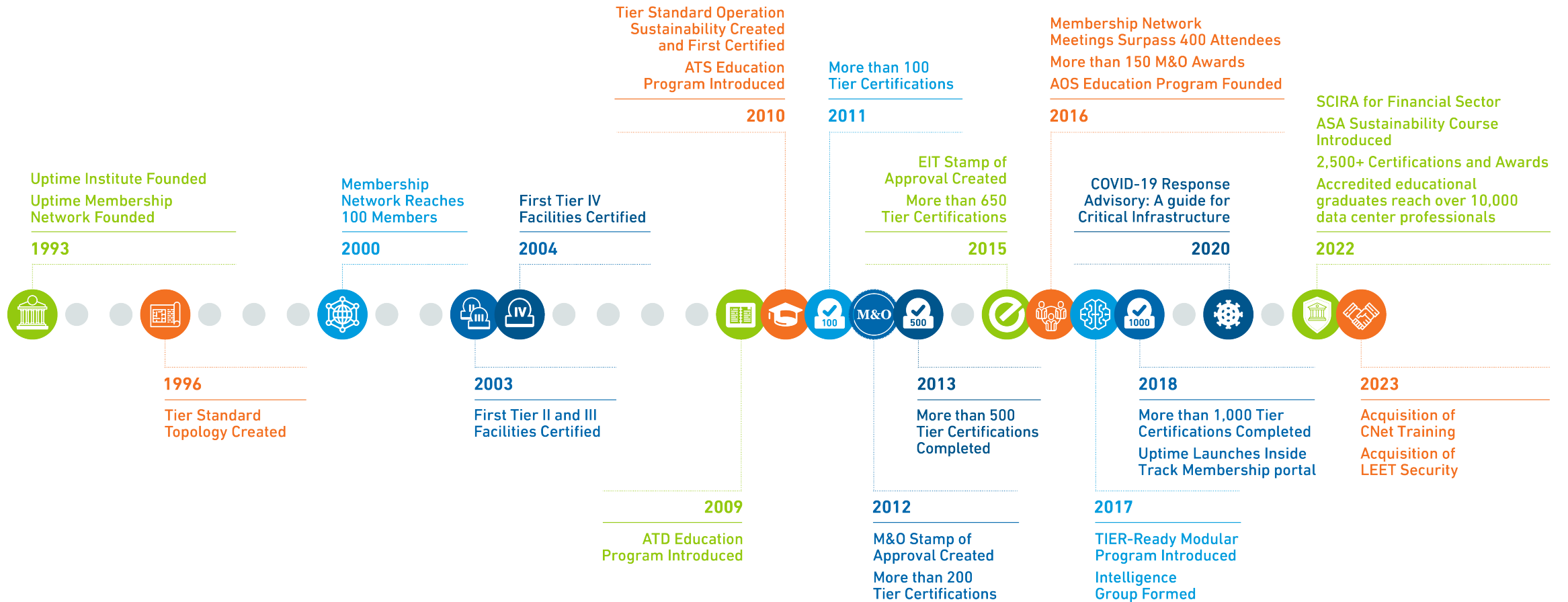


Five Data Center Predictions for 2025

Tomas Rahkonen PhD

Research Director
Uptime Institute

Uptime Institute Timeline



What We Stand for

Global Experience and Expertise

Infrastructure Risk Reduction

Operational Effectiveness and Resiliency

IT Advisory Services

Infrastructure Cost Management

Sustainability Services

Professional Development

Intelligence and Research

Membership Community



Uptime Institute means trust,
and we take pride in our clients' success.

Actionable insight for the digital infrastructure ecosystem

This information is based on research by the Intelligence team in conjunction with dozens of other Uptime experts.

The Intelligence service includes access to research reports with 150+ new publications annually, as well as exclusive webinars and roundtables, and analyst consultations.

Topics include:

- Artificial Intelligence in data centers
- Enhancing resiliency strategies
- Innovations in cooling systems
- Sustainability and efficiency

Evaluation Access:

Request Evaluation – Uptime Intelligence

<https://intelligence.uptimeinstitute.com/request-evaluation>



Intelligence Predictions for the Year (and Years) Ahead



We aim to highlight crucial, yet often overlooked topics

Many challenges facing the industry result from the ongoing success of the IT sector

Prediction 1

Data Center Resource Use Will Raise Deep Questions — And Opposition

Prediction One: Data Center Resource Use Will Raise Deep Questions — And Opposition

Data center developments will face greater scrutiny as they consume ever more energy and water - and emit more carbon.

Unqualified Governments support for data center investment, often because of AI's potential, will be challenged.

Developers and operators will increasingly have to be more transparent, more collaborative, and contribute more.



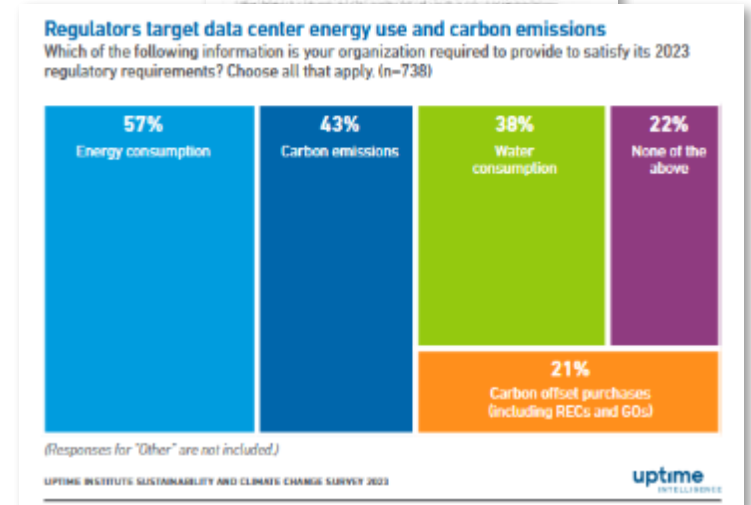
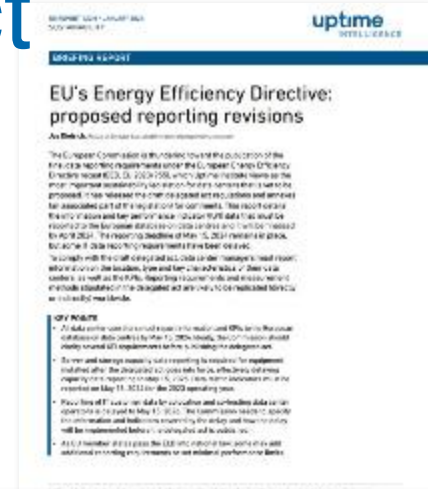
Everything Is Bigger, With More Impact

With or without AI, data centers:

- Are growing dramatically in size and number
- Use 20%-30% of power in some regions
- Consume more water
- Excluding offsets, release ever more carbon

Many data center operators:

- Currently enjoy government support and favored status
- Receive tax breaks and have Critical National Infrastructure status
- Will miss or push out Net Zero targets
- Employ few people despite incentives
- May not be perceived as local or benign, except by governments



The New Oil...but Not in a Good Way

Governments experience policy conflicts: Promote AI and investment v. prioritize sustainability

Sustainability groups will target data centers directly – often through the courts.

Governments and stakeholders will increase ask: What are you doing in your data center? Transparency will be a new battleground.

Data center operators will seek to turn the focus onto the IT operations – causing disputes and policy disagreements.

High profile will lead to more security threats



Prediction 2

Most AI Models Will Be Trained in the Cloud

Components of AI Training

Before an AI can be used (inferencing), it must be trained


Model

Training data


Infrastructure

Retraining strategy


A “Good” AI

Increasing customization and control 	
	Good
Model	Foundation (pre-trained) model
Training data	Included, general-purpose
Infrastructure	Included
Retraining strategy	Included

A “Better” AI

	Increasing customization and control 	
	Good	Better
Model	Foundation (pre-trained) model	Foundation model + fine-tuning, RAG
Training data	Included, general-purpose	Addition of custom data
Infrastructure	Included	Likely cloud
Retraining strategy	Included	Updates plus occasional fine-tuning

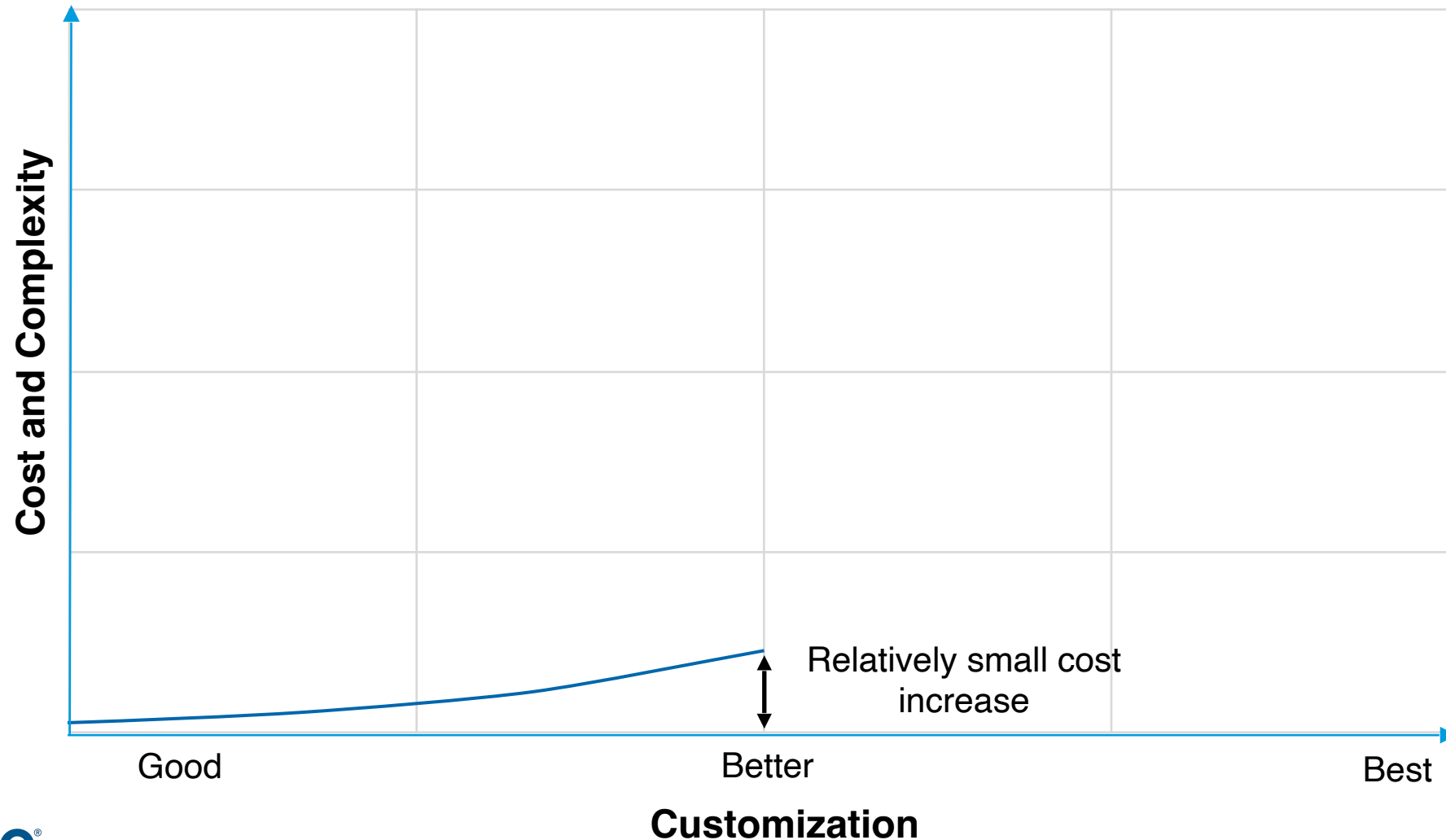
The “Best” AI

	Increasing customization and control 		
	Good	Better	Best
Model	Foundation (pre-trained) model	Foundation model + fine-tuning, RAG	New fully trained model
Training data	Included, general-purpose	Addition of custom data	All custom
Infrastructure	Included	Likely cloud	Dedicated
Retraining strategy	Included	Updates plus occasional fine-turning	Regular fine-tunes and full retrains

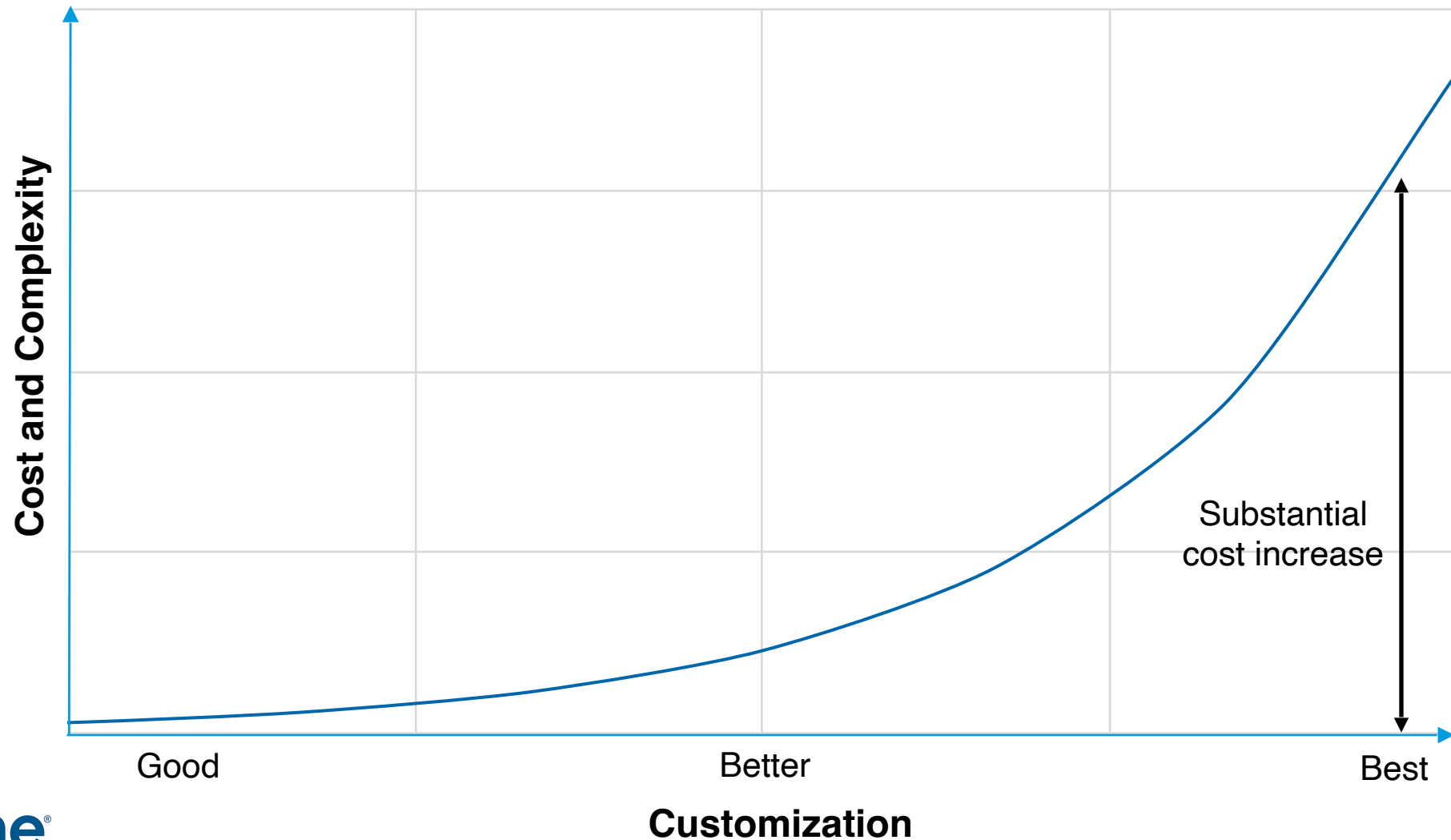
Improvement Comes at a Cost

	Increasing customization and control		
	Good	Better	Best
Model	Foundation (pre-trained) model	Foundation model + fine-tuning, RAG	New fully trained model
Training data	Included, general-purpose	Addition of custom data	All custom
Infrastructure	Included	Likely cloud	Dedicated
Retraining strategy	Included	Updates plus occasional fine-tuning	Regular fine-tunes and full retrains
	Exponential increase in cost and complexity		

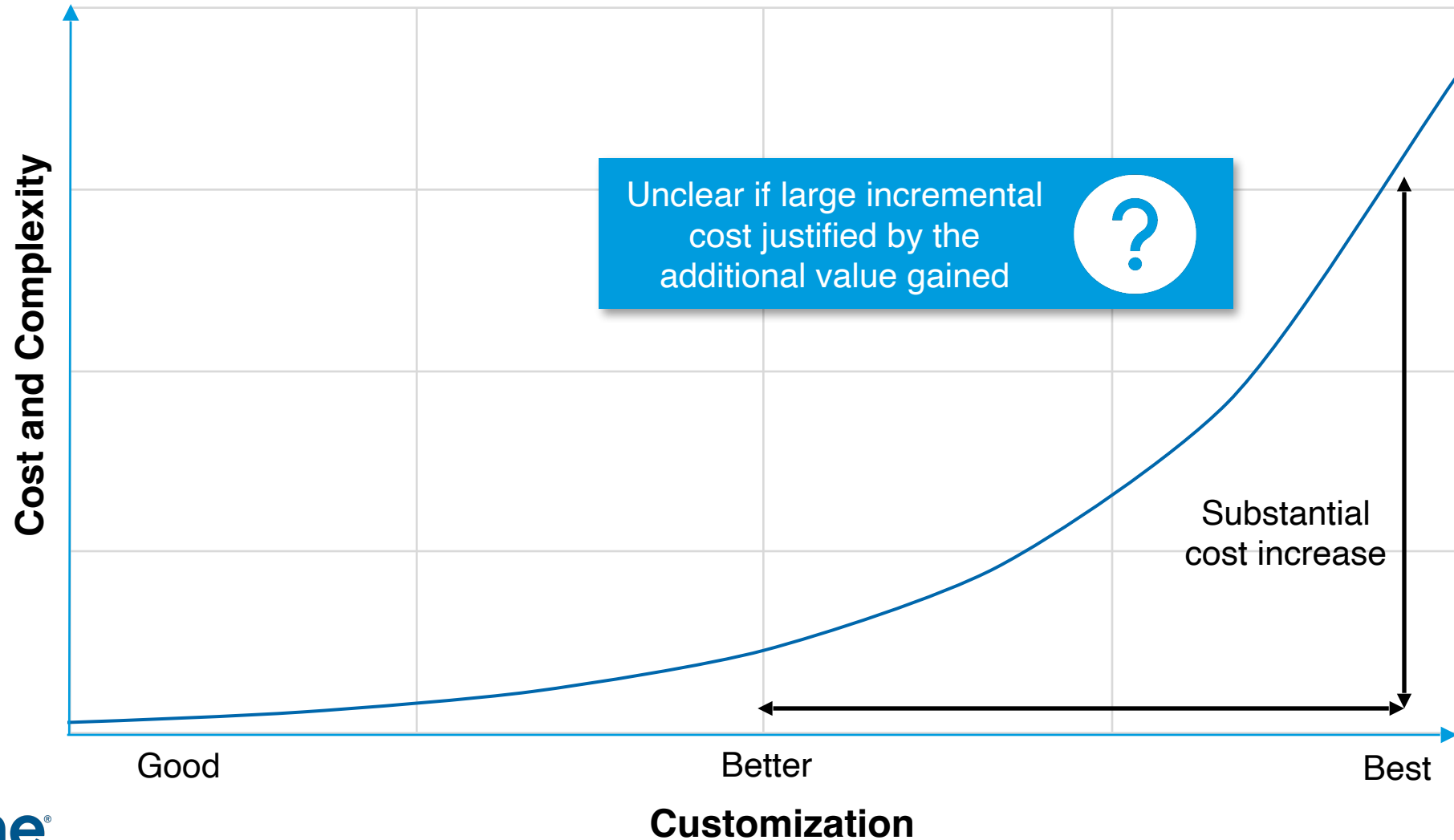
Some Customization Possible for Low Cost



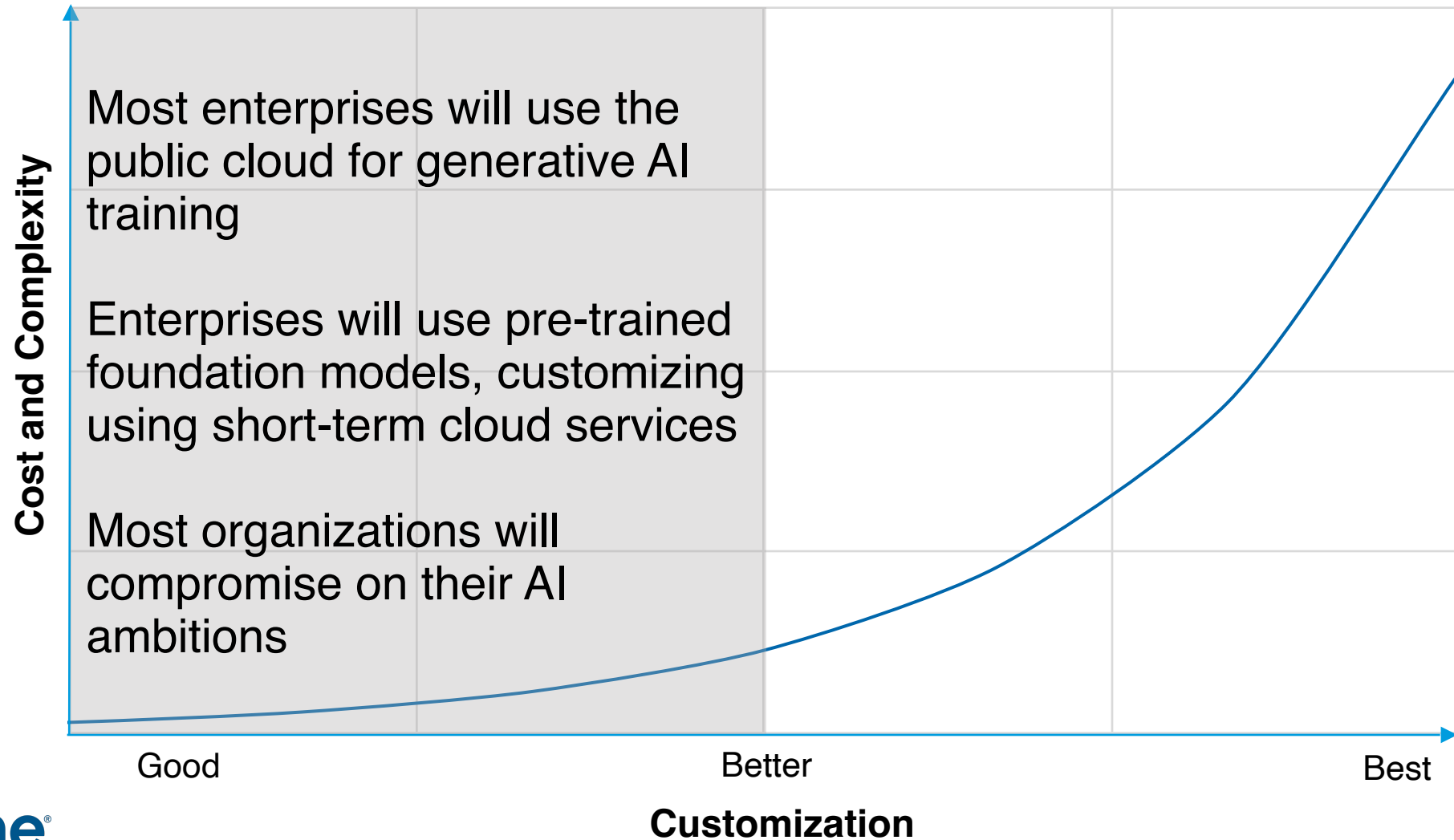
Fully Customized is Very Expensive



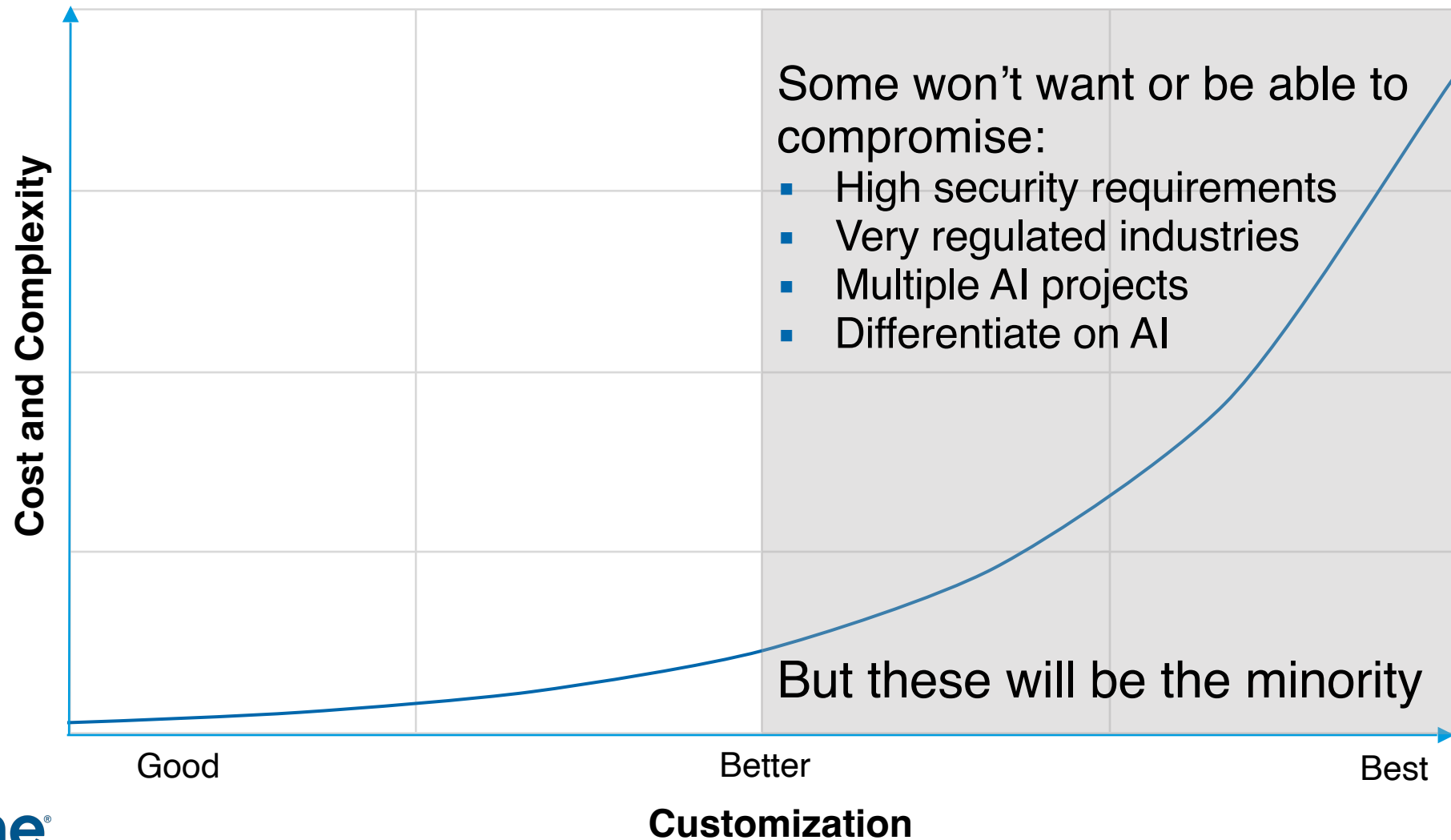
And Not Necessarily Justified



In 2025



In 2025



Prediction 3

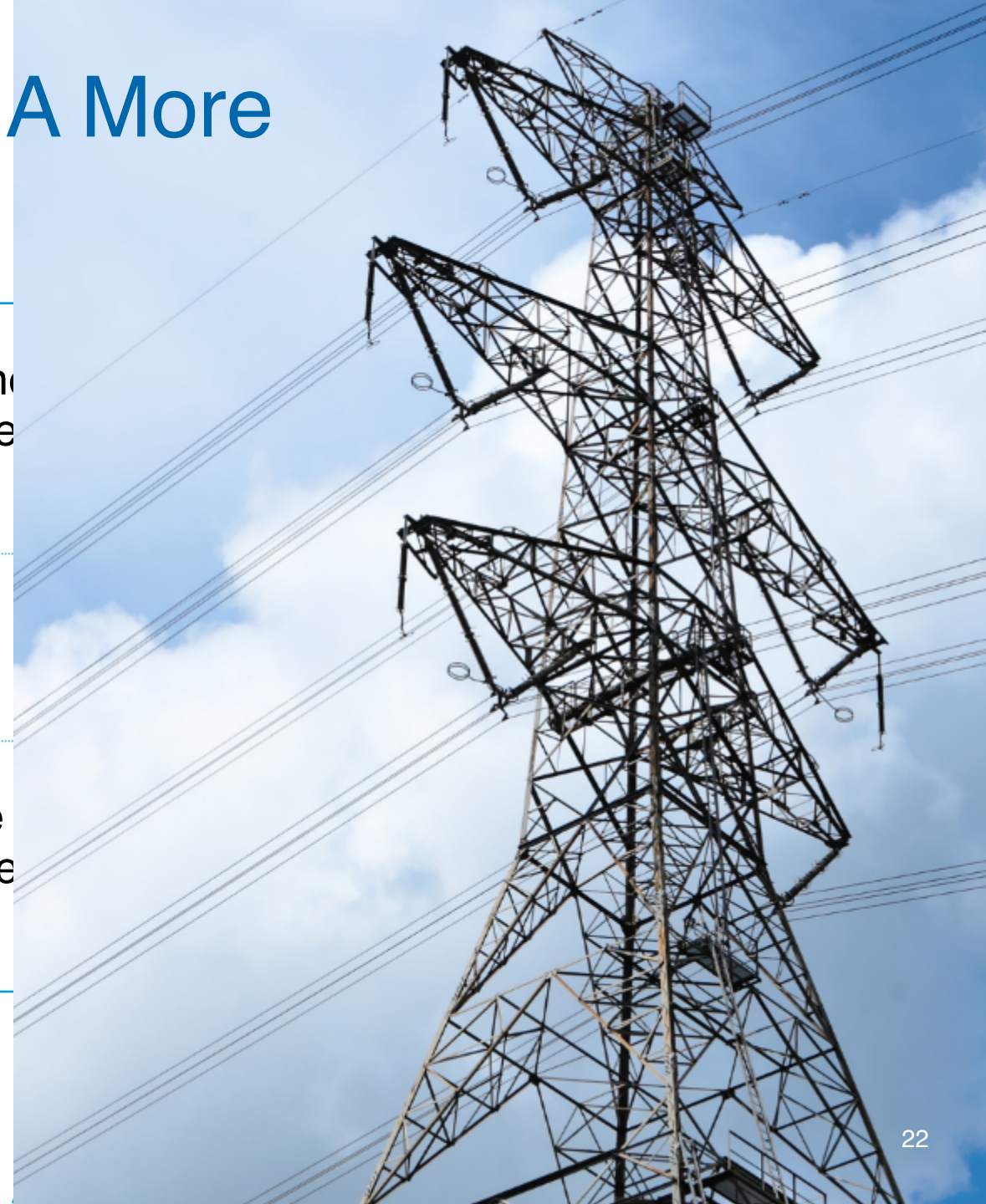
Grid Demand Will Require Active Participation From Data Centers

Data Centers and the Grid: A More Transactive Relationship?

Data center planners will find it difficult to find or generate sufficient power. Power companies will apply more leverage and seek more collaboration.

AI will present challenges because of its large and sometimes erratic demands.

New and expanded data centers will increasingly be expected to provide or store power, and possibly even shed loads.



The Context: Power Grids Under Pressure

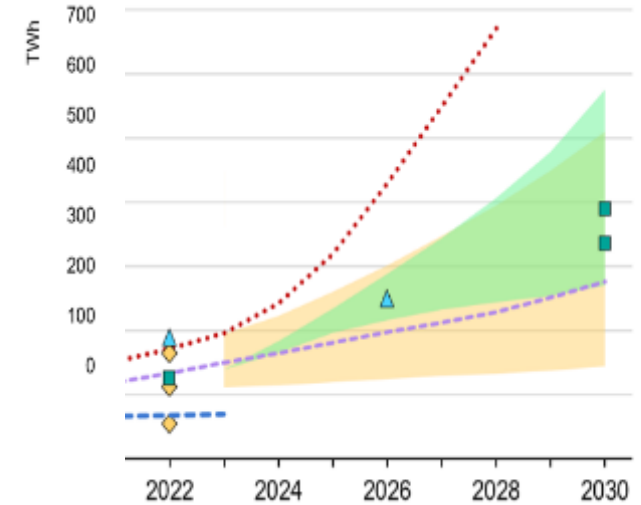
Power companies need to meet growing demand while also reducing carbon emissions.

Data centers are not the only challenge – EVs, air conditioning, decarbonization drive up demand.

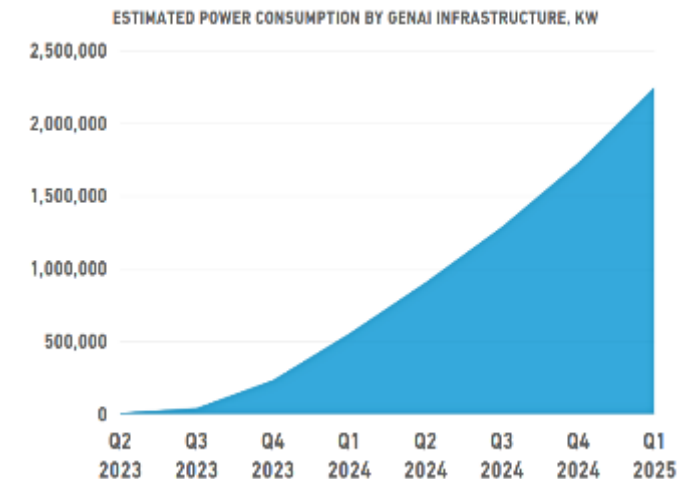
Intermittent renewable power sources and unstable demand add reliability challenges.

Data center concentration in some grid regions has become unmanageable.

Onsite power – such as gas turbines – will need to be shared back to the grid.



US power use by data centers, various estimates, compiled by IEA, October 2024



Uptime Intelligence model of NVIDIA based GenAI power use, based on shipment forecasts.

Utilities Need a More Transactive, Open Relationship

Collaboration: Sharing information on likely demand earlier, and in more detail

Load shedding using on-site power: More power companies want data centers to load shed by using local, on-site power sources

IT load shedding by time or location shifting: Utilities want more operators to power down some workloads (time-based) or move the work to an alternative site or region (location-based).

Pay more: Power companies (sometimes mandated) may charge data centers a premium for power.



Prediction 4

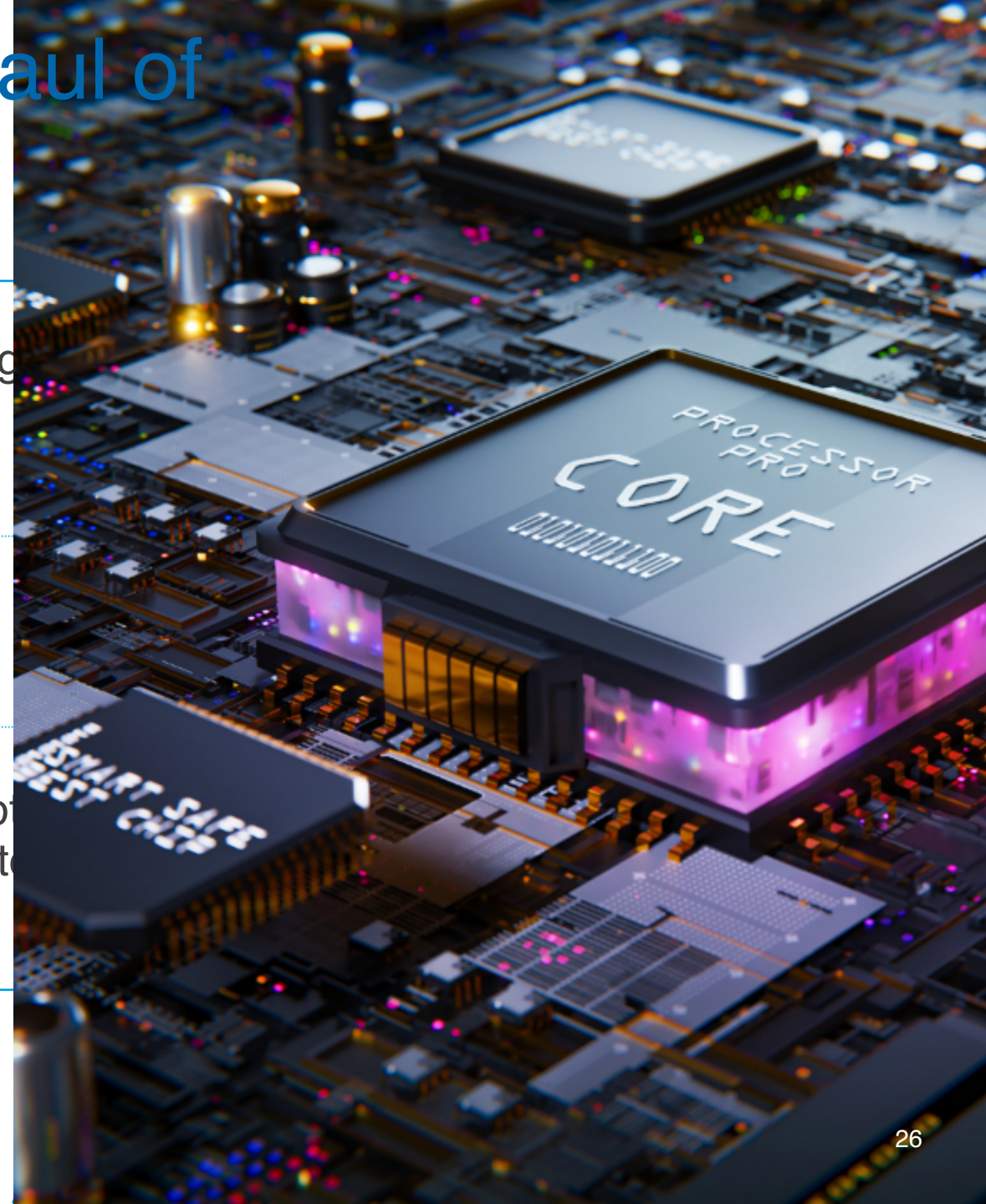
AI to Trigger Radical Overhaul of Data Center Electrification

AI to Trigger Radical Overhaul of Data Center Electrification

AI training hardware is pushing rack densities to heights previously unseen in mainstream facilities. Today, 40 kW AI training compute racks are typical,

As a result, IT space is requiring increasingly larger support infrastructure.

Medium-voltage gear, solid-state transformers and other power delivery innovations will emerge in response to these challenges.



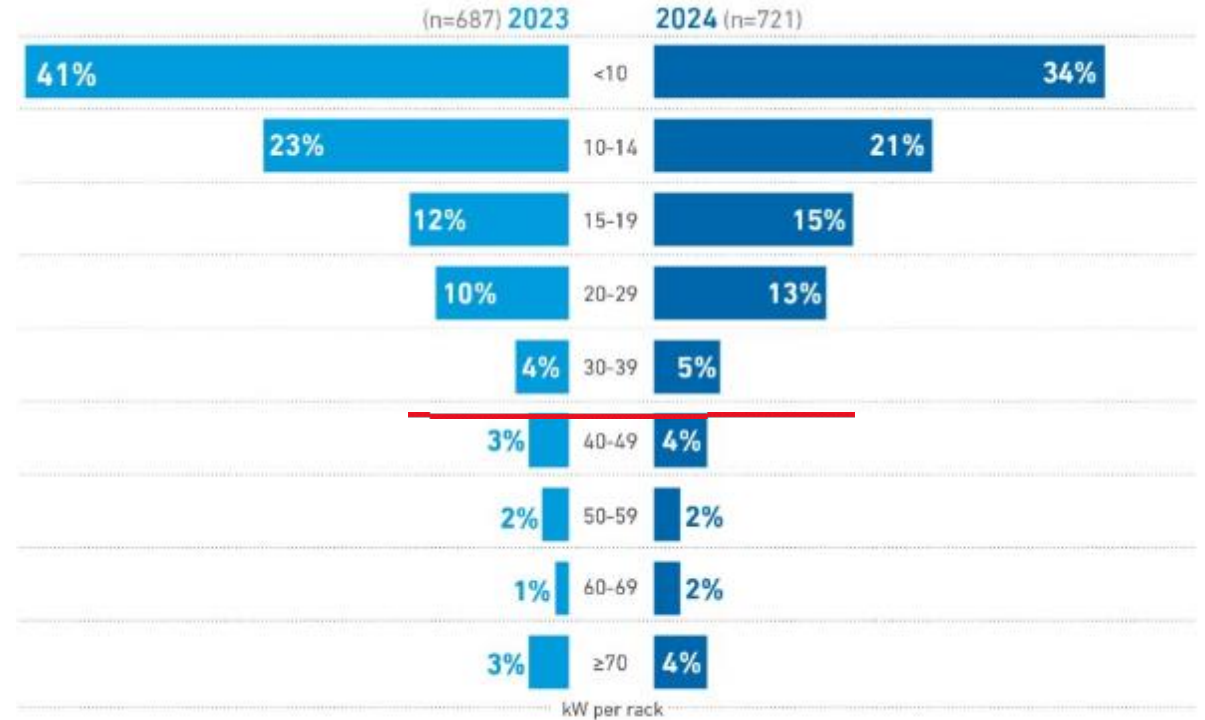
The Crushing Forces of AI

Current generation AI compute racks are around 40 kW per standard 19-inch rack – mostly within standard power distribution equipment.

Rack densities for AI hardware will rise significantly with each new generation of chip upgrades.

Unannounced product roadmaps envision rack-scale systems at and above 300 kW — densities only seen before in supercomputing.

What is the highest server rack density deployed in your data center?



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2024

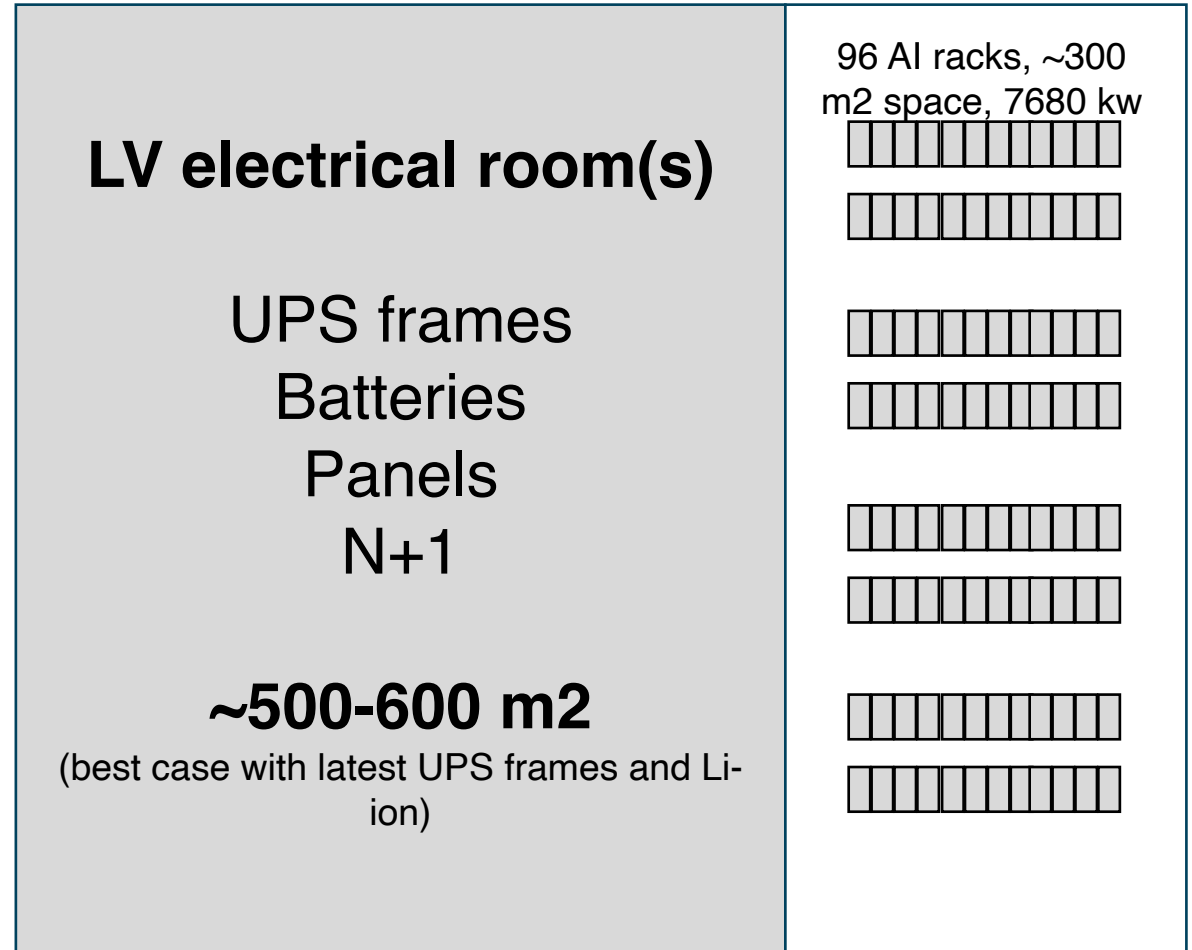
uptime
INTELLIGENCE

Power Distribution Challenges at High-density

A large UPS system for every couple of aisles of AI compute racks today — for every row tomorrow.

Plus, redundancy for concurrent maintainability.

Are data centers becoming large electrical (and mechanical) plants with some IT space?



Power Distribution Challenges at High-density



Number and size of conductors will grow because

- Voltage limitations (400/480/600V) mean that
- Currents go up with power (e.g. 250A busway to 1000A)

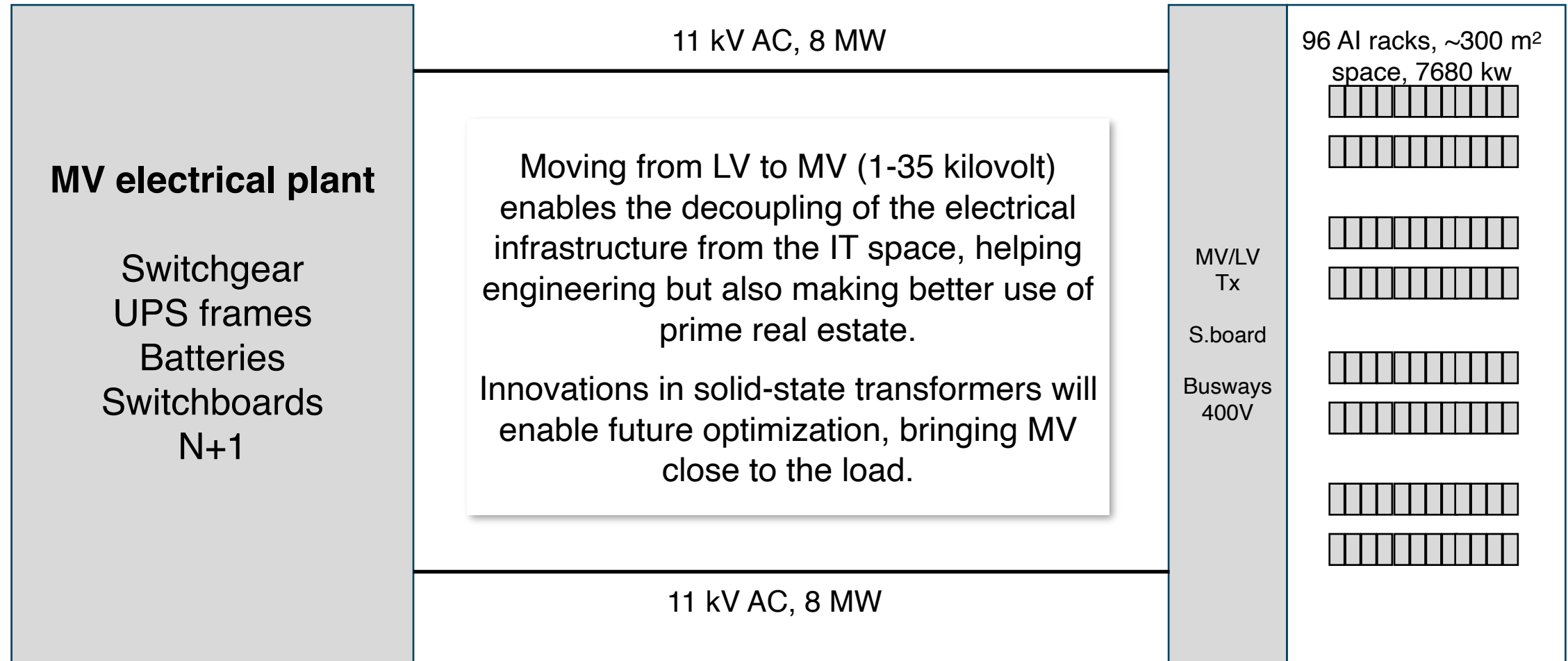
Busway can weigh 55 kg per meter.

Tap-out cables for 10 high-density racks can add 225 kg.

This is without calculating for losses, which makes conductors even larger to compensate.

Run length and weight quickly becomes an issue/added cost.

An Overhaul Of Data Center Power – An Illustration



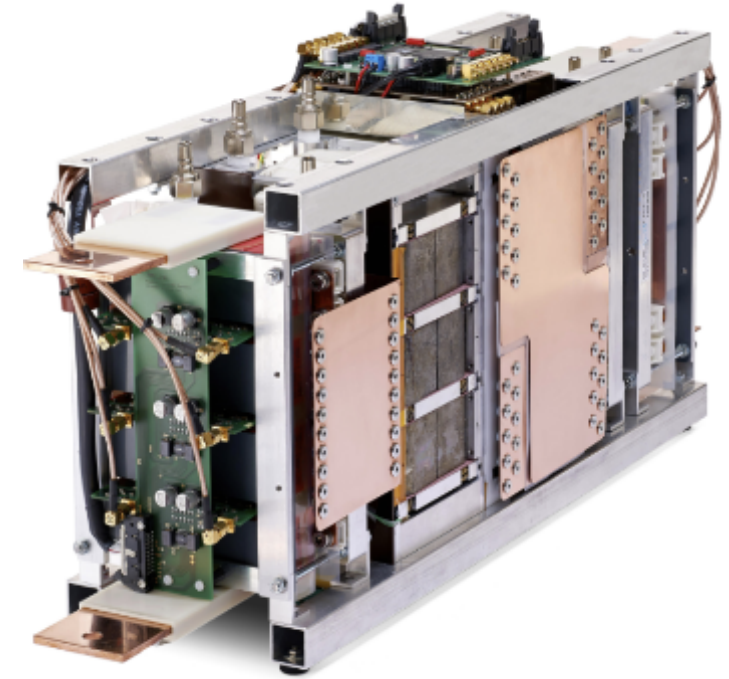
Transforming Data Center Power Systems

Innovations that will help operators optimize power distribution include:

- MV distribution to the IT space
 - Novel IT power distribution topologies
 - Solid state transformers
-

Benefits of switching to MV UPS systems and downstream distribution include:

- Reduces conductor sizes and distribution losses
 - Enables greater distances between UPS rooms and IT load
 - Compresses total electrical plant footprint
-



Credit: EHT Zurich

Prediction 5

NVIDIA's Vision for Data Centers Is Not Without Alternatives

NVIDIA GB200 NVL72

- 42U rack-scale system
- 72 Blackwell GPUs, 36 Grace CPUs
- Power consumption: Up to 120kW
- 5,000+ network cables
- 3,000+ pounds of static weight
- Direct-to-chip liquid cooling

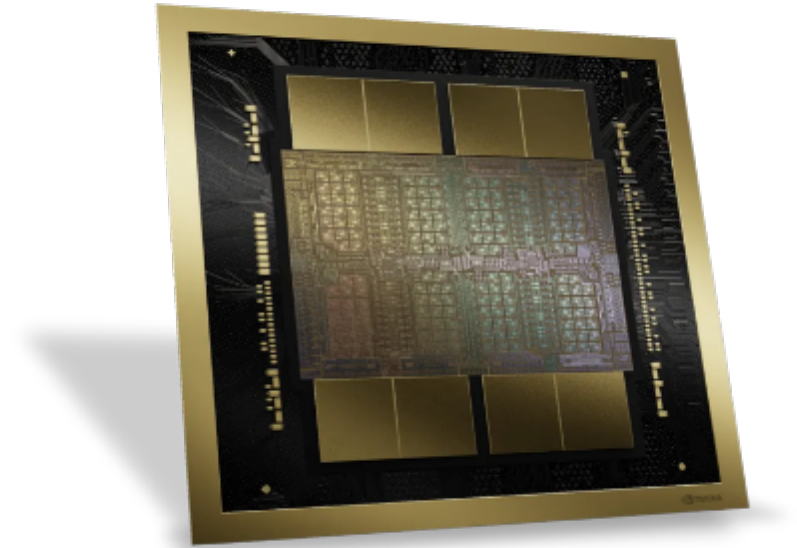


Dense or Distributed?

NVIDIA's vision for the future of AI involves highly performant, power-hungry GPUs in very dense server configurations. These systems are expensive, in low supply, and challenging to deploy.

Alternative approaches to hardware for AI deliver less performance — but do not require a redesign of the data center.

A wider choice of AI development and deployment platforms will benefit accessibility and innovation.



Alternative Hardware Choices for AI

AMD:

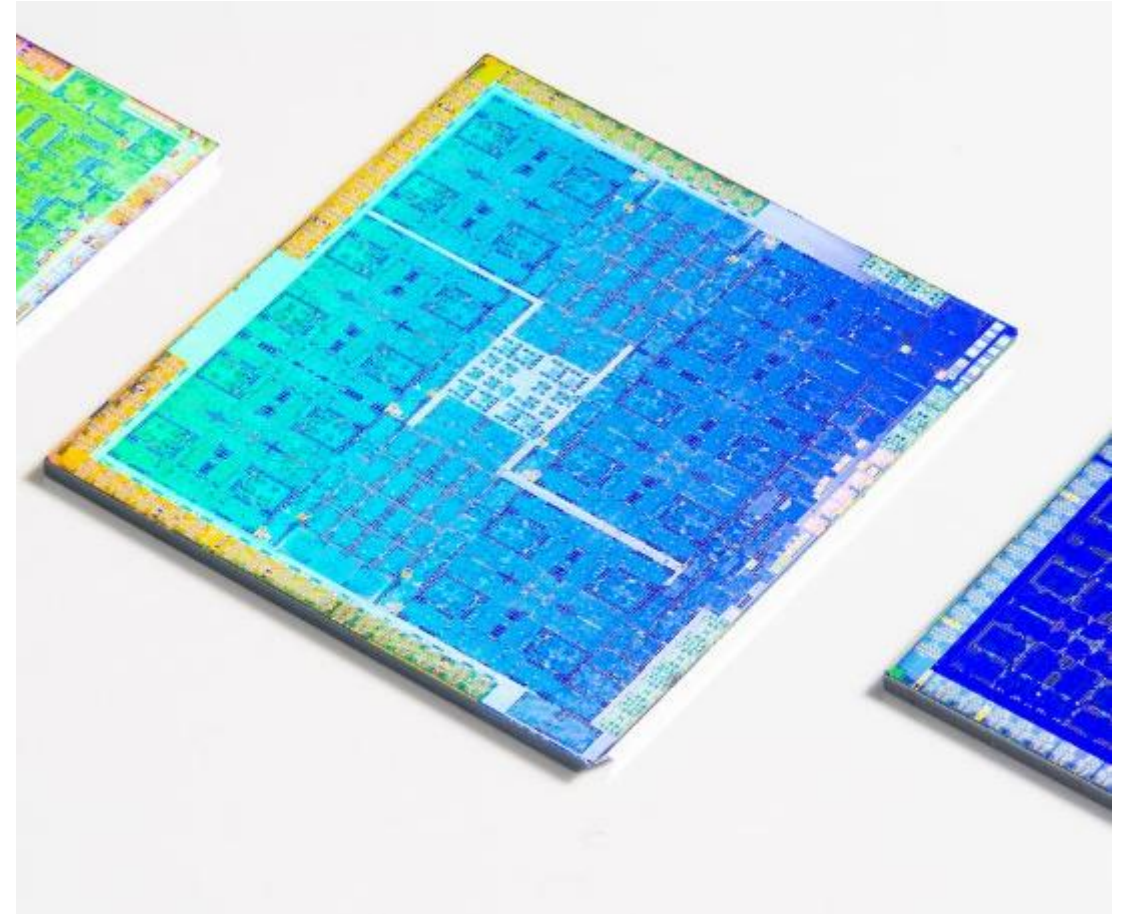
- Instinct MI325X launched in October 2024
- Instinct MI350 expected in 2H 2025

Public cloud providers:

- Trainium2 (AWS)
- Trillium (Google Cloud)
- Maia 100 (Microsoft Azure)

Startups:

Groq, Cerebras, Enflame, SambaNova, Untether, Blaize, FuriosaAI, Recogni, Etched, and many others



NVIDIA versus the world?

A wider choice of hardware platforms for model training and inference will put competitive pressure on NVIDIA and help reduce GPU prices.

It will empower researchers and organizations with limited budgets to build more impressive prototypes and proofs of concept at a lower cost.

Proliferation of alternative hardware platforms will enable many more data centers to run enterprise AI workloads without adaptations in power and cooling delivery.



Visit www.uptimeinstitute.com for more information.



Uptime Institute is the Global Digital Infrastructure Authority.
Uptime Institute is headquartered in New York, NY, with offices in
London, São Paulo, Dubai, Riyadh, and Singapore.

©2025 Uptime Institute, LLC. All Rights Reserved.

Uptime Institute
405 Lexington Avenue
New York, NY 10174

uptime[®]
INSTITUTE