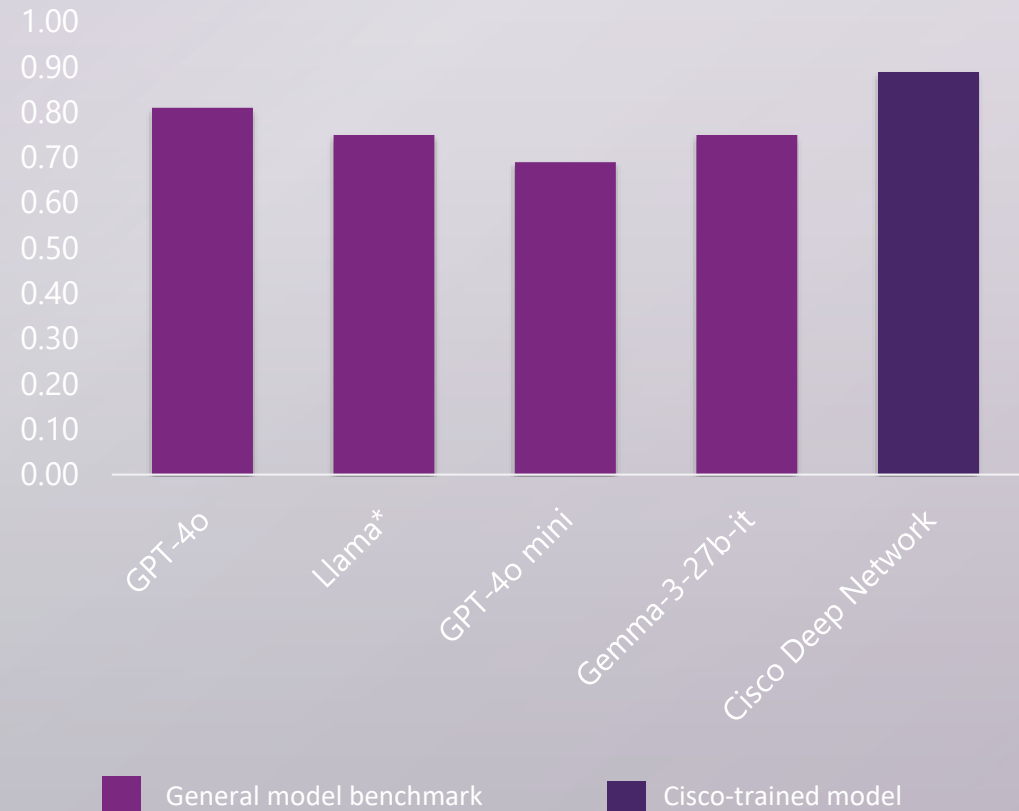
Trusted Training

Fine-tuned on 40+ years of Cisco expertise and expert-vetted for accuracy.

Continuous Learning

Evolves with live telemetry and real-world Cisco TAC and CX insights.

Outperforms general models by ~20%



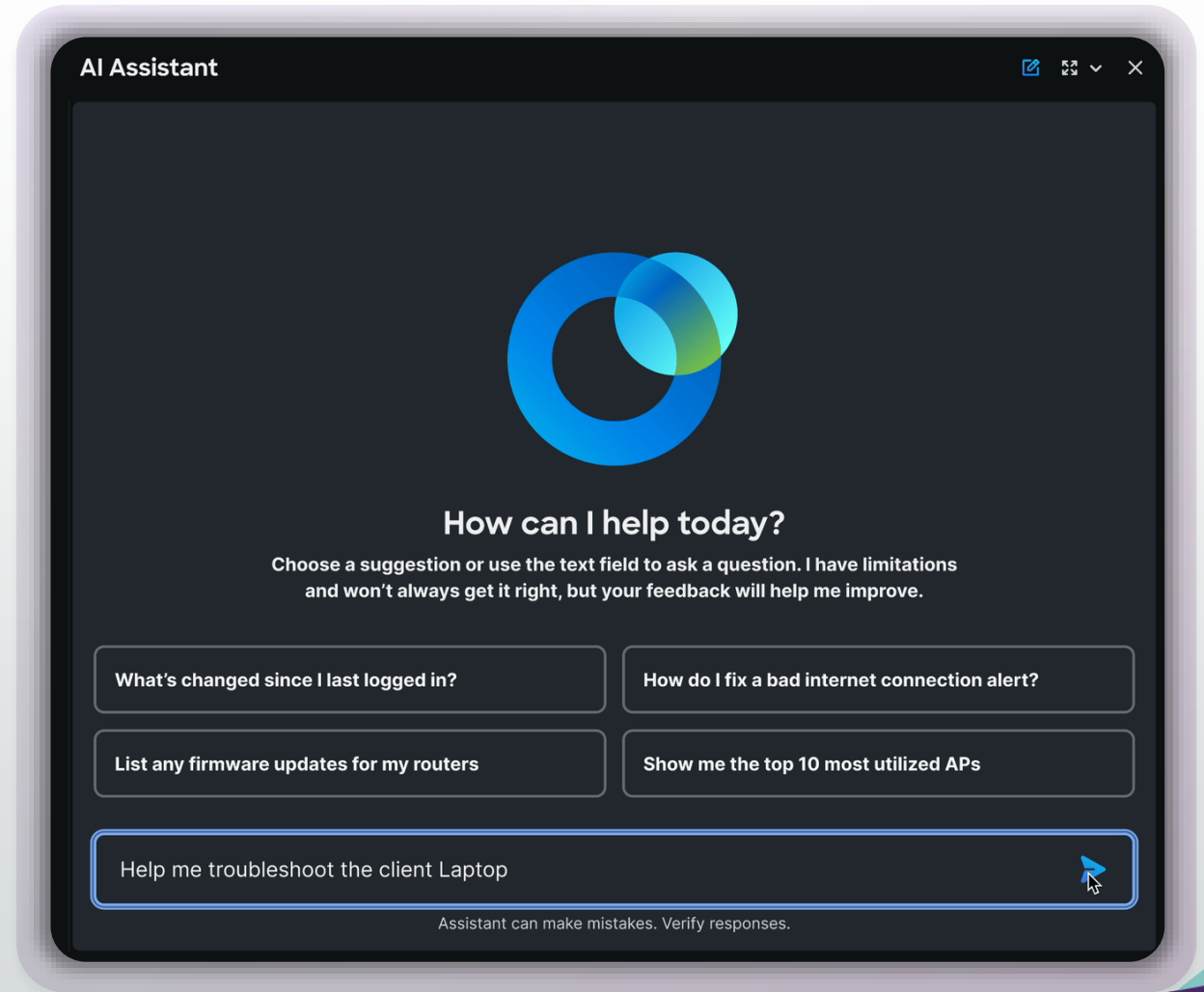
Accuracy on CCIE-style MCQs (590-question benchmark, May 2025)
*3.2-90B-Vision-instruct

AI Assistant Rapid troubleshooting and RCA

Finds root cause—not just the alert

Provides full stack visibility

Pinpoints the issue and guides the fix

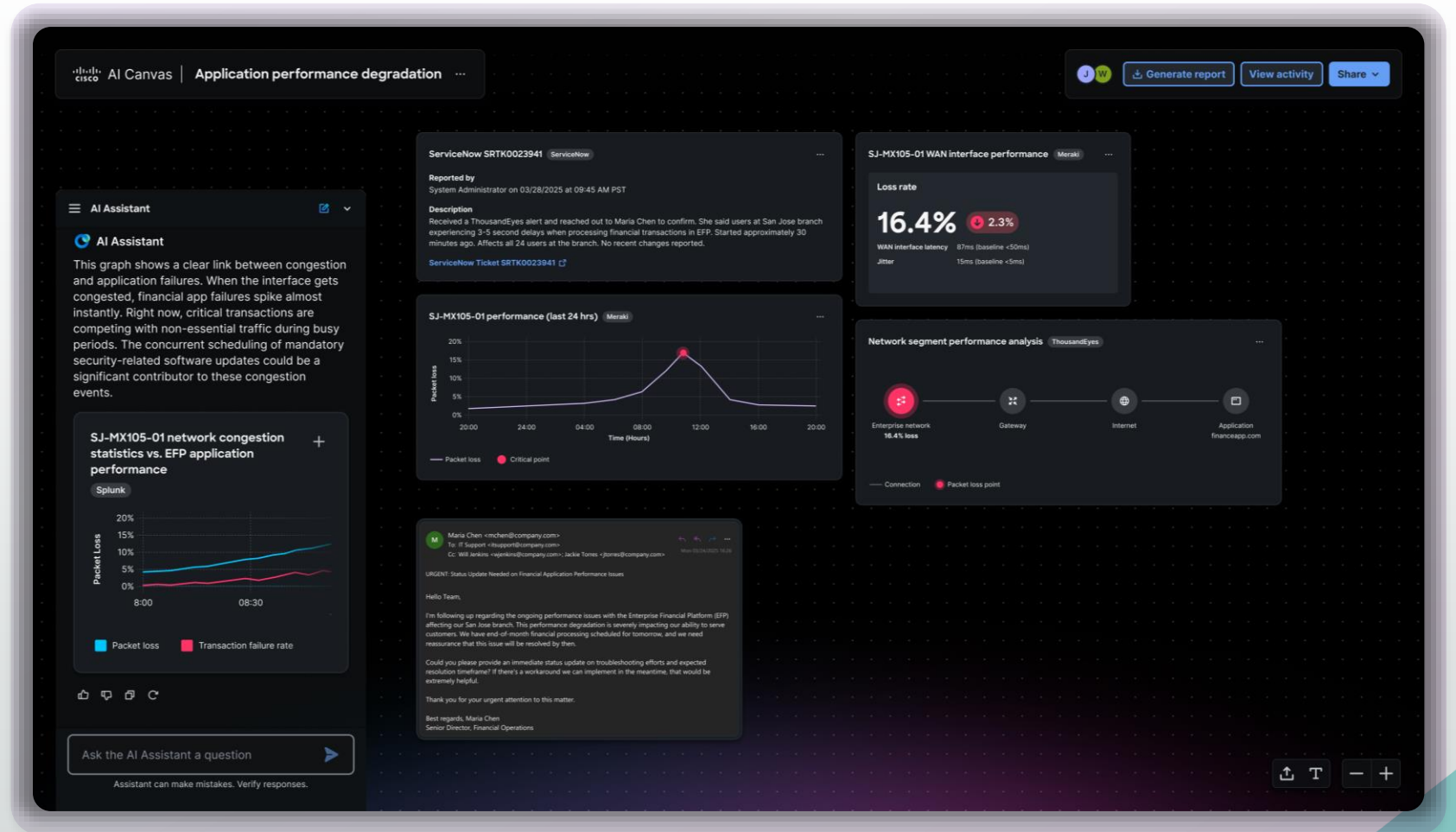


AI Canvas

Troubleshooting and execution across multiple domains

Collaboration across multiple users (NetOps, SecOps and execs)

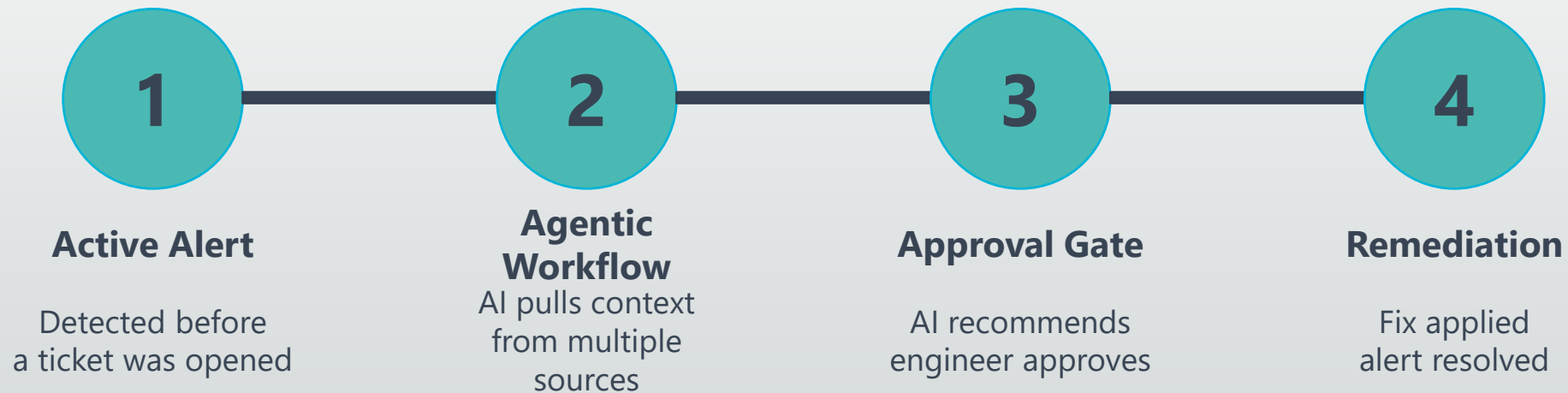
Built on the foundation of the Deep Network Model



D E M O

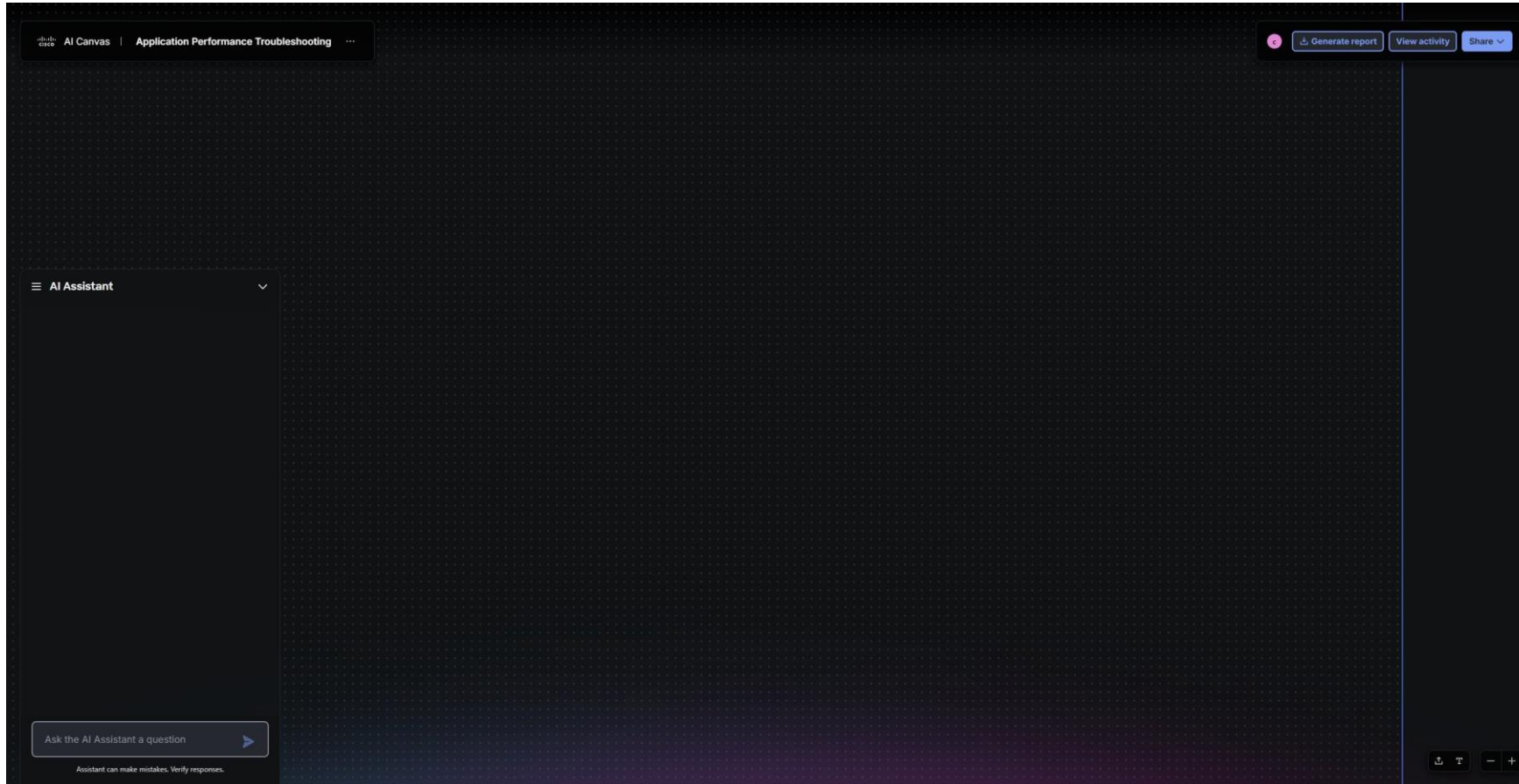
Cisco AI Canvas

Agentic AI in action — correlating data, finding issues, resolving them in one platform



Cisco AI Canvas — The Workspace

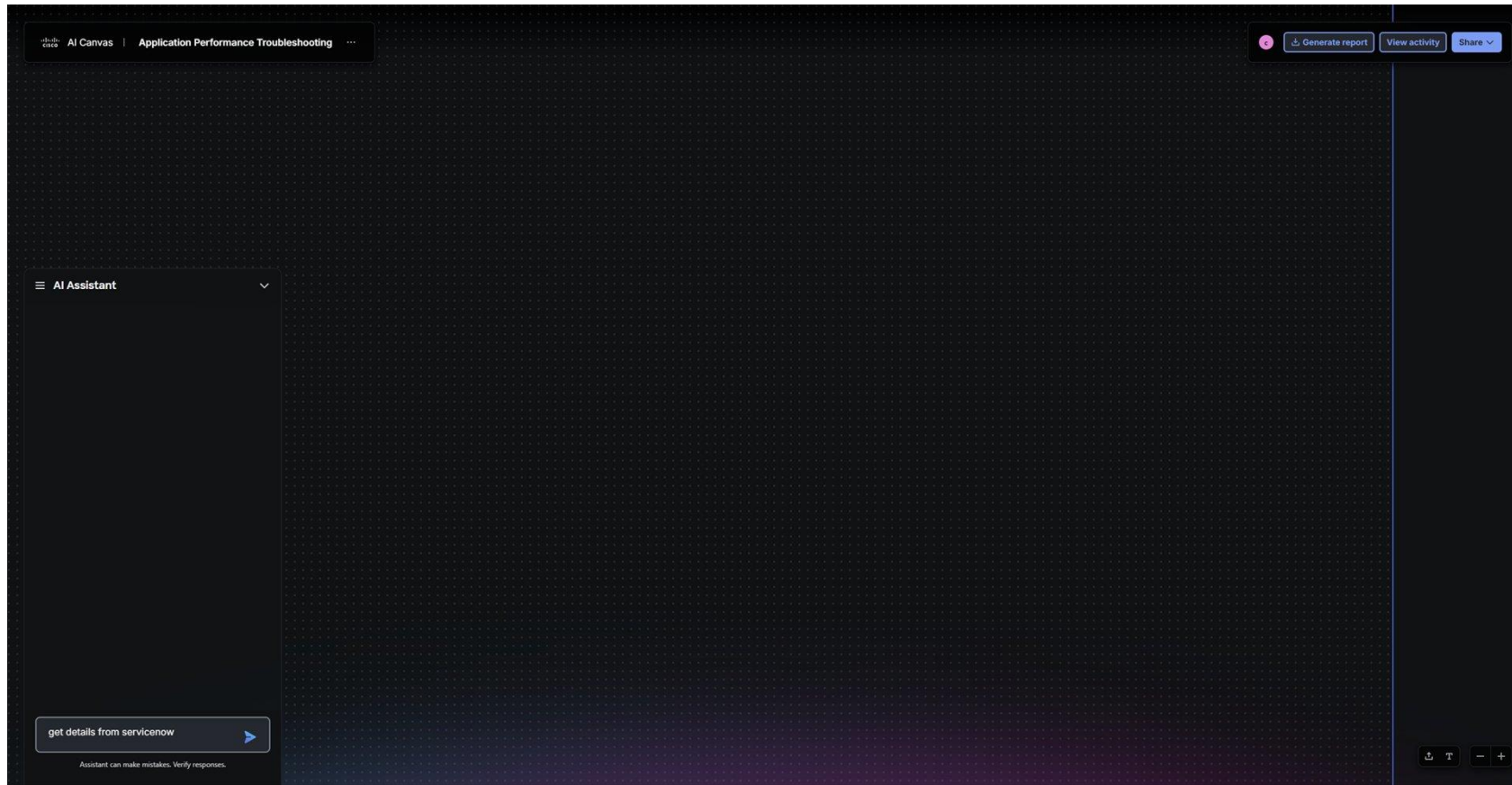
A blank canvas. The AI Assistant is embedded in the corner — ready to be your investigation partner.



AI Assistant docked bottom-left — accepts natural language prompts

Pull the Ticket — In Plain English

No portal switching, no API calls. Just type the ask and let the agent reach into ServiceNow.

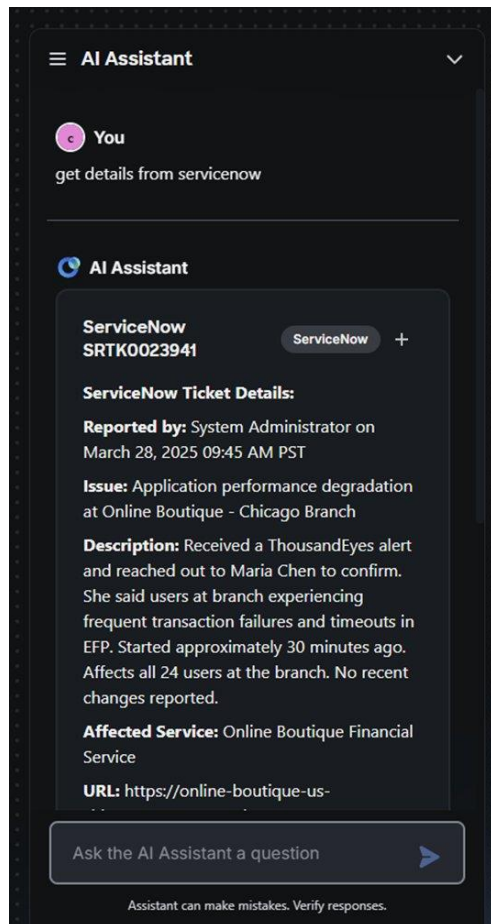


NATURAL LANGUAGE PROMPT

"get details from servicenow" — the agent recognizes intent, finds the connector, fetches the ticket

Ticket Pulled — Full Context

AI Assistant returns the complete ticket — reporter, issue, automated diagnostics, and Meraki/ThousandEyes IDs.

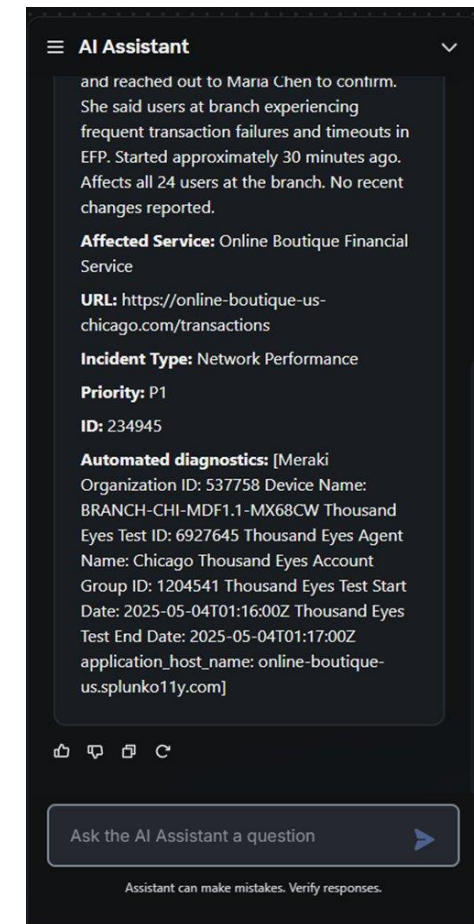


Top of ticket

WHY THIS MATTERS

No tab-switching to ServiceNow.

Full ticket context arrives directly in the canvas — ready for the next step.



Diagnostics + device IDs

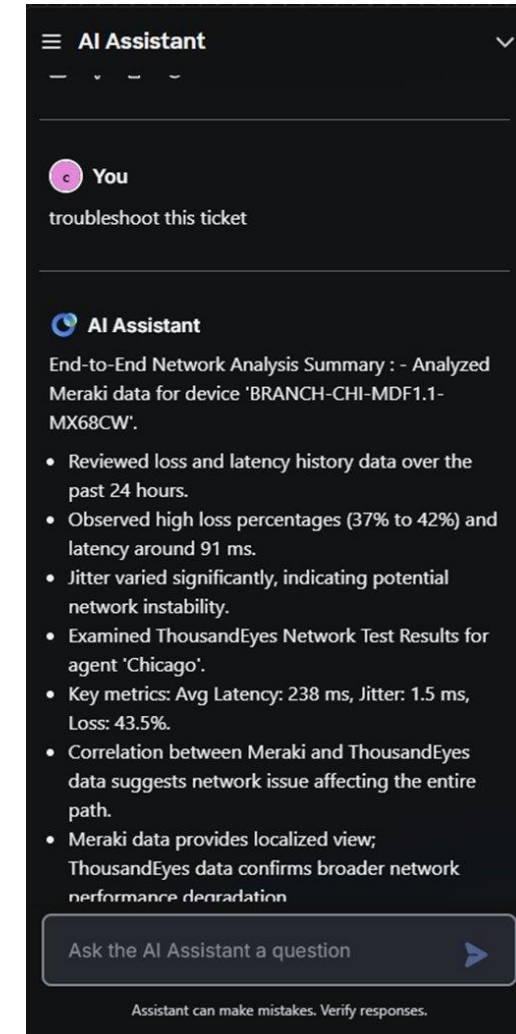
Troubleshoot This Ticket

Two-word prompt. AI Assistant kicks off an end-to-end network analysis pulling Meraki + ThousandEyes data automatically.

"troubleshoot this ticket"

What the AI does behind the scenes

- Pulls 24 hours of Meraki loss & latency data for the named device
- Examines ThousandEyes test results from the Chicago agent
- Calculates key metrics: 238 ms latency, 1.5 ms jitter, 43.5% loss
- Correlates Meraki (local view) with ThousandEyes (path view)
- Identifies the issue is internal — not an ISP problem



Remediation Suggestions

AI Assistant moves from diagnosis to prescription — with three concrete fixes the engineer can act on.

Three actionable fixes

1

Configure QoS on MX device

Prioritize application traffic at the branch gateway

2

Limit non-essential traffic

Reduce bandwidth contention during business hours

3

Add DSCP tags

Apply network-wide priority marking for app traffic

AI Assistant

- Jitter varied significantly, indicating potential network instability.
- Examined ThousandEyes Network Test Results for agent 'Chicago'.
- Key metrics: Avg Latency: 238 ms, Jitter: 1.5 ms, Loss: 43.5%.
- Correlation between Meraki and ThousandEyes data suggests network issue affecting the entire path.
- Meraki data provides localized view; ThousandEyes data confirms broader network performance degradation.

Remediation Suggestions : The path view confirms our internal network issue. This is completely within our control to fix. Here are some solutions you can implement immediately:

- Configure QoS on the MX device to prioritize the application traffic.
- Limit non-essential traffic during business hours.
- Add DSCP tags to application traffic for network-wide priority.

The end-to-end network analysis is complete. Act on the provided remediation suggestions. Please escalate to an expert if the problem still persists.

👍 🗨️ 📄 🔄

Ask the AI Assistant a question

Assistant can make mistakes. Verify responses.

KEY POINT

AI moves from "what is wrong" to "what to do about it" — with confidence in each step.

The Full Canvas Comes Alive

Loss rate, WAN performance, network path analysis, AI summary — all assembled in one view.

AI Assistant

- Jitter varied significantly, indicating potential network instability.
- Examined ThousandEyes Network Test Results for agent 'Chicago'.
- Key metrics: Avg Latency: 238 ms, Jitter: 1.5 ms, Loss: 43.5%.
- Correlation between Meraki and ThousandEyes data suggests network issue affecting the entire path.
- Meraki data provides localized view; ThousandEyes data confirms broader network performance degradation.

Remediation Suggestions : The path view confirms our internal network issue. This is completely within our control to fix. Here are some solutions you can implement immediately:

- Configure QoS on the MX device to prioritize the application traffic.
- Limit non-essential traffic during business hours.
- Add DSCP tags to application traffic for network-wide priority.

The end-to-end network analysis is complete. Act on the provided remediation suggestions. Please escalate to an expert if the problem still persists.

Ask the AI Assistant a question

Assistant can make mistakes. Verify responses.

Incident Type: Network Performance
 Priority: P1
 ID: 234945
 Automated diagnostics: [Meraki Organization ID: 537758 Device Name: BRANCH-CHI-MDF1.1-MX68CW Thousand Eyes Test ID: 6927645 Thousand Eyes Agent Name: Chicago Thousand Eyes Account Group ID: 1204541 Thousand Eyes Test Start Date: 2025-05-04T01:16:00Z Thousand Eyes Test End Date: 2025-05-04T01:17:00Z application_host_name: online-boutique-us.splunko11y.com]

BRANCH-CHI-MDF1.1-MX68CW - WAN Interface Performance

Loss Rate
 42.78% 5.04%

WAN interface latency 90.91ms (baseline < 50ms)
 Jitter 1.1ms (baseline < 5ms)

BRANCH-CHI-MDF1.1-MX68CW - WAN interface performance (Last 24 hrs)

Loss Percent

Network path performance analysis

172.21.10.207 ...38.0ms latency) — Connection

172.21.10.1 — Connection

BRANCH-CHI-MDF1.1-MX68CW (Packet loss point)

198.27.154.1 ...Latency: 86 ms) — Connection

198.27.135.249 ...Latency: 91 ms) — Connection

Multi-hop — Connection

Application - on... ...tency: 237 ms) — Connection

Invite collaborators

Everyone you invite will have edit access.

djsam+meraki@cisco.com

+ DJ Sampath

Who has access
 Only you

Regenerate AI summary

The issue at hand involves network performance degradation affecting the Online Boutique's Chicago branch. A ServiceNow ticket was raised due to application performance issues, with users experiencing frequent transaction failures and timeouts. The problem was identified through a ThousandEyes alert, indicating a network performance issue with a high priority (P1).

The analysis focused on the network path performance, particularly the device 'BRANCH-CHI-MDF1.1-MX68CW'. Data from Meraki and ThousandEyes showed significant packet loss (37% to 43.5%) and high latency (up to 238 ms), suggesting network instability. The analysis confirmed that the issue is within the internal network and can be addressed by configuring Quality of Service (QoS) on the MX device, limiting non-essential traffic, and prioritizing application traffic with DSCP tags.

The most recent action taken was a comprehensive end-to-end network analysis, which provided remediation suggestions to resolve the network issue. The user is advised to implement these solutions and escalate to an expert if the problem persists.

Cancel **Invite**

LOSS RATE: 42.78%

Confirmed packet loss spike

NETWORK PATH

Red node = packet loss point

AI SUMMARY

Plain-English RCA — sharable

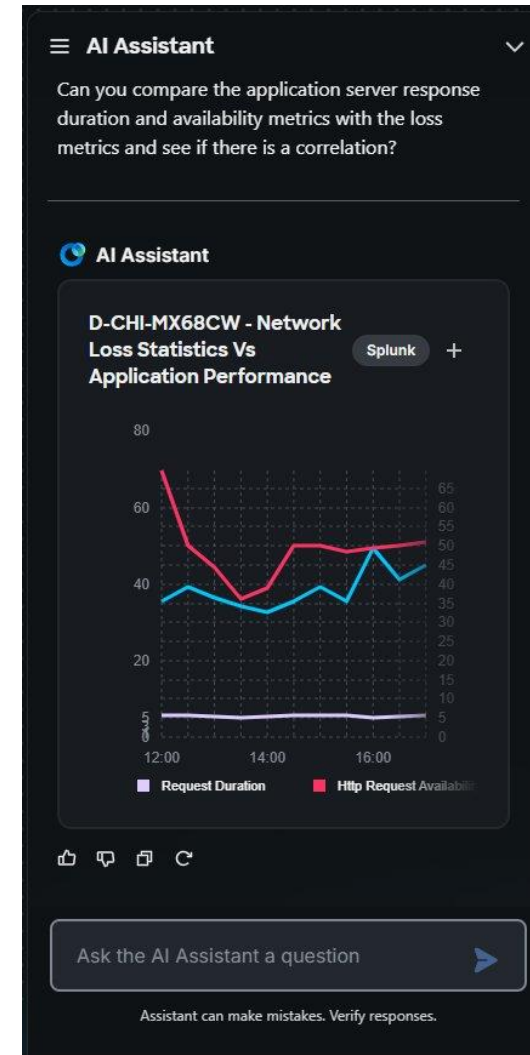
Cross-Domain Correlation

AI Canvas pulls data from Splunk too — correlating app server response time with network loss in one view.

"Can you compare the application server response duration and availability metrics with the loss metrics and see if there is a correlation?"

What you're looking at

- HTTP Request Availability dips when network loss climbs
- Request Duration rises in lockstep with degradation
- Single chart proves what previously took multiple tools
- Splunk data pulled inline — no separate dashboard



Apply the Fix — From the Canvas

AI Canvas doesn't just recommend — with approval, it executes. QoS policy pushed to Meraki MX in one step.

"Can you apply the QoS policy update on the Meraki MX device with name MX68CW?"

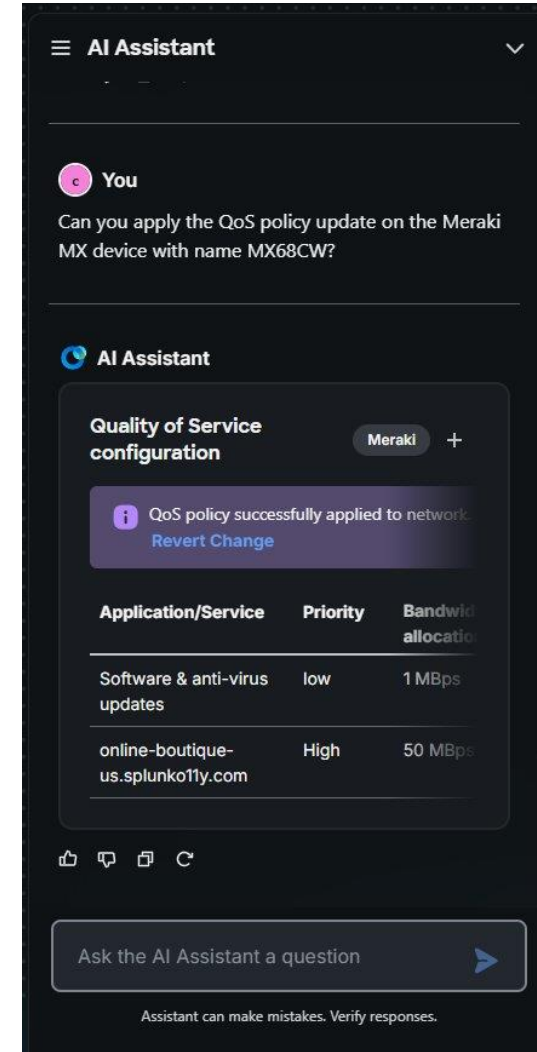
 QoS policy successfully applied to network

What got pushed

APPLICATION	PRIORITY	BANDWIDTH
online-boutique	High	50 Mbps
anti-virus updates	Low	1 Mbps

APPROVAL GATE

AI proposed → engineer confirmed → AI executed. The "Revert Change" button stays available for instant rollback.



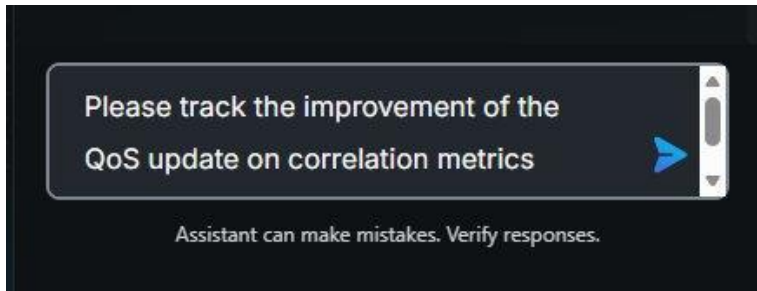
The screenshot shows the AI Assistant interface with a dark theme. At the top, it says "AI Assistant" with a dropdown arrow. Below that, a message from "You" asks: "Can you apply the QoS policy update on the Meraki MX device with name MX68CW?". The AI Assistant responds with a confirmation: "Quality of Service configuration" for the "Meraki" device, followed by an information box stating "QoS policy successfully applied to network" with a "Revert Change" button. Below this is a table of the applied configuration:

Application/Service	Priority	Bandwidth allocation
Software & anti-virus updates	low	1 Mbps
online-boutique-us.splunko11y.com	High	50 MBps

At the bottom, there are social sharing icons and a button that says "Ask the AI Assistant a question" with a right-pointing arrow. A small disclaimer at the very bottom reads: "Assistant can make mistakes. Verify responses."

Did It Actually Work?

AI Canvas tracks the remediation in real time — and shows you the moment things change.



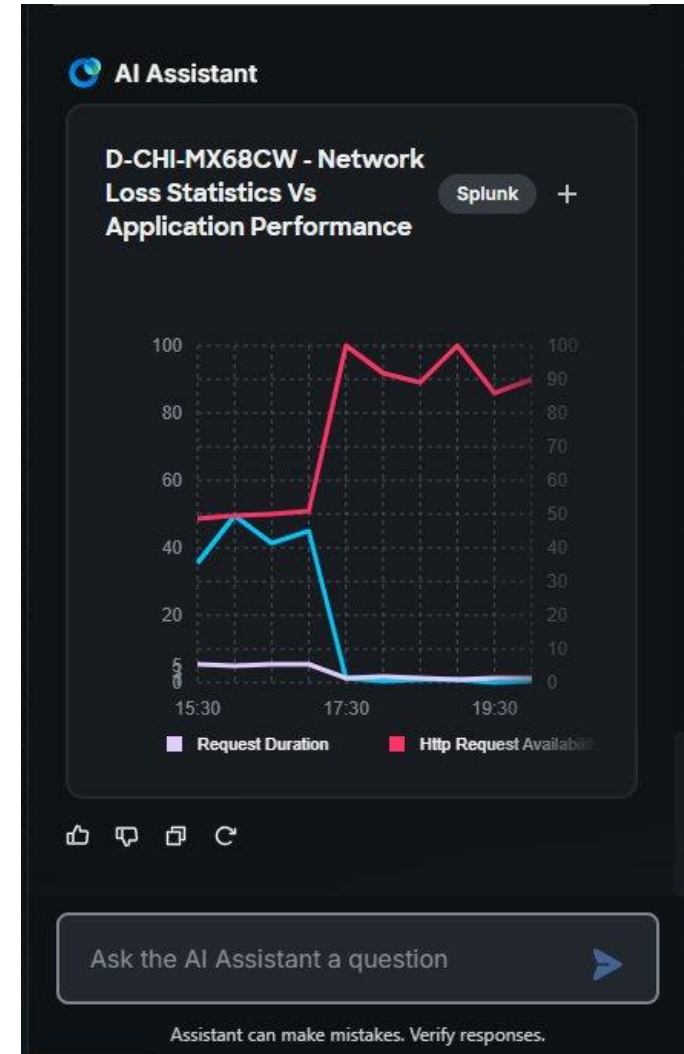
What you're seeing in the graph

BEFORE HTTP availability fluctuating around 36–50%

AFTER Availability climbs to ~95% — loss drops to near zero

CLOSED-LOOP REMEDIATION

Detect → diagnose → recommend → apply → verify. AI Canvas closes the loop the engineer used to walk manually.



Generate the Report

Click one button. AI Canvas writes the full root cause analysis with the fix and validation — ready to share.

What's in the report

Application Performance Metrics

HTTP availability + request duration during incident

QoS Configuration Applied

Bandwidth allocations and priorities pushed to MX

Root Cause Analysis

High loss + latency, internal network path

Recommendations

Configure QoS · limit non-essential · DSCP tags

Conclusion

Validates the fix and outlines escalation criteria

Application Performance Metrics

The application performance was evaluated using metrics such as HTTP request availability and request duration:

- **HTTP Request Availability:** Fluctuated between 36.11% and 100%, with notable dips during the incident period.
- **Request Duration:** Varied slightly, with a general trend of increased duration correlating with higher loss percentages.

Quality of Service (QoS) Configuration

A QoS policy was successfully applied to the network to prioritize critical application traffic. The configuration details include:

- **High Priority:** Allocated 50 MBps bandwidth to the Online Boutique service.
- **Low Priority:** Allocated 1 MBps bandwidth to software and anti-virus updates.

Root Cause Analysis and Recommendations

The analysis suggests that the network performance degradation is primarily due to high loss rates and latency, affecting the entire network path. Recommendations for remediation include:

- **Configure QoS:** Prioritize application traffic on the MX device.
- **Limit Non-Essential Traffic:** Reduce bandwidth usage during peak business hours.
- **Add DSCP Tags:** Implement network-wide priority for critical application traffic.

Conclusion

The end-to-end network analysis indicates that the performance issues are within the internal network's control to resolve. Implementing the recommended QoS configurations and traffic management strategies should mitigate the observed performance degradation. Further escalation may be necessary if the problem persists after these measures are applied.

Close Download as PDF

"This is exactly the kind of correlation that would have solved my 3-month problem in about 3 minutes."



ANM TECH DAY

Questions?

Connect With Me!

Email: Anthony.Budetti@anm.com

LinkedIn:

