

Advancing AI with Edge Computing: Best Practices for Scalable, Real-Time Intelligence

ITLGEN-2002

Daniel McGinniss

VP, Product Management, Cisco Compute,
Cisco

cisco Live !

IT Leadership Program

Everyone's talking about value.

82%

Feeling the heat on ROI in the last 6 months

69%

Making AI a top IT budget priority



But not everyone is ready to act

Only

13%

Are fully AI-ready



Questions your teams are asking...

Is our infrastructure
truly ready to scale
with AI demands?

Is our infrastructure
aligned to the
actual AI models
we need to run?

Are we prepared to
manage AI across
hybrid and edge
environments?

Can we access,
govern, and
trust our data for
AI use cases?

Can we secure
sensitive data
across the
AI lifecycle?

Are we building
to support new
AI without having
to rip and replace
in a year?

Are our AI tools
and platforms
integrated—or
creating friction?

Can we control
costs while proving
ROI on AI
investments?

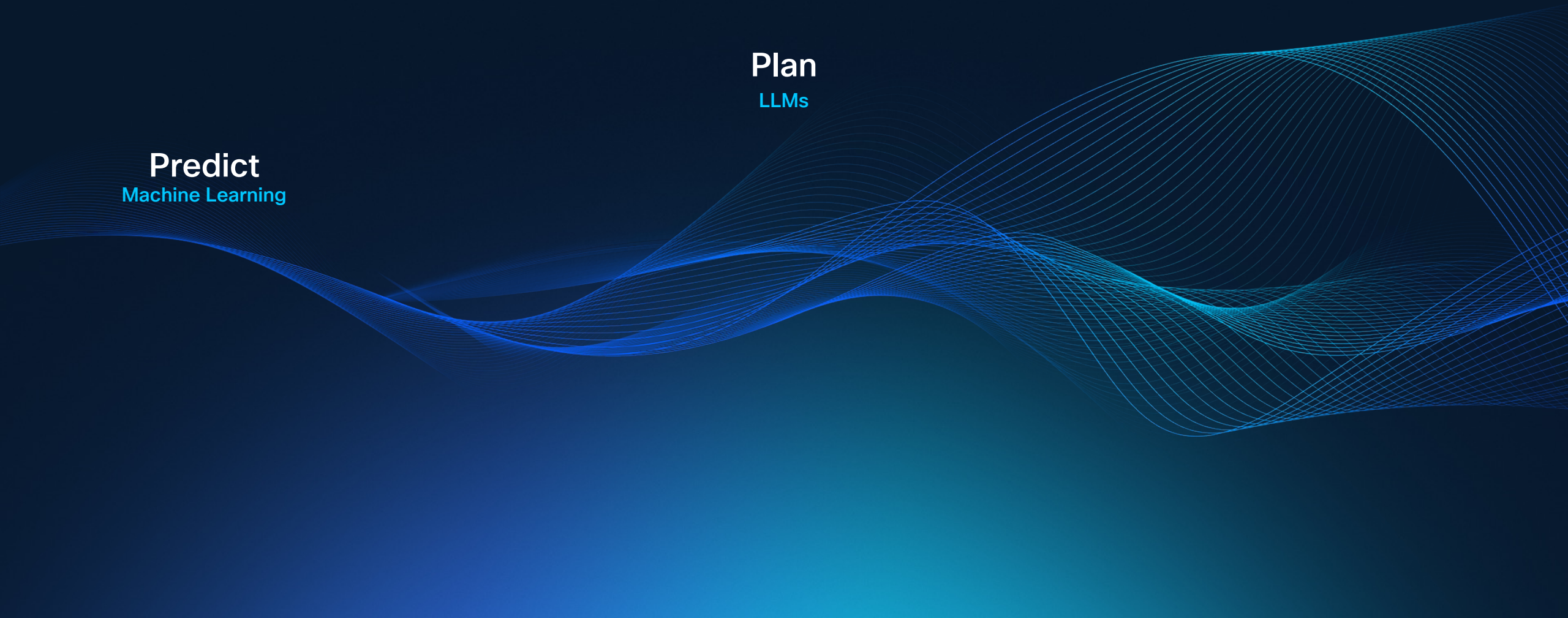
Do we have the
talent—and
buy-in—to
operationalise AI?

The next wave of AI

Predict
Machine Learning

Plan
LLMs

Act
Agents



AI workloads are moving to the edge

Data centre

DATA PROCESSING

AI is increasing demand for
local processing and networking

On premises edge

DATA PROCESSING

90%

2021

10%

25%

2027

75%

Why?

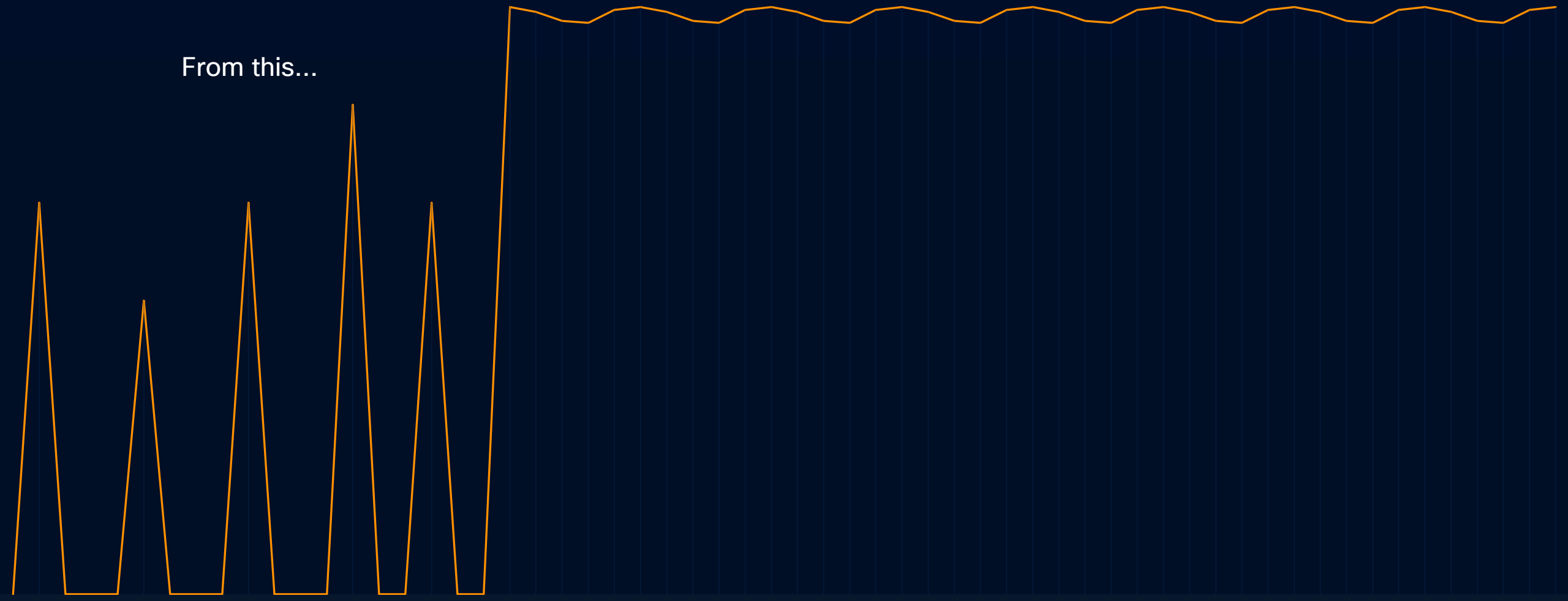
To process data where it's created



Agentic AI is a stress test for infrastructure

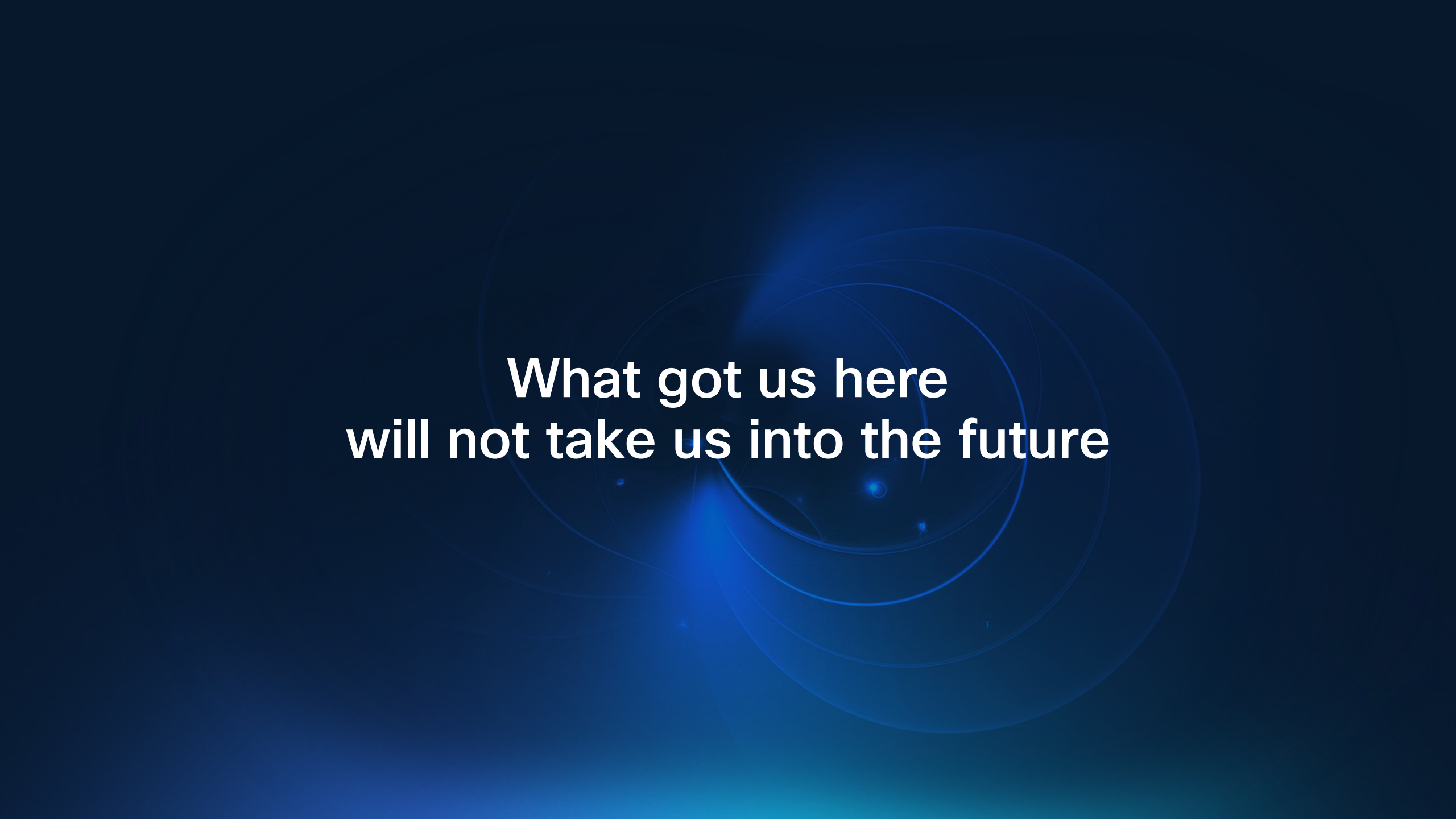
From this...

...to this?



The core and the edge are no longer separate worlds

**To compete in the AI era,
enterprises need an infrastructure strategy
that spans both**

The background is a dark blue gradient. In the center, there are several concentric, semi-transparent circles. A bright blue light flare or lens flare effect is visible, emanating from the center and spreading outwards, creating a sense of depth and movement. The text is centered over this graphic.

**What got us here
will not take us into the future**

IT's new role

Optimising AI from the core to the edge



What to do about it

Five best practices for AI success

1



Strategy

2

010110
110010
001011

Data

3



Scale

4



Economics

5



Trust

Start with strategic clarity

Define Use Cases for AI Strategically



● Pacesetters ● All companies

What this looks like in practice

- 1 Align on outcomes
- 2 Define KPIs
- 3 Create a prioritisation framework for use cases
- 4 Build for repeatable value
- 5 Measure the impact

Run AI where it makes the most sense

		BEST FOR	SECURITY ANGLE
AND not OR	 Cloud	Large-scale training, public data, scalability	Shared model – use Cisco tools for control and security
	 On-Premise	Sensitive data, regulatory compliance, tuning	Full-stack ownership, zero-trust enabled
	 Edge	Real-time inference, private/local decisions	Local data protection, fast + secure ops

Build Infrastructure that scales



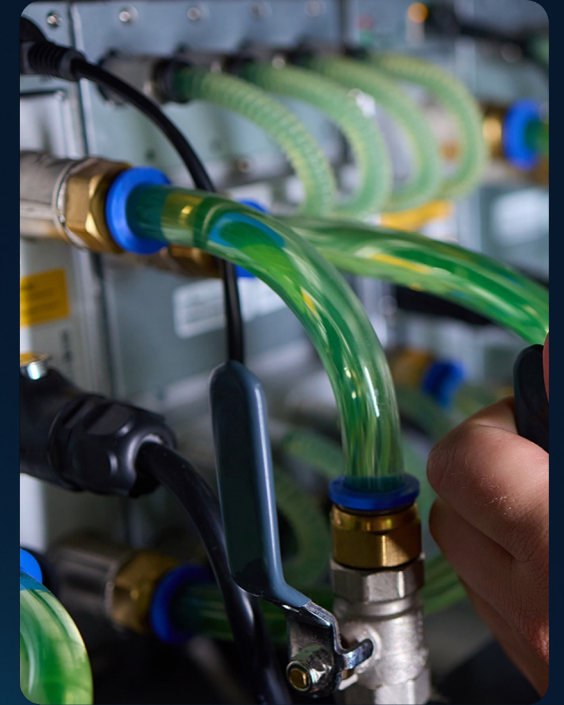
Build for
multi-tenancy



Design
for flexibility



Prevent AI
infrastructure debt



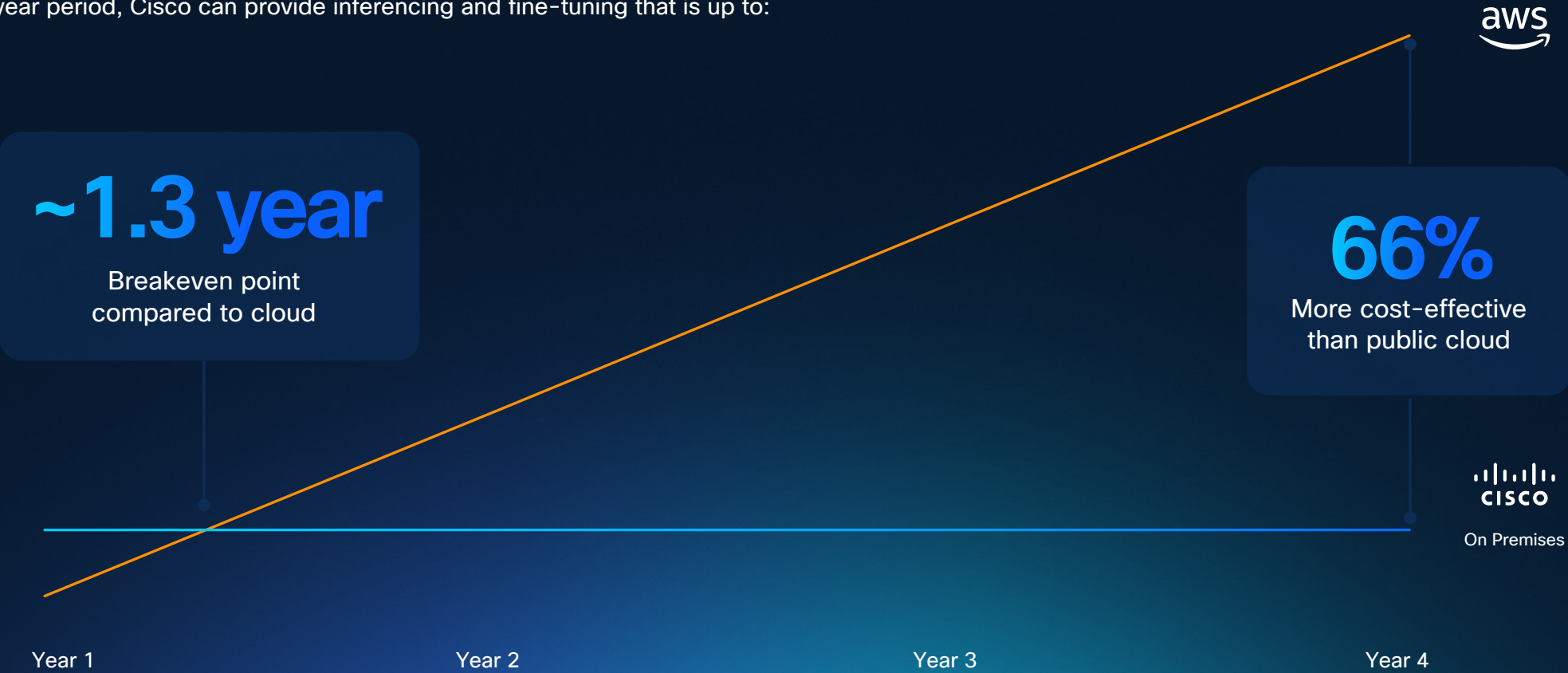
Optimise for
energy efficiency

The Economics of AI

Inferencing and fine-tuning on premises with Cisco is more cost-effective than cloud

Expected costs of a four-year period, with consideration of power, networking, and infrastructure

Over a four-year period, Cisco can provide inferencing and fine-tuning that is up to:



Best Practice #5

Build Trust into Every Layer

Security and visibility are what make AI at the edge real



Executive Strategy Questions

1

Where does latency
monetise in your business?

2

What data is too sensitive—
or too heavy—to move?

3

What's your real
egress profile?

4

How will you
govern thousands of
distributed agents?

5

Where should
your brain live?



Cisco can help you get there

Future-ready foundation. Unified operations. Trust everywhere.

