

Gain Control of AI's Impact on Network Traffic and User Experience

cisco Live !

Leveraging Provider Connectivity Assurance (formerly Accedian Skylight)

Paul Thompson
Technical Leader – Technical Marketing Engineering

Cisco Webex App

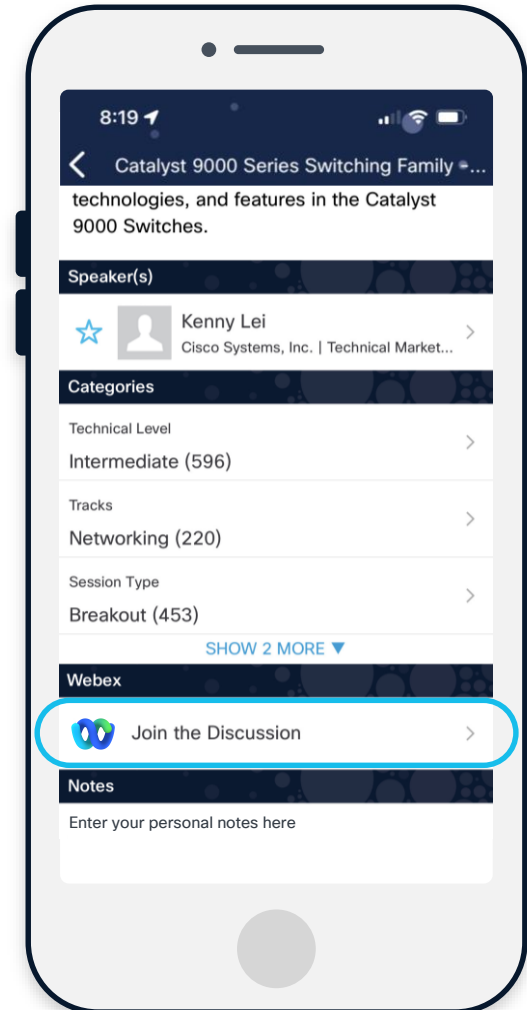
Questions?

Use Cisco Webex App to chat with the speaker after the session

How

- 1 Find this session in the Cisco Live Mobile App
- 2 Click “Join the Discussion”
- 3 Install the Webex App or go directly to the Webex space
- 4 Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 13, 2025.



<https://ciscolive.ciscoevents.com/ciscolivebot/#AIHUB-2013>

AI-generated Traffic Presents Itself To The Network In Different Ways When Compared To Other Traffic Types

Assure and Optimize AI-experiences With
Provider Connectivity Assurance (PCA)

Agenda

- 01 Are networks ready for AI traffic?
- 02 What's different about AI - generated traffic?
- 03 How do you assess a network for AI - readiness?
- 04 Future Plans
- 05 Would you like a demo of our LLM assurance agent?

Thank you to Virginia Teixeira for some of the content in this session.

Impact of AI Traffic in Transport Networks - BRKSPG-1180
(June 9th, 11am PDT)

01: Are Networks Ready For AI Traffic?

66%

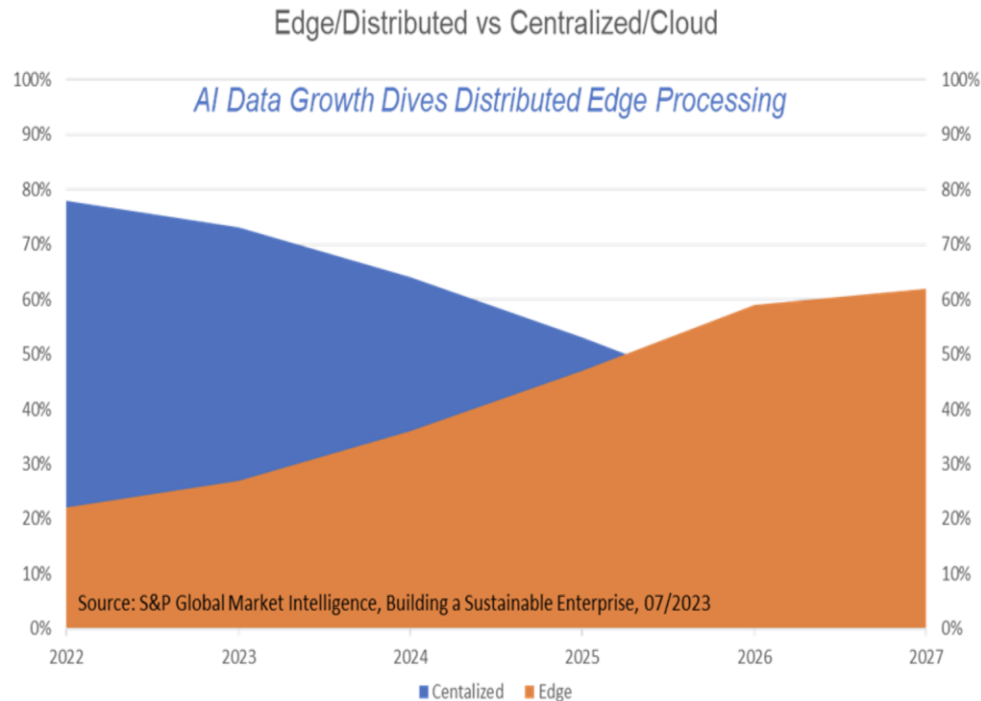
of enterprises list GenAI workloads as one of their top use cases for using multi cloud networking

IDC report

New AI-Assistants

will drive an increase in uplink traffic that is unprecedented, beyond the capacity of current 5G networks as soon as 2028

Mobile Experts



By 2027, 62% of total data will be computed at edge

S&P global market intelligence

01: Are Networks Ready For AI Traffic?

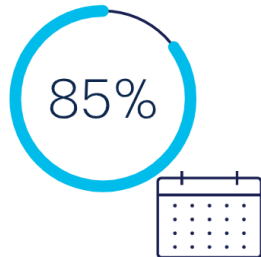
Worldwide, companies know the AI deployment clock is ticking. And the pressure is relentless.



98% feel increased urgency over the past year.



For 50% of organizations, CEOs and leadership – more so than any other factor – are driving urgency for AI.

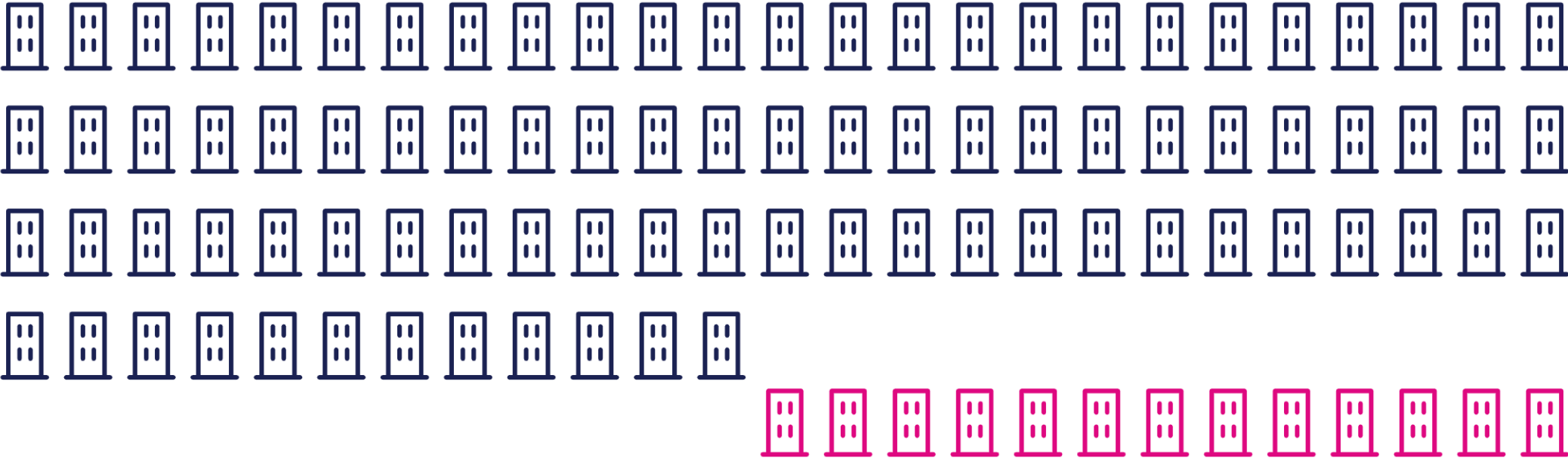


85% say they have less than 18 months to deploy an AI strategy, or they will see negative business effects.



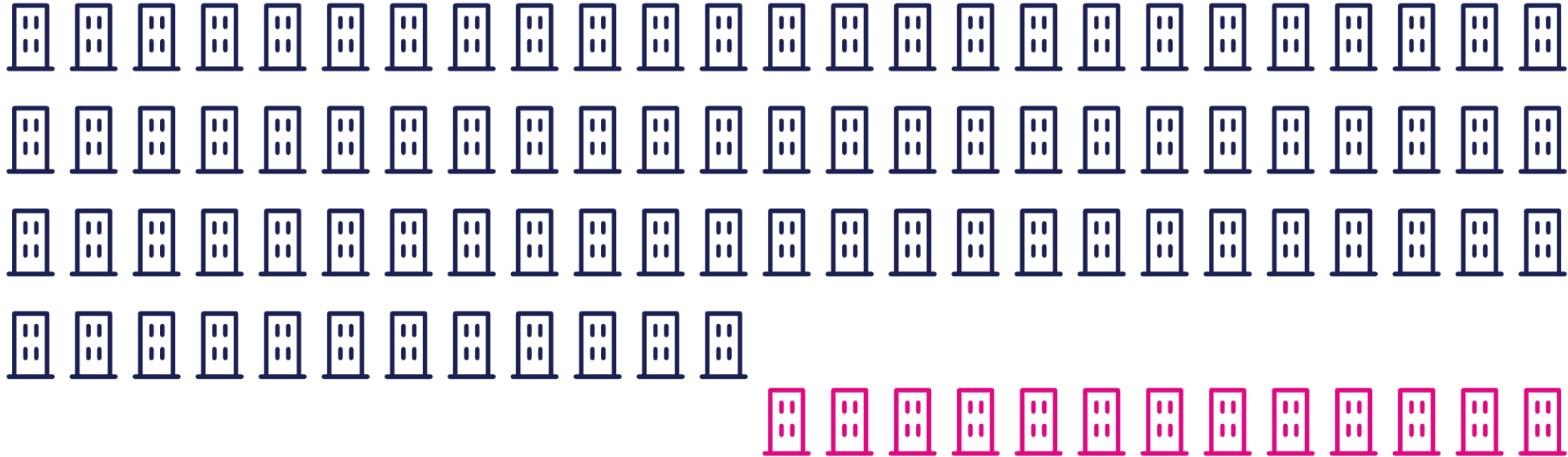
AI is a priority spend for IT budgets: 50% of companies say they've already dedicated 10-30% of their budget to AI.

01: Are Networks Ready For AI Traffic?



13% of companies globally are ready to leverage AI and AI-powered technologies to their full potential, a 1-point decline compared to last year.

01: Are Networks Ready For AI Traffic?



13% of companies globally are ready to leverage AI and AI-powered technologies to their full potential, a 1-point decline compared to last year.

Global readiness



Cisco, AI readiness, 2024

01: Cisco Portfolio Innovations To Power AI Connectivity



Routing infrastructure

1

Cisco 8000 Cisco Silicon One

Unparalleled performance, TCO and operational simplicity

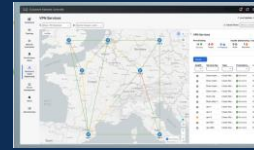


IP / optical convergence

2

Cisco Routed Optical Networking

Dramatic cost and energy savings with coherent pluggable optics



Network programmability

3

Cisco Crosswork Network Automation

SLA differentiation and policy control across DC, access, metro, backbone and cloud



Observability and insight

4

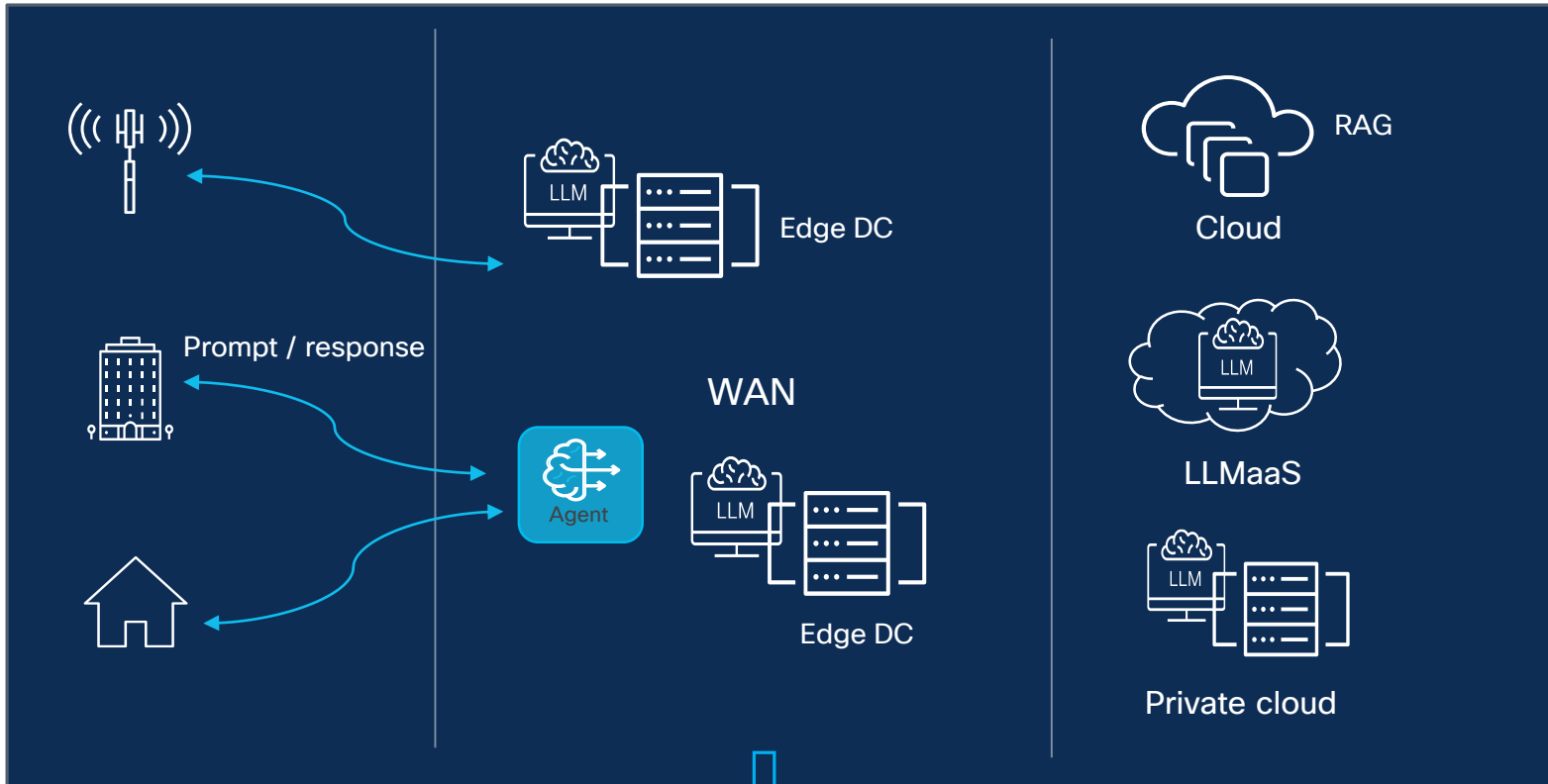
Cisco Provider Connectivity Assurance

Real time visibility and insights for proactive and predictive experience assurance

02: What's Different?

AI Inference

Edge, RAG, AI agents, AI assistants



North-South transactions from client to LLM generate significant East-West activity between DCs across the WAN

Bandwidth Demand

- AI traffic is unique, dynamic, non-cacheable
- Upstream = 10x downstream
- AI traffic is growing fast
- Peering links can be expensive and hard to scale

Latency

- AI assistants are chatty, interactive – every 300msec increase leads to 30% drop engagements
- AI agents require multiple processing steps before reply– latency x10s or higher with multi-agents
- Less predictable latency due to multi-modal varying request/response sizes
- Meta has the goal to have sustained <30msec rtt for real-time AI/immersive experiences

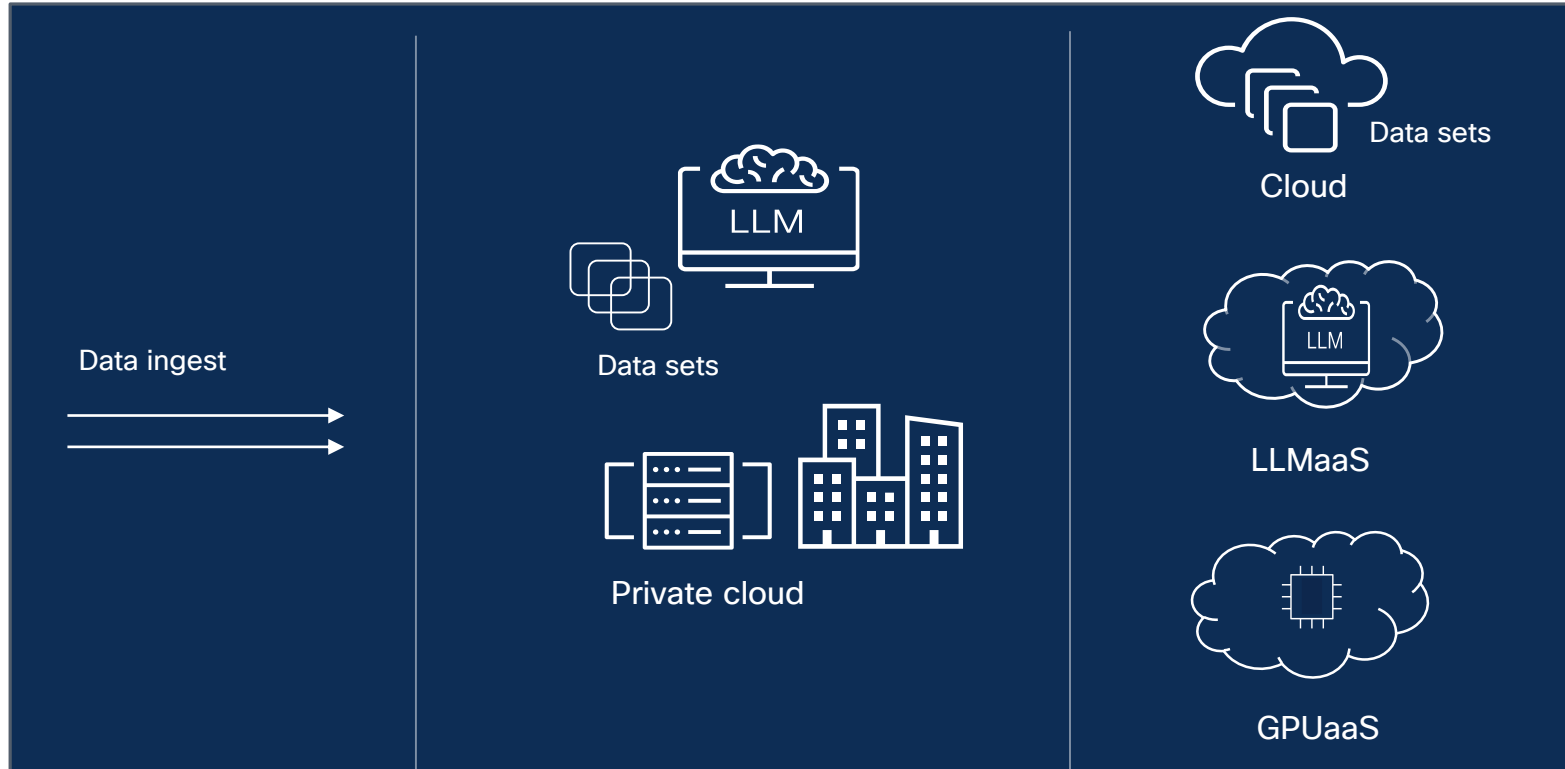
Resiliency & security

- Reduce blast radius – service continuity
- Process data locally, regulatory compliance, protect IP
- Failures/error rate can significantly impact real-time, interactive experiences

02: What's Different?

AI Training

Training, Fine-tuning, Federation, Grounding



Dataset movement & replication

- Very High bandwidth needs
- AI Federation require iterations of training with transfer of models (100s of GB or TB at a time)
- Fresh data ingest – collect from where it's generated to where it will be used

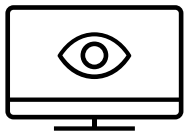
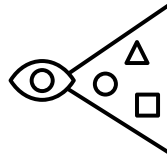
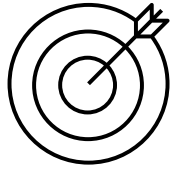
Secure high throughput transfers

Reliable

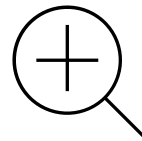
- Good quality data and network resiliency is critical for training

03: Visibility Is Key – Detect and Identify

Accuracy and Granularity



2 x one-way




Focus on KPIs that are critical for AI-traffic on the network

The performance of the network has a major influence on user Quality of Experience (QoE)


03: Visibility Is Key – Detect and Identify

Three Pillars Of Assurance for AI-generated traffic

Synthetic Network/Service Probing


 **Network performance**

- Synthetic probing
- Standards-based
- 24/7 continuous
- 50+ KPIs per session
- 2 x one-way
- End-to-End or
- Segment-by-Segment


 **Network Testing**

- On-demand or
- Scheduled
- L2, L3 or L4
- Used at turn-up or for troubleshooting

LLM Assurance Agent (+other agent types)


 **Throughput**

- Requests/sec
- Input tokens/sec
- Output tokens/sec

 **Token Handling**

- Time to first token (TTFT)
- Time per output token (TPOT)


Passive Per-Packet Intelligence

 **Network performance**

- How is network used?
- How is it performing?
- Identify and isolate network-level errors

Application delivery (TCP performance)

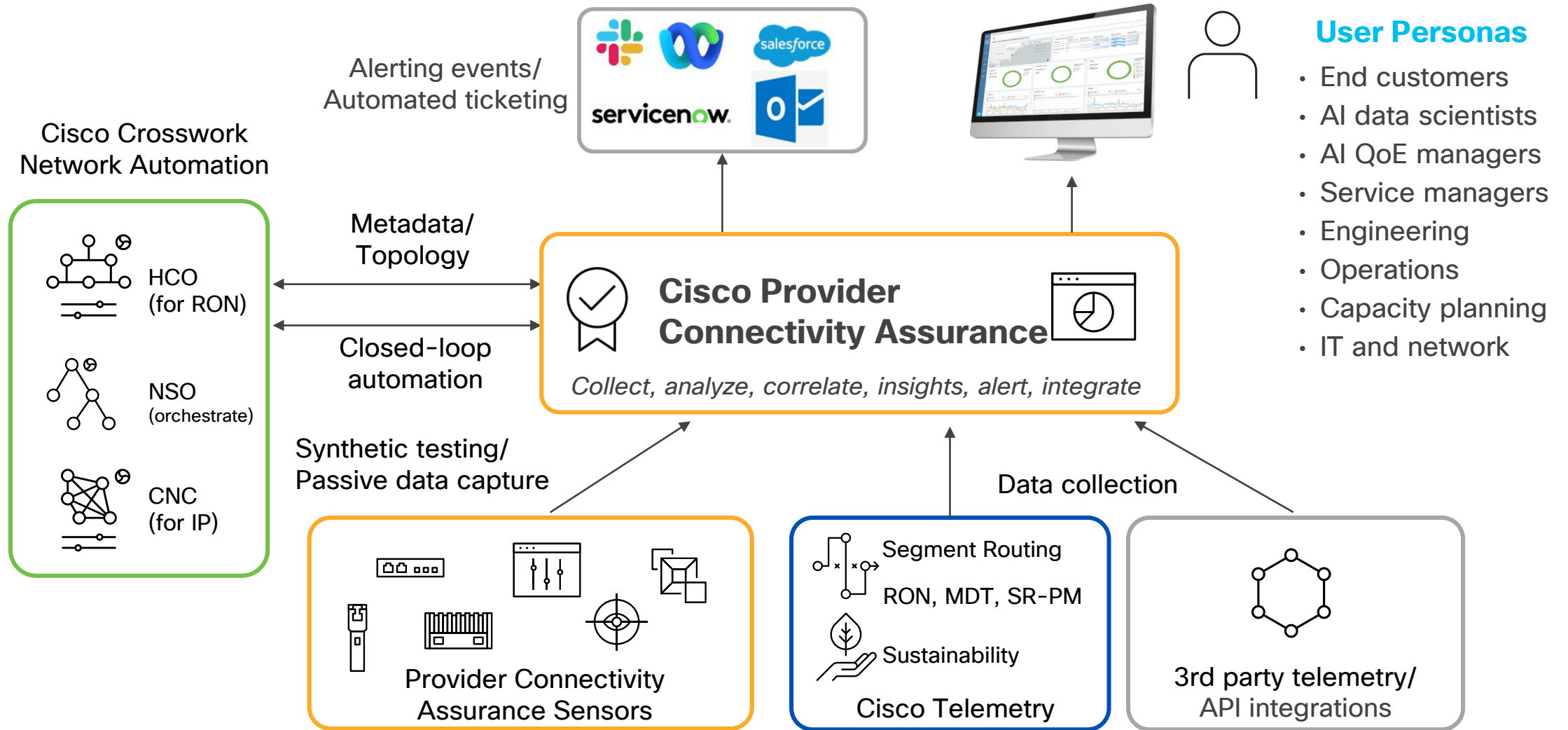
- End-user experience
- Application delivery over end-to-end infrastructure
- Degradation scope and fault domain origin

 **Application performance**

- Transaction-level response times
- Application errors
- Back-end performance

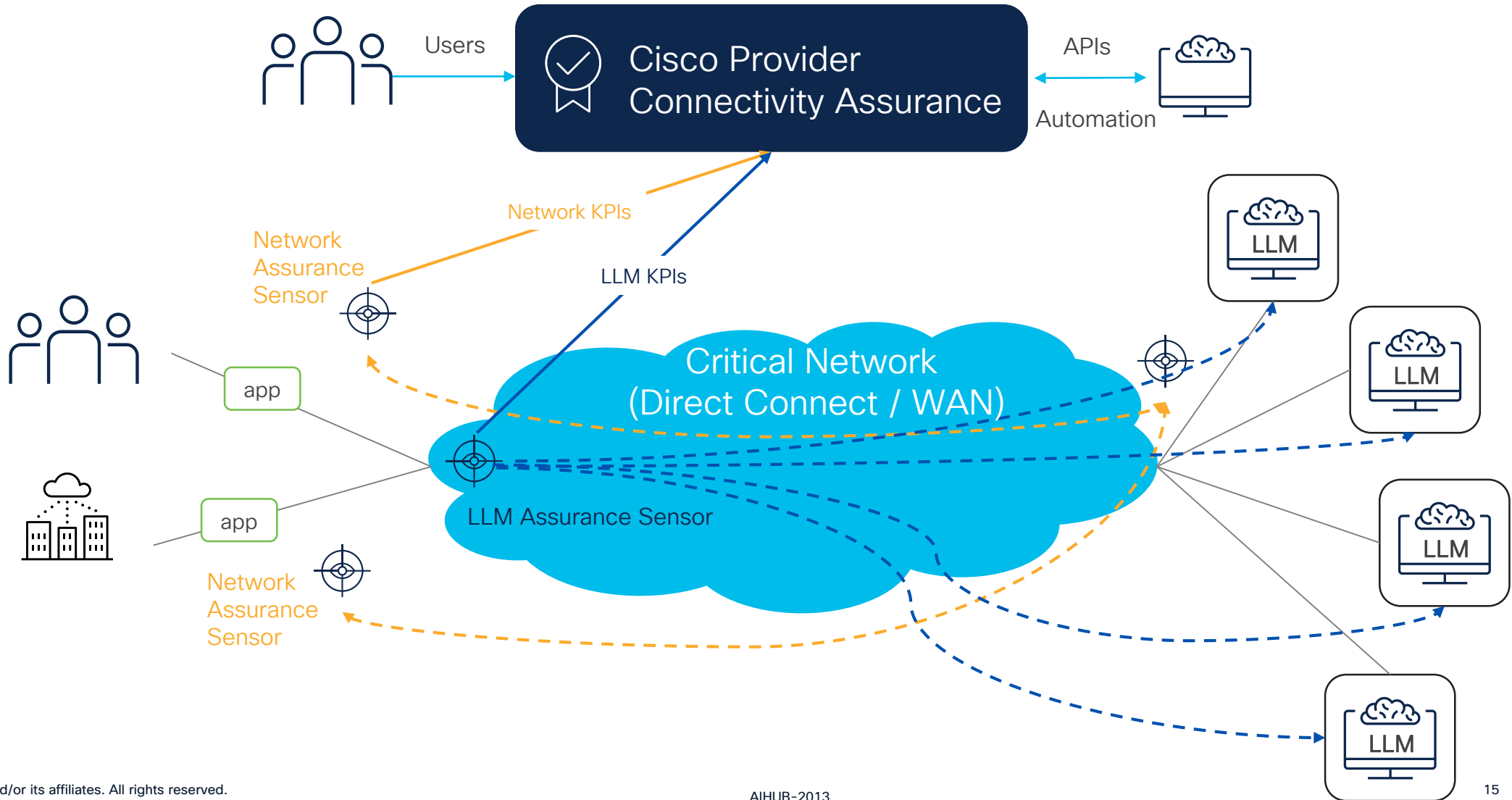
100% Transactions, applications, users

03: Cisco Provider Connectivity Assurance



Leverage integrations for further visibility and insights

03: LLM Assurance Agent (sensor)



03: Visibility Drives Insights – Assure and Optimize

Flexible Multi-Vendor and Multi-Layer Reporting and Analytics

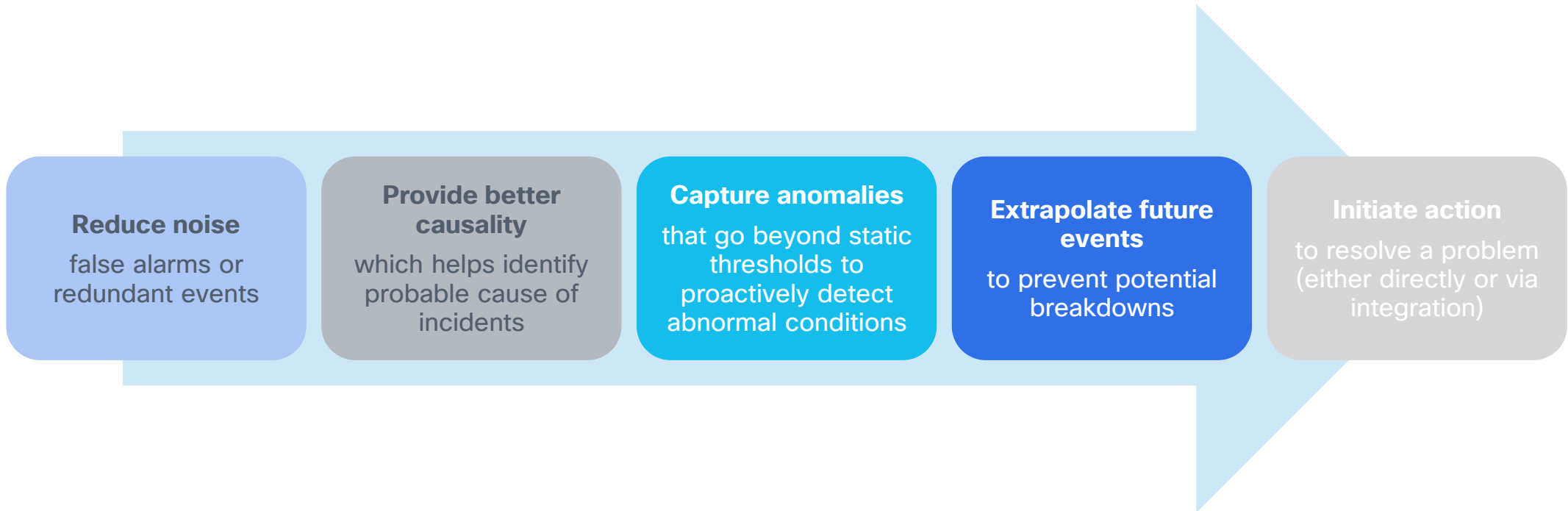
See, understand,
and improve your network

- Network-wide and end-to-end service visualization
- Granular, real-time metrics
- Multisource data ingestion
- Correlate performance across layers and data sources
- Ability to create SLAs for AI-driven services (monetization)
- AI-enabled, rapid anomaly detection and predictive forecasts



03: Visibility Drives Insights – Assure and Optimize

AI-Native Analytics



Provider Connectivity Assurance is built on a big data stack allowing it to leverage machine learning to enrich our customer's end-user experiences

03: Agentic AI Use-Cases

Reactive

Minimize impact of network incidents

- Time series anomaly detection
- Outlier detection
- Event clustering
- Root Cause Analysis

Improve operational KPIs
Reduce churn

Preventive

Reduce the occurrence of network incidents

- Toxic Factor Identification
- Automated maintenance
- Modeling and analysis
- Goal-based optimization

Improve Customer Experience
Optimized Operations

Prescriptive

Automatically align network operations to policy/intent goals

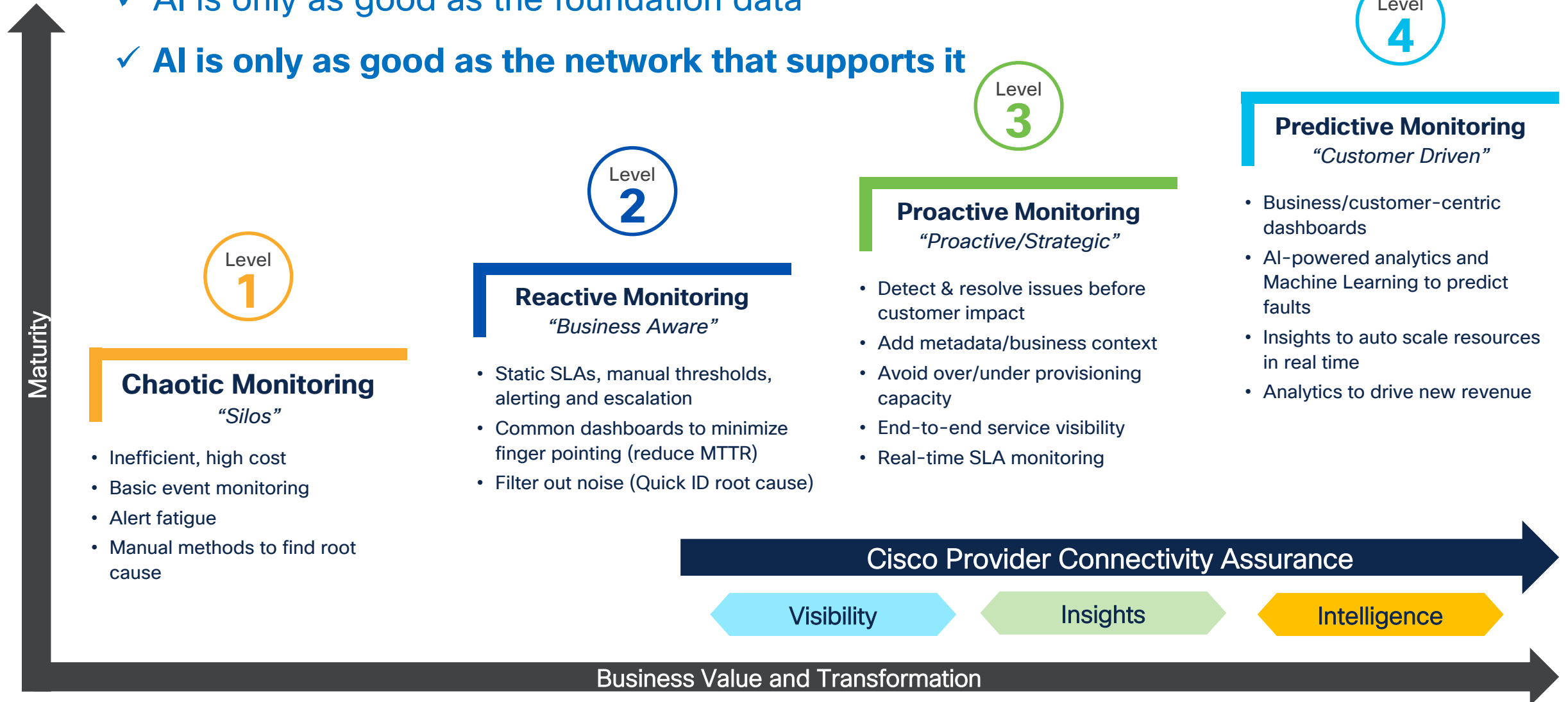
- Configuration recommendations
- Operational recommendations
- Continuous testing agents
- Co-Pilot code & config generation

Simplified Operation Lifecycle
Reduced Time-to-Market

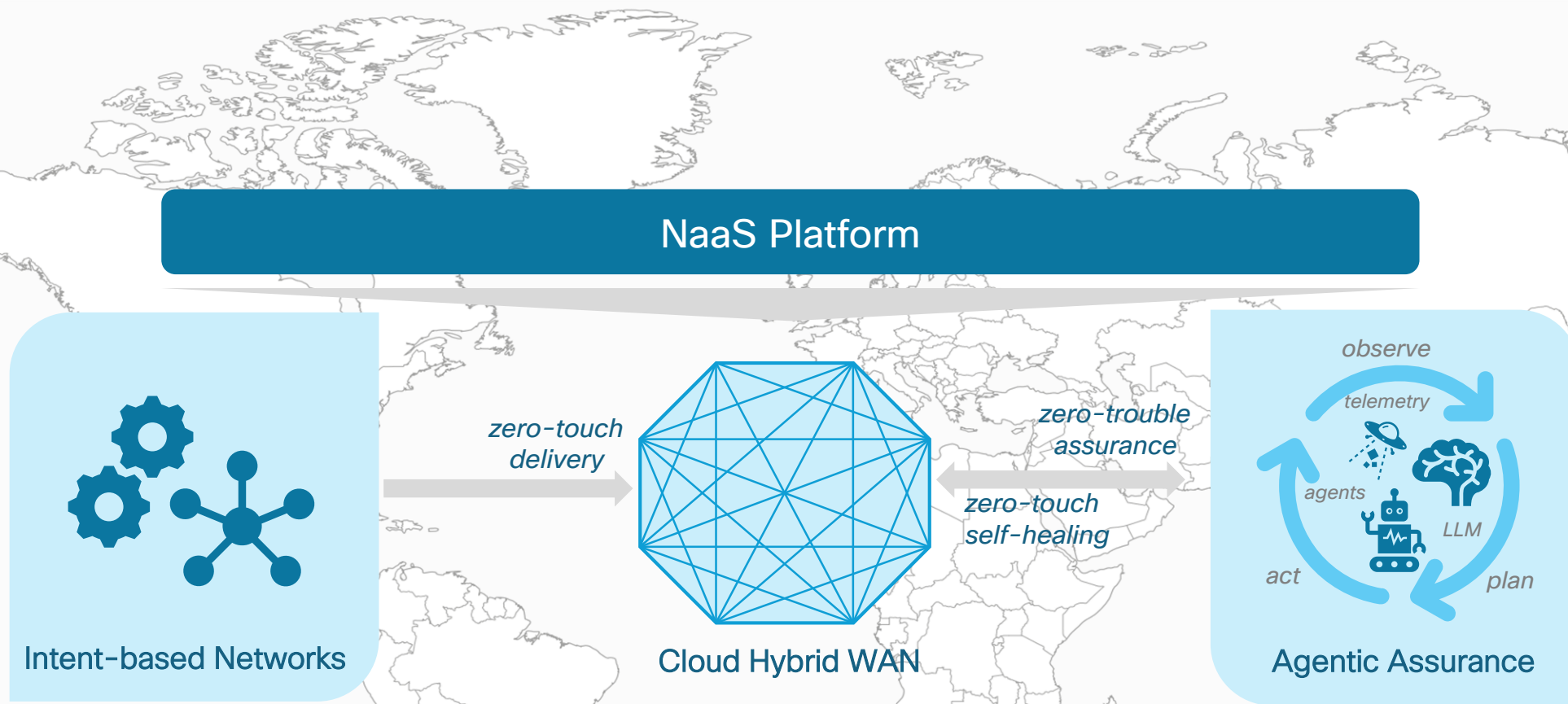
04 - Future Plans

It's A Journey To Predictive Assurance

- ✓ AI is only as good as the foundation data
- ✓ AI is only as good as the network that supports it



What Are Our Customers Asking For?



**05 - Would you like a demo of
our LLM Assurance Agent?**

Complete your session evaluations



Complete a minimum of 4 session surveys and the Overall Event Survey to be entered in a drawing to win 1 of 5 full conference passes to Cisco Live 2026.



Earn 100 points per survey completed and compete on the Cisco Live Challenge leaderboard.



Level up and earn exclusive prizes!



Complete your surveys in the Cisco Live mobile app.

Continue your education



Visit the Cisco Showcase for related demos



Book your one-on-one Meet the Engineer meeting



Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs



Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

Contact me at: pathomp2@cisco.com
<https://uk.linkedin.com/in/pbtradio>

Thank you

CISCO Live !

