

Artificial Intelligence

Einordnung, Situation und Use Cases

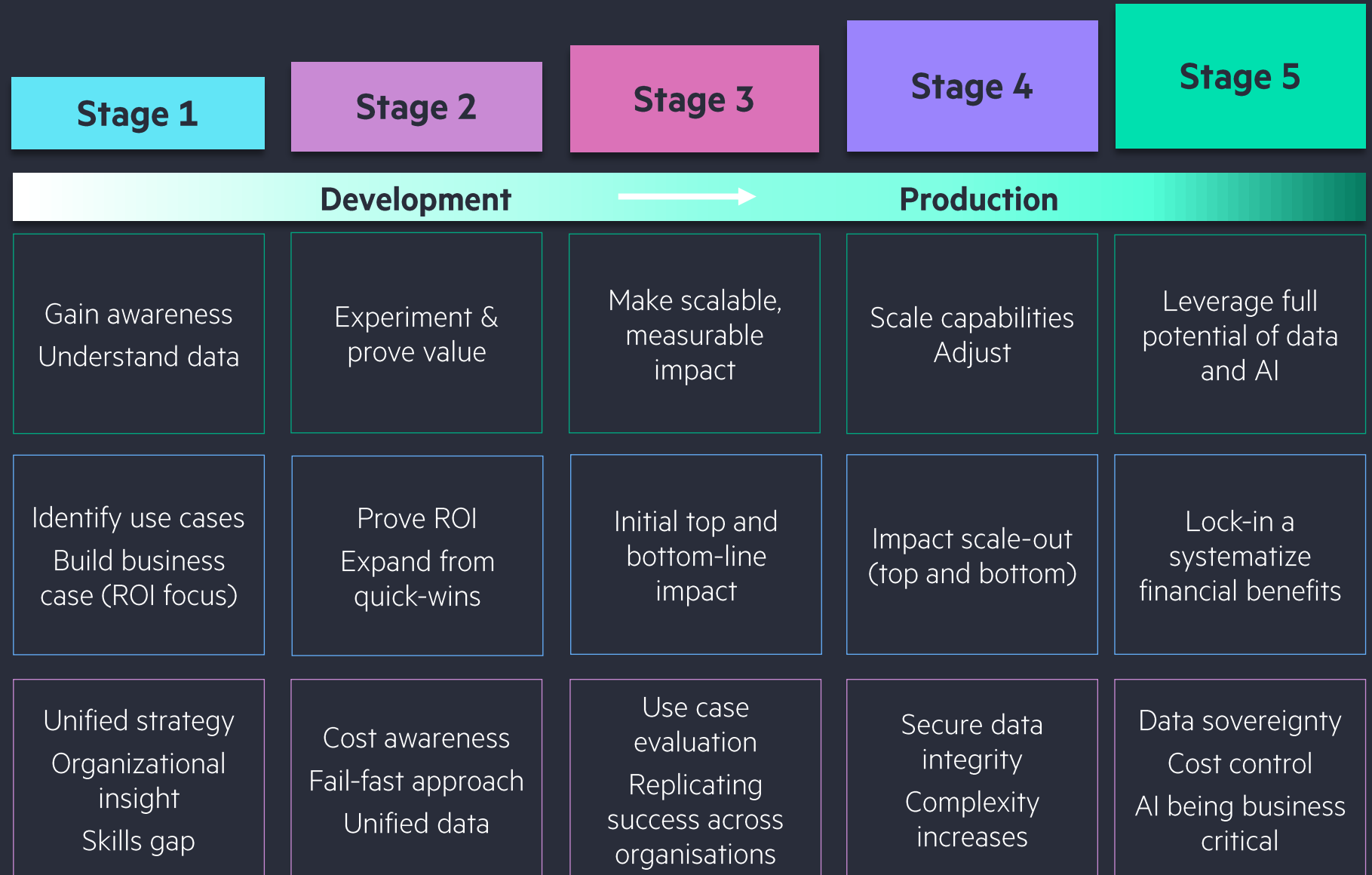
Stefan Brock
HPE AI Excellence
Center

92% of companies investing in
AI; **only 1%** have reached
deployment maturity₁

AI adoption reality check

1. **Maturity varies greatly across organizations**
2. **Differentiation depends on private data**
3. **Security & Sovereignty**
4. **AI is fundamentally a hybrid workload**
5. **Power and sustainability**

Where are enterprises on their AI journey?



The success factors in the implementation of AI

Developing a **clear AI strategy** for the organization

Building an **AI ecosystem** of partners and customers with shared AI ambitions

Creating **technical conditions**

Recognize the importance of data: establish a data-centric architecture and mindset

AI use cases aligned with strategic business

Strategic AI usage

Establishing a **corporate culture** and a growth mindset that is open to for innovation

Structural requirements for the implementation of AI

Create the right **organizational setup** to realize the potential of AI

Involve employees at all levels and further qualify them through AI training

AI's impact can span every industry

Finance



Fraud detection
Personalized banking
Investment insights

Healthcare



Molecule simulation
Drug discovery
Clinical trial data analysis

Retail



Personalized shopping
Automated catalog descriptions
Automatic price optimization

Telecommunications



AI virtual assistants
Network performance tuning
Remote support capabilities

Media & entertainment



Character development
Video editing & image creation
Style augmentation

Manufacturing



Factory simulation
Product design
Predictive maintenance

Federal



Document summarization
Audit compliance
AI virtual assistants

Energy



Knowledge base Q&A
Predictive maintenance
Customer service

Virtual assistants

Automation

Product and content creation

Common enterprise AI use-cases



Code generation

“Co-pilot” for SW engineers to generate code faster—on prem with own codebase and coding standards



Document creation

Automate form-filling—insurance, prescriptions and more; create marketing briefs, parts manuals and more



Search engine

Enhance search results with internal documents and information



Q & A chat

Create an internal support chatbot with your own data for an enhanced knowledge base and faster resolutions



Customer service

Boost productivity for customer service teams and enhance customer experiences with optimized call centers

From vision to value

Barriers to achieving scalable AI



#1 Data

Data strategy

Sovereignty
and privacy

Ensuring regulatory
compliance



#2 Complexity

Gaps in specialized
expertise

Slow ROI



#3 Scale

Growing computational
demand

Exploding data volumes



#4 Cost

Inference at
massive scale

Spiraling cloud costs



#5 Security

Risks of exfiltration
or IP leakage

Ensuring AI governance

AI factories provide a blueprint for AI success

From pilots to profits and purpose-driven outcomes

AI today

High failure rate

Scattered across silos and teams

No clear plan for operational scale

Fragmented operating model

AI factory



Measurable and predictable



Centralized workflows and reusable assets



Speed and predictability of deployment



Unified environment

Outcomes

Faster time to AI ROI

Repeatable AI success

AI innovation at scale

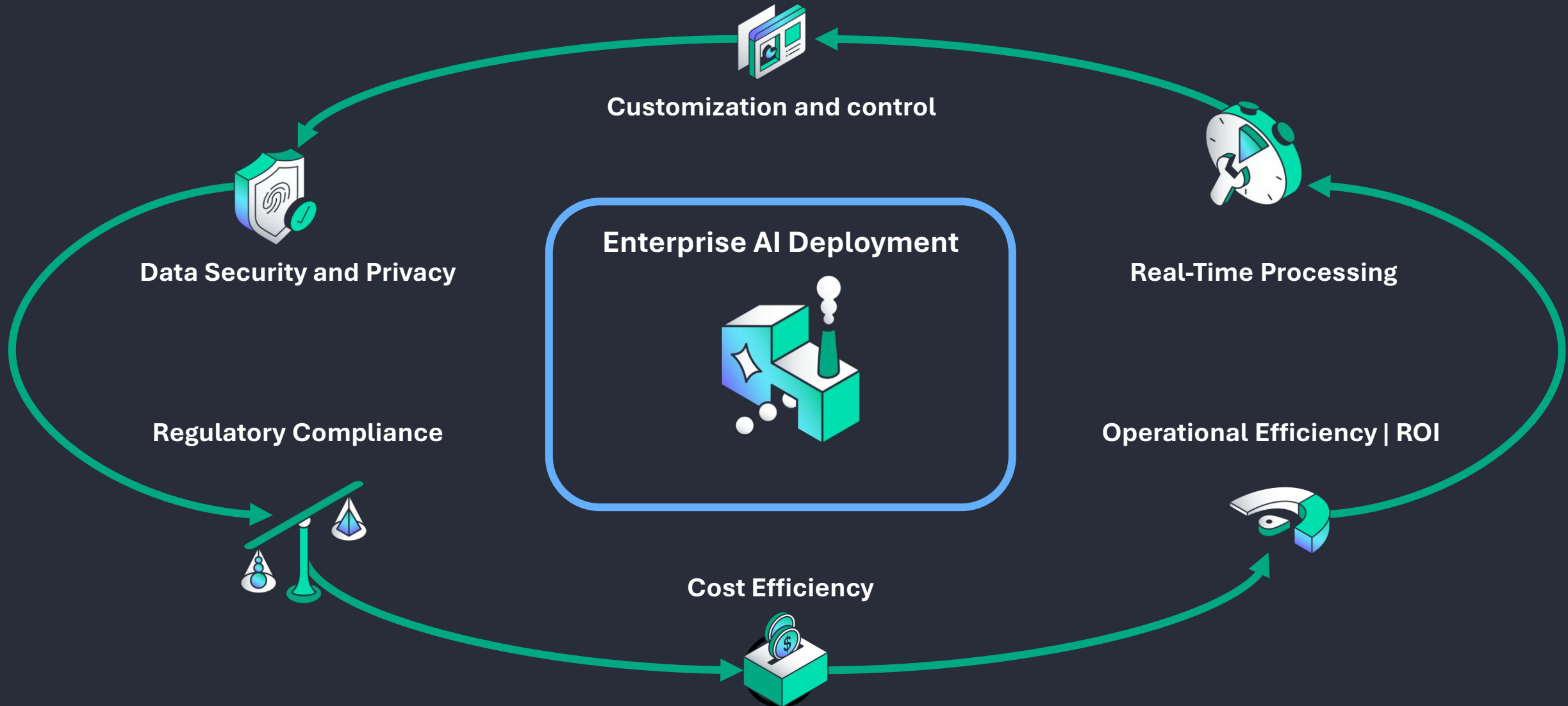
Secure governance & control

Welcher KI Ansatz?

Choosing the infrastructure and services for your AI ambitions

Stefan Brock
HPE AI Excellence
Center

Key drivers for hybrid AI factory adoption



Expanded AI Factory Portfolio from HPE

for every AI ambition, across clouds, cores and countries

Turnkey AI factory

General enterprises

AI factory at scale

Model builders & SPs

Sovereign AI factory

Governments, public sector

Common control plane: HPE Morpheus and HPE OpsRamp

- Demand rapid ROI
- NVIDIA Software-preference
- Inference & tuning
- Air-cooled
- Up to 64 GPUs: RTX pro 6000, H100/200

- Tailor to your scale and scenario
- Services integrated software stack
- Model dev, training & inference
- Direct liquid & air-cooled
- 100s-10,000s GPUs: H200, B200

- Independence from others
- Strict data sovereignty
- Model dev, training & inference
- Direct liquid & air-cooled
- 100s-10,000s GPUs: H200, B200

Turnkey, engineered systems

Customized, validated solutions

Infrastructure | Software | Services | Ecosystem | Sustainability

AI Implication Matrix

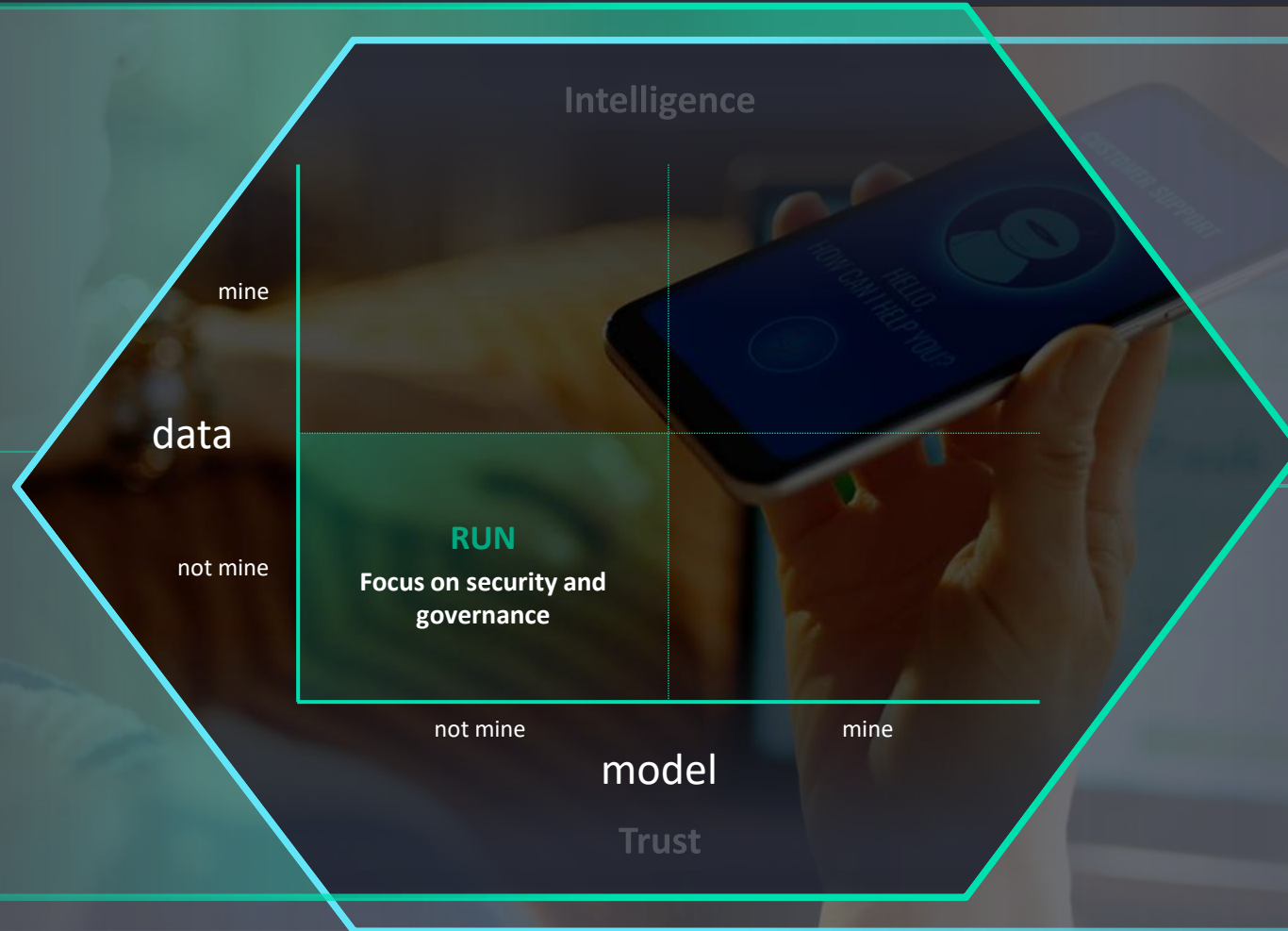
RUN: functionality infused into existing applications

Experience

Digitization

Engineering

Platforms



HPE OpsRamp Software

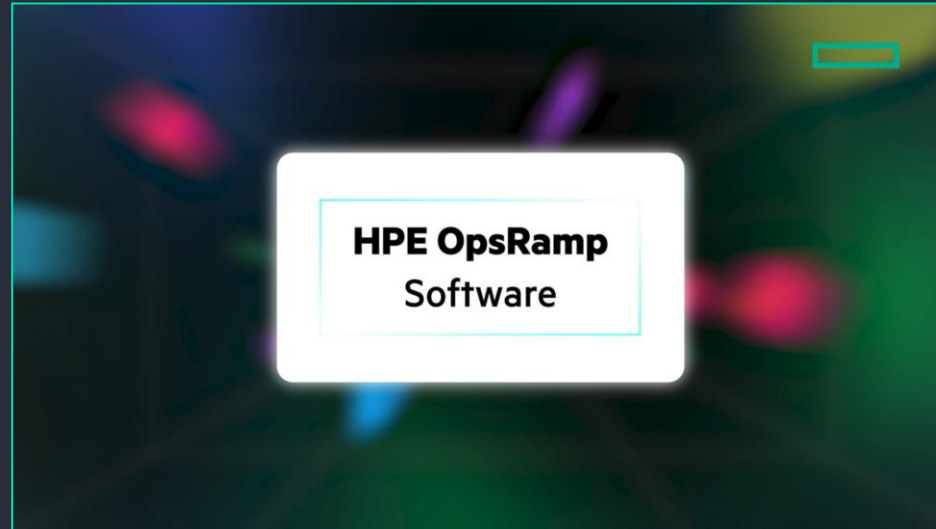
Transform IT operations with deep visibility, AIOps and smart automation

Autonomous IT ops, powered by full-stack observability and AI

Empower enterprise IT teams with full-stack observability, predictive insights, and intelligent automation across complex hybrid environments — on-premises, multi-cloud, and cloud-native.

Speed issue detection and resolution

Identify anomalies in telemetry data and correlate alerts with AIOps and machine learning to quickly identify root cause and accelerate issue resolution.



Intelligent Automation

Act faster on the insights generated by AI-powered analytics, automating routine tasks such as event and incident management, configuration, and capacity management.

HPE GreenLake Intelligence

Supercharge your IT operations with a mesh of intelligent AI agents

Deep intelligence where it matters

AI agents trained to understand enterprise IT from workloads to infrastructure and act on your intent.

Break down the silos

Agents solve problems across layers and domains, working together with continuity and context.

Bring the opportunities to you

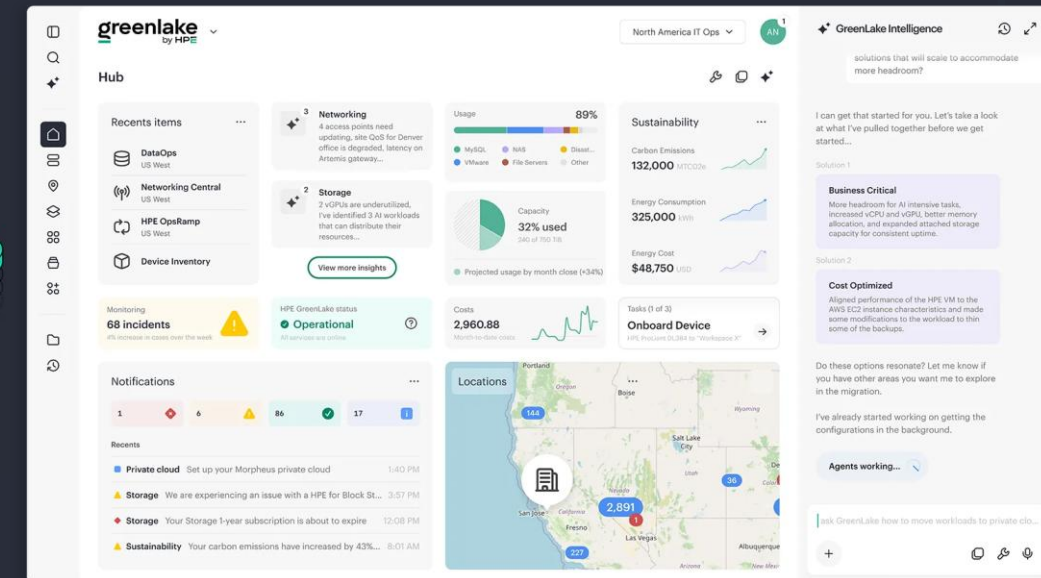
Always-on, proactive recommendations enable intelligent optimization.

Action, not just answers

Eliminate manual tasks with autonomous IT. Agents take action, subject to your approval.

HPE GreenLake Intelligence

Acting like a 24/7 intelligent operations team that can reason to solve problems across your hybrid IT estate. It's like giving your infrastructure the ability to think, learn, and act - so your teams can move faster, reduce risk, and unlock more value from your data.



AI Implication Matrix

RAG: integrate a model (or multiple models) with corporate data

Experience

Engineering

Digitization

Platforms



HPE ProLiant Compute + NVIDIA RTX Pro 6000

Level-up your compute infrastructure and develop your own solution

DIY

Turnkey AI factory

AI factory at scale

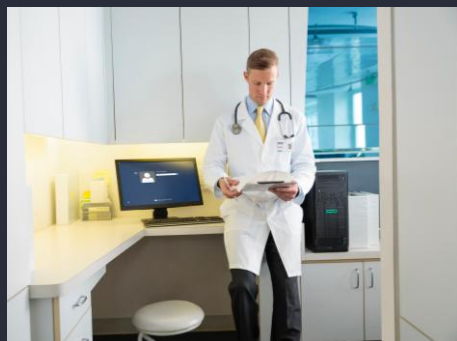
Sovereign AI factory

Bring your own tools

GPU consolidation

Reduce TCO and power

Use cases



Multi-modal AI inference



Fine tuning



Design & video



Physical AI



Edge

Agentic AI: Enabling autonomous intelligence at the edge



- **Architecture:** NVIDIA agentic AI software stack with HPE ProLiant edge platforms
- **Agentic AI blueprints** so agents that can reason, plan, and act rather than simply respond to prompts
- **Flexible right-sizing, low-latency inference** on HPE ProLiant DL145, DL345, and DL380 Gen12 & more with NVIDIA RTX PRO 4500 Blackwell

Local Inputs
Sensors, cameras, files, telemetry



NVIDIA Blueprint + Microservices Layer

+
NVIDIA Blueprints
(Agentic AI reference workflows)
+
NVIDIA NIM microservices
deploy/serve LLM/VLM inference as standardized microservices



NVIDIA RTX PRO Server Software Kit

NVIDIA AI Enterprise
NVIDIA Omniverse
vGPU



HPE ProLiant Compute

HPE iLO + HPE Compute Ops Management
NVIDIA RTX PRO 4500 Blackwell
32GB, 165W, PCIe Gen5, up to 2 MIG, confidential compute



Ruggedized



ProLiant DL145

Entry Rack



ProLiant DL345

Performance



ProLiant DL380

HPE Private Cloud AI

Enterprise AI infrastructure simplified

DIY

Turnkey AI factory

AI factory at scale

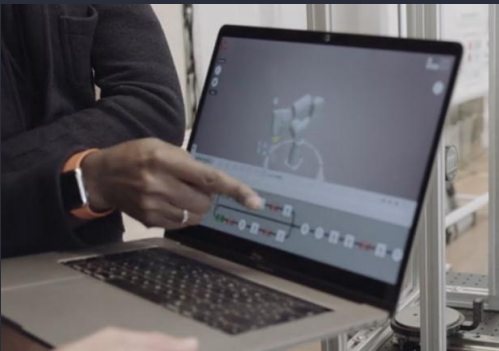
Sovereign AI factory

Instant AI
productivity

Secure and unified
data access

End-to-end AI
software platform

Use cases



Development



Computer vision



Agentic AI



Physical AI



RAG

HPE Private Cloud AI

Turnkey AI factory

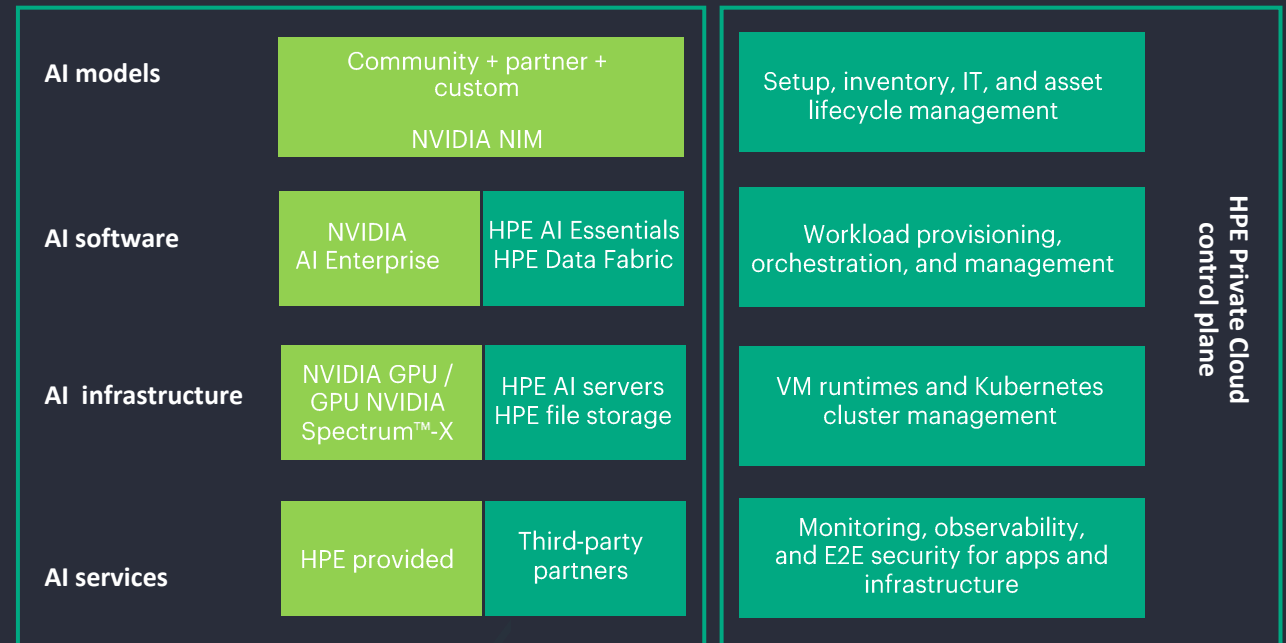
AI factory at scale

Sovereign AI factory

AI ecosystem that rapidly deploys generative and agentic AI without constant IT intervention

Rapidly move from AI pilot to production with a turnkey AI private cloud

Leverage an out-of-the-box AI ecosystem and low-code tools that empower teams to rapidly deploy generative and agentic AI solutions without constant IT intervention.



HPE Private Cloud AI

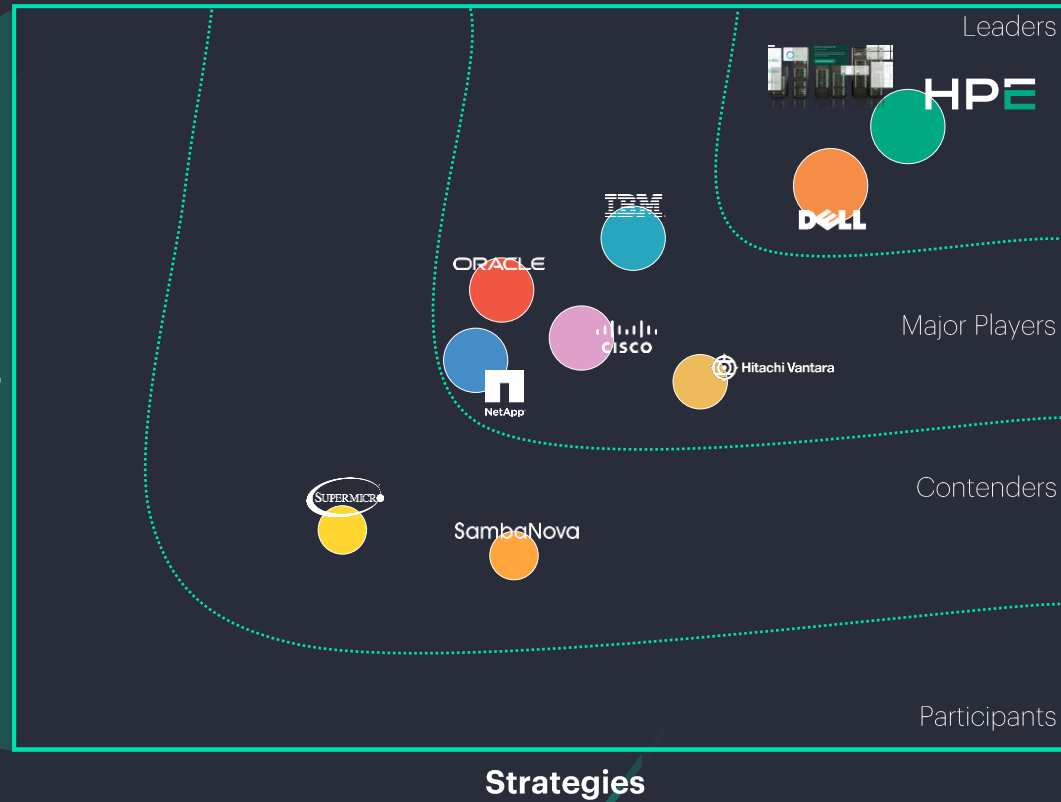
IDC Leader in Worldwide Private AI Infrastructure Systems 2025



Worldwide Private AI Infrastructure
Systems 2025
Leader



Capabilities



Meet the family – HPE Private Cloud AI

Next Gen AI-optimized and scalable hardware

Start fast

Learn | Prototype | Prove



Developer system

2 NVIDIA H100 NVL GPU's

32 TB Integrated

Customer Network

up to 2.2 kW

Inference | RAG



Small

8 NVIDIA RTX Pro 6000 GPU's

109 TB

400GbE NVIDIA Networking

up to 12 kW rack

Inference | RAG | Fine Tune



Medium

8 NVIDIA H200 GPUs

109 TB

400GbE NVIDIA Networking

up to 13 kW rack

Inference | Large RAG | Fine Tune



Large

16 NVIDIA H200 GPUs

217 TB

400GbE NVIDIA Networking

up to 17.4 kW rack

Scale fast



NEW

Expansion rack

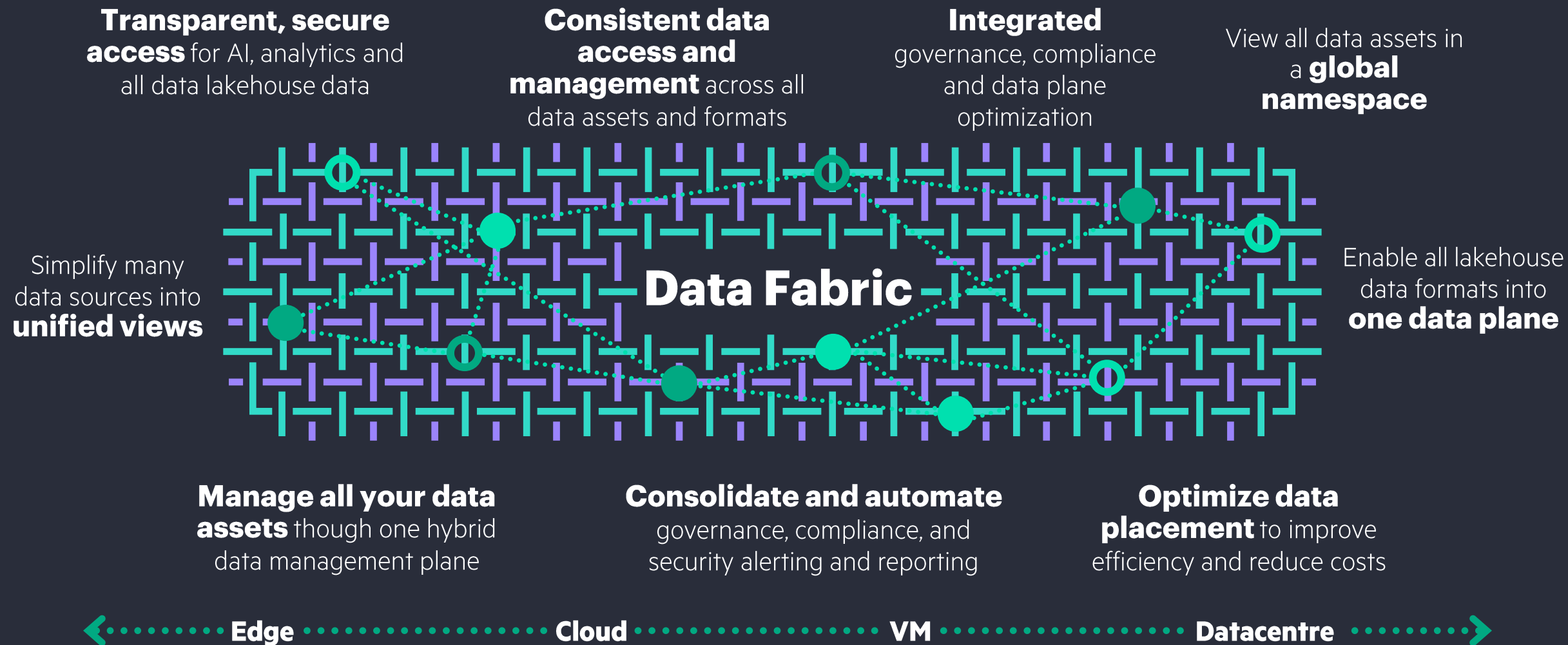
16 NVIDIA H200 GPUs

-

400GbE NVIDIA Networking

up to 12 kW rack

Delivering a unified approach to hybrid data lakehouse architectures



Thank You

Stefan.brock@hpe.com