Economic Availability: The Metric the AI Factory Era Requires

"Technical availability tells you the facility is running. Economic Availability tells you whether it is producing."

ABSTRACT

The data center industry has measured availability in nines. These metrics capture whether systems are powered on and accessible. They do not capture whether those systems are converting their power envelope into economic output. As the industry transitions to AI factory infrastructure at revenue densities exceeding $10 million per megawatt per year, the gap between technical availability and economic availability has become the primary determinant of asset performance, return on invested capital, and competitive position.

This paper introduces Economic Availability (EA) as a formal framework for measuring AI factory performance. It documents EA's five loss components using peer-reviewed research, national laboratory data, and production operator results. It demonstrates that a Tier IV facility at 99.995% technical availability may simultaneously operate at 65 to 78% Economic Availability, losing up to $6.6 billion of annual economic output to structural, measurable, and largely invisible coordination failures.

Publication	Synestra Research Series | White Paper No. 1
Date	June 2026
Status	Public — For Distribution to Industry Analysts and Investors
Contact	research@synestra.ai | contact@synestra.ai

SECTION 1

The Limits of Technical Availability

1.1 What Technical Availability Measures

Technical availability is the fraction of calendar time during which a system is operational and accessible. The Uptime Institute's Tier classification system has been the international standard for three decades, defining four progressive Tiers and the redundancy architecture required to achieve each target availability level.

Tier	Classification	Availability	Max Annual Downtime
I	Basic Capacity	99.671%	28.8 hours/year
II	Redundant Components	99.741%	22.0 hours/year
III	Concurrently Maintainable	99.982%	1.6 hours/year
IV	Fault Tolerant	99.995%	26.3 minutes/year

Most hyperscale AI facilities target Tier III or Tier IV equivalent infrastructure. They invest substantially to achieve it. By the technical availability standard, they succeed. Tier IV facilities routinely demonstrate annual downtime measured in minutes. Their SLA compliance rates approach 100%.

And yet, as this paper documents, a Tier IV facility operating at 99.995% technical availability may simultaneously operate at 65 to 78% Economic Availability, losing 22 to 35% of its potential economic output to coordination failures that technical availability was never designed to detect.

1.2 What Technical Availability Does Not Measure

Technical availability was designed for a world where the economic output of a data center was binary: either the system was accessible (producing value) or inaccessible (producing no value). Measuring time-in-service was a sufficient proxy for economic productivity in a general-purpose colocation environment.

In the AI factory era, this binary assumption fails structurally. A data center can be fully operational, all systems green, all alerts nominal, all SLAs met, all Tier IV requirements satisfied, while simultaneously exhibiting these economically consequential failure modes, none of which appear in a technical availability report:

- Operating cooling systems at 40 to 70% of rated load: the steepest, least efficient region of the efficiency curve, consuming disproportionate power overhead relative to the IT load served.
- Running GPU and CPU servers at 50% compute utilization while consuming 65% of peak power, due to the well-documented energy-proportionality gap in modern server hardware (Barroso and Holzle, 2007).
- Leaving 20% or more of provisioned power capacity stranded and unmonetized: paid for in capex and energy infrastructure, allocated on paper, unable to be deployed to productive compute loads.
- Throttling GPU compute performance due to thermal headroom limitations caused by uncoordinated cooling response, where cooling systems react to measured temperatures rather than anticipate workload-induced thermal demand.
- Processing AI training workloads at suboptimal batch sizes that increase total training energy by up to 4x versus optimally scheduled alternatives, on identical hardware, with identical model outcomes (Latif et al., IEEE Access 2025).
- Operating power, cooling, compute, and workload scheduling systems as fully independent domains with no unified optimization surface: a configuration Synestra terms the system zoo, in which no single system is aware of the causal relationships between its decisions and the economic consequences that cascade across other domains.

None of these losses appear in a technical availability report. All of them appear in the economics. The gap between what the monitoring layer sees and what the income statement records is precisely the Economic Availability gap.

1.3 The Revenue Density Transformation

The reason these invisible losses have become the dominant concern of AI infrastructure operators is not that they are new. Versions of each loss category have existed for decades. It is that the revenue density of data center infrastructure has increased by one to two orders of magnitude in the transition to AI factory operation, transforming the economic consequence of each percentage point of efficiency loss from manageable to material.

Infrastructure Era	Revenue Density (Per MW/Year)	Cost of 10% EA Gap (Per MW/Year)
Traditional colocation (2020)	$0.5M to $2.0M	$50K to $200K
Cloud compute (2022 to 2024)	$2.0M to $5.0M	$200K to $500K
AI factory / GPU cloud (2025 to 2026)	$10.0M to $15.0M	$1.0M to $1.5M
Projected AI factory (2027 to 2028)	$15.0M to $20.0M+	$1.5M to $2.0M

The infrastructure did not change. The economics did. A 10% Economic Availability gap on the same physical megawatt costs 5 to 30x more in the AI factory era than it did in the colocation era. At 2.5 GW campus scale, a 10% EA gap costs $3.13 billion per year in unrealized economic output. The technical availability report shows green.

Technical availability was built to answer the question the industry asked in 1995: Is the system on? Economic Availability is built to answer the question the industry must ask in 2026: Is the system producing?

SECTION 2

Introducing Economic Availability

2.1 The Formal Definition

Economic Availability is the ratio of a facility's actual annual economic output to its maximum possible annual economic output given its provisioned physical infrastructure.

EA = Actual Annual Revenue Output / Maximum Possible Annual Revenue Output

Expressed as a percentage. EA = 100% represents full conversion of provisioned power to economic output at design efficiency. EA below 100% represents coordination losses across one or more of the five EA loss categories.

Maximum Possible Revenue Output is the revenue the campus would generate if: all provisioned power were converted to IT load at the target PUE; all IT load were converted to productive compute at full optimal utilization; all workloads were optimally scheduled for their infrastructure demand profiles; and all cooling and power systems operated at their design efficiency points across all load conditions. This benchmark is physically achievable. It requires no new hardware. It requires cross-domain coordination that no current operational architecture provides.

Unlike PUE, which measures facility-level energy efficiency without reference to economic output, EA is denominated in revenue terms, making it directly legible to operators, investors, and landlords simultaneously.

2.2 The Five Components of Economic Availability Loss

EA is eroded by five measurable, independently documented, and research-validated loss categories. Each represents a distinct failure of the operational architecture to convert physical infrastructure into economic output. Each has been studied in isolation by major research institutions. None has been measured in aggregate, because no production system has attempted to coordinate all five simultaneously.

Component 1: Stranded Capacity Loss

Power provisioned at the facility level but unable to be deployed to productive IT load. Industry analysis documents 20% or more of provisioned capacity stranded under normal operating conditions. At AI factory revenue densities of $10 to $15M per MW per year, each stranded megawatt represents $10 to $15M of annual unrealized revenue potential on infrastructure already paid for in capex and ongoing energy costs.

Component 2: Cooling Inefficiency Loss

Cooling infrastructure operating at 40 to 70% rated load runs in the steepest, least efficient segment of the cooling efficiency curve. Chillers, cooling towers, and CRAHs achieve design efficiency only at or near full load. At partial load, efficiency drops nonlinearly, increasing energy consumed per kilowatt of heat removed. The Uptime Institute's 2024 Global Data Center Survey found industry average PUE has remained essentially flat at 1.58 for five consecutive years despite sustained hardware investment: a finding that implicates coordination failure, not hardware inadequacy, as the binding constraint.

Component 3: Compute Utilization Loss

Servers consuming 50 to 65% of peak power while delivering 10 to 50% of peak compute throughput. This energy-proportionality gap was formally quantified by Barroso and Holzle in 2007 and confirmed repeatedly in the modern fleet. Gartner estimates that more than 30% of enterprise and colocation servers are comatose: powered on, drawing power, delivering no productive compute output. In AI factory environments, GPU servers cost $200,000 to $400,000 per unit. Comatose GPUs are an extraordinary concentration of wasted capital.

Component 4: Workload Scheduling Loss

The newest and most consequential EA loss category. Latif et al. at Brookhaven and LBNL (IEEE Access, March 2025) found that changing batch size from 512 to 4,096 images during ResNet training on an 8-GPU NVIDIA H100 HGX node produced a 4x difference in total training energy. A software scheduling decision, requiring no hardware change, produced a fourfold swing in facility power and cooling demand. Chen et al. (2025) extended this across all four AI workload stages: training, inference, fine-tuning, and data preparation, finding that scheduling across stages without infrastructure coordination produces systematic inefficiency across all of them.

Component 5: Coordination Loss

The compounding consequence of operating power, cooling, compute, and workload management as independent domains with no unified optimization surface. SCADA sees power. BMS sees cooling. DCIM sees infrastructure. Workload schedulers see job queues. No system sees the relationships between them. Sakalkar et al. (Google, ASPLOS 2020) demonstrated that software-coordinated power oversubscription of 25% or higher is achievable in production without compromising workload availability: saving hundreds of millions of dollars over multiple years of production deployment. The savings required no new hardware. They required coordination intelligence.

2.3 Economic Availability at Representative Campus Scale

The following applies the EA framework to a 2.5 GW AI factory campus, consistent with the scale of current hyperscaler infrastructure projects.

Parameter	Value	Basis
Provisioned capacity	2,500 MW (gross)	Current hyperscaler AI campus scale
Target PUE	1.30	Best-in-class design target for liquid-cooled AI workloads
IT load at target PUE	1,923 MW	2,500 MW / 1.30
Revenue density	$12.5M per MW/year	Mid-range AI GPU cloud rate, 2025 to 2026
Maximum possible annual revenue	$24.04B/year	1,923 MW x $12.5M, at 100% EA

EA Loss Component	Loss Rate	Annual Economic Impact
1. Stranded Capacity	20% of provisioned capacity undeployable	$4.81B potential foregone
2. Cooling Inefficiency	Actual PUE 1.45 vs target 1.30	$121M additional energy cost annually
3. Compute Utilization	10 ppt below optimal GPU utilization	$2.40B revenue foregone
4. Workload Scheduling	5% energy waste from suboptimal batch/scheduling	$1.20B revenue foregone
5. Coordination Loss	2 to 4% compounding loss from domain isolation	$480M to $960M foregone
Total EA Gap (Estimated)	EA = 72 to 78%	$4.8B to $6.6B/year unrealized
Coordination-Recoverable	Conservative 3 to 5 ppt EA recovery	$2.28B to $3.15B/year recoverable

$6.2M daily economic loss per 2.5 GW campus from recoverable EA gap. Based on conservative $2.28B annual recovery estimate / 365 days. No new hardware. No workload disruption. No additional power required.

2.4 Why EA Is the Right Metric for the AI Factory Era

The data center industry has produced a succession of efficiency metrics: PUE, DCiE, CUE, WUE. Each represented a genuine advance. Each also has a structural limitation: it measures one dimension of facility performance in isolation, without reference to the economic output the facility is designed to produce.

Metric	Introduced	What It Measures	Revenue-Referenced?	Captures Workload Coupling?
Uptime / TA	1990s	System on/off time	No	No
PUE	2007	Facility energy efficiency	No	No
DCiE	2007	Inverse of PUE	No	No
CUE / WUE	2010+	Carbon and water usage	No	No
Economic Availability	2026 (Synestra)	Revenue realized vs revenue possible	Yes	Yes

SECTION 3

The Research Record

3.1 The Evidence Is Settled on the Problem

The five EA loss categories are not theoretical constructs. Each has been independently documented across multiple research traditions. The evidence is substantial, convergent, and not seriously contested.

National Laboratory Research

Lawrence Berkeley National Laboratory's 2024 U.S. Data Center Energy Usage Report (Shehabi et al., LBNL-2001637) is the most authoritative systematic survey of U.S. data center energy characteristics. It documents persistent compute utilization gaps across the commercial fleet, confirms that the server energy-proportionality limitations identified by Barroso and Holzle in 2007 remain present in modern hardware, and projects continued growth in AI workload energy demand at rates sensitive to infrastructure efficiency choices. The report is sponsored by the U.S. Department of Energy.

The Brookhaven National Laboratory and LBNL joint study (Latif et al., IEEE Access 2025) provides the most direct empirical evidence for workload scheduling loss. On an 8-GPU NVIDIA H100 HGX node, a single scheduling parameter change reduced total training energy by a factor of 4 while producing identical model outcomes. The maximum observed power draw of 8.4 kW was 18% below the manufacturer-rated 10.2 kW, confirming that even hardware efficiency assumptions are subject to operational variance with direct economic consequences.

Peer-Reviewed Engineering Literature

Barroso and Holzle's 2007 paper in IEEE Computer established the foundational energy-proportionality framework underlying EA Component 3. Their finding, that servers consume 50 to 65% of peak power at 10 to 50% load, has been repeatedly confirmed in subsequent literature.

Sakalkar et al.'s ASPLOS 2020 paper provides the most rigorous production evidence for EA Component 5 (Coordination Loss). Over several years of production deployment at Google scale, software-coordinated power oversubscription of 25% or higher was sustained across tens of megawatts, saving hundreds of millions of dollars in data center costs while preserving workload availability and performance. The finding is significant because it demonstrates that the coordination opportunity is not marginal. It is architectural.

Chen et al. (arXiv:2509.07218, 2025) extended workload-infrastructure analysis across all four AI workload stages, documenting distinct infrastructure demand profiles for each stage and establishing that AI factory infrastructure faces a multi-stage scheduling coordination problem that is qualitatively more complex than what prior data center management research addressed.

Production Operator Data

Google DeepMind's 2016 production deployment of machine learning-based cooling control demonstrated that software coordination of cooling systems could reduce cooling energy by up to 40% on infrastructure that had already been extensively optimized by skilled engineering teams, producing a 15% facility-level PUE improvement. This result is significant because it demonstrates that the coordination opportunity is additive to, not dependent on, prior hardware efficiency investment. Even optimized facilities have substantial EA headroom that hardware cannot capture.

3.2 The Recovery Evidence Is Also Settled

Each EA loss category has a validated recovery vector. The following table summarizes the five primary recovery mechanisms.

Recovery Vector	Documented Magnitude	Source
Cooling Coordination	Up to 40% cooling energy reduction; 15% PUE improvement	Google DeepMind (2016)
Power Oversubscription	25%+ oversubscription; hundreds of millions of dollars saved	Sakalkar et al., Google ASPLOS 2020
Compute Utilization Recovery	15 to 30% fleet energy from zombie/idle elimination	LBNL 2024; DOE FEMP 2024; Gartner
Design and Construction Efficiency	$250B of $1.7T projected global capex recoverable	McKinsey, August 2025
Workload Scheduling Optimization	Up to 4x energy difference from batch size alone	Latif et al., Brookhaven/LBNL, IEEE Access 2025

The recovery evidence is not theoretical. Each mechanism has been validated in production at significant scale. The distinguishing feature of all five mechanisms is that they require coordination intelligence, not new hardware, not additional power infrastructure, and not workload disruption.

3.3 The Gap in the Literature

What the research record does not contain is a production study that quantifies the economic cost of coordination failure across all five EA loss categories simultaneously, at AI factory scale, in a single unified measurement framework. Individual interventions have been studied with rigor. The compounding effect of full-stack coordination across power, cooling, compute, and workload domains simultaneously has not been studied, because no production system has implemented it.

This gap is not incidental. It exists because the operational architecture required to implement full-stack coordination does not exist as a commercially deployed product category. Economic Availability is the metric. Synestra’s coordination architecture is the instrument. Section 5 quantifies the return.

SECTION 4

The Coordination Architecture

4.1 Why Monitoring Is Insufficient

The data center industry's response to operational complexity has been to invest in monitoring infrastructure: more sensors, more dashboards, more data streams, more visibility. DCIM provides real-time visibility into physical infrastructure capacity, utilization, and power draw. BMS provides granular visibility into temperature, airflow, and cooling system state. SCADA provides real-time power grid and distribution visibility.

The industry is not suffering from insufficient monitoring. It is suffering from insufficient coordination. The distinction is precise:

Monitoring tells operators what is happening. Coordination changes what happens next.

More visibility into independent domains does not close the EA gap, because the gap is not created by ignorance of individual domain states. It is created by the structural inability to translate observations across domains into coordinated optimizing decisions. A DCIM operator who can see that cooling systems are at 60% efficiency and GPU utilization is at 55% cannot, without a cross-domain coordination model, translate those observations into a workload scheduling recommendation that simultaneously improves both metrics.

4.2 The Five Requirements of an EA Platform

01 Cross-Domain Observability

Simultaneous real-time visibility across power, cooling, compute, network, and workload systems, not as independent data streams, but as a unified campus model in which the relationships between domains are continuously maintained. This is the necessary precondition for coordination. It is not coordination itself.

02 Cross-Domain Intelligence

The ability to model causal relationships between workload scheduling decisions and infrastructure outcomes in real time. Scheduling a batch of 4,096-image training jobs in Hall C will increase power draw by X kilowatts, increase cooling demand by Y kilowatts, and interact with the thermal state of the adjacent hall in the following ways. Cross-domain intelligence transforms monitoring data into actionable coordination signals.

03 Campus-Scale Optimization

Optimization across all halls, all zones, all systems, and all workloads simultaneously. Local optimization of individual domains produces suboptimal campus-level outcomes because it cannot account for cross-domain interactions and opportunity costs.

04 Lifecycle Learning

Learning and retention of facility-specific behavioral patterns over time. How this campus responds to thermal load in July. How this tenant's ML training jobs interact with cooling headroom in Hall B. What the actual energy-proportionality curve of this server generation looks like in this thermal environment. Lifecycle learning transforms each day of deployment into a more accurate campus model.

05 Economic Translation

Converting infrastructure metrics into revenue impact in real time, denominated in dollars per day, per hall, per tenant, and per workload stage. Cooling efficiency is 3% below target in Hall C becomes: Hall C is generating $18,000 per day less than its EA potential, recoverable by the following two coordination actions.

4.3 Synestra as an EA Platform

Synestra is designed and built to satisfy all five EA platform requirements through six integrated modules:

Resource Intelligence

Addresses Requirements 01 and 02. Maintains a real-time, cross-domain campus model that integrates data from BMS, SCADA, DCIM, EPMS, workload schedulers, and network monitoring. Continuously models causal relationships between workload decisions and infrastructure outcomes across power, compute, and network domains. Generates optimization recommendations ranked by economic recovery value.

Compute Intelligence

Addresses Requirement 03 from the workload perspective. Provides ML engineering teams with real-time visibility into the infrastructure cost of their scheduling decisions, enabling workload placement and batch sizing decisions that align with facility-level EA targets. Optimizes GPU and CPU utilization to maximize revenue output per deployed megawatt.

Operations Management

Addresses Requirements 01 and 03 across the full facility lifecycle. Establishes an EA baseline during facility commissioning, before the first production workload, by characterizing campus infrastructure behavior under controlled load conditions across all halls and zones. Identifies EA optimization opportunities before revenue operations begin and provides ongoing fault detection, predictive alerting, and workflow orchestration in production.

Lifecycle Analytics

Addresses Requirement 04. A longitudinal learning system that accumulates facility-specific behavioral data over time, continuously refines the campus model, and preserves institutional knowledge against staff turnover, configuration changes, and hardware evolution. Provides predictive maintenance and asset health scoring across the full hardware lifecycle.

Thermal Intelligence

Addresses Requirements 01 and 03 for the thermal domain. Treats the chiller plant, cooling towers, computer room air handlers, and liquid cooling circuits as a single coordinated optimization surface. Anticipates workload-induced thermal demand rather than reacting to measured temperatures, delivering pre-cooling signals and PUE minimization across all halls and zones.

Edge Data Intelligence

Addresses Requirement 05 from the asset owner perspective. Provides OEM-agnostic sensor ingestion and hardware abstraction across all facility infrastructure types. Translates infrastructure performance metrics into asset-level EA scores, hall-by-hall NOI impact analysis, and tenant-level

economic contribution reporting. Makes EA directly legible to investors, lenders, and asset managers.

4.4 Integration Architecture

Synestra does not replace existing operational tools. It integrates with them, reading from the existing sensor and system infrastructure that AI factory operators have already deployed, and returning coordination signals to those systems without requiring new sensor networks, new hardware platforms, or changes to existing operational workflows.

Existing System	Synestra Integration	Data Consumed	Signal Returned
BMS	Read + Write	Temperature, airflow, CRAH state, chiller load	Cooling setpoint recommendations; predictive pre-cooling signals
SCADA (Power)	Read + Write	Hall-level power draw, breaker state, UPS load, generator status	Power oversubscription recommendations; load-balancing signals
DCIM	Read	Rack-level power, server inventory, capacity utilization	Stranded capacity recovery recommendations
EPMS	Read	Sub-metered power at PDU and rack level	Anomaly detection; compute utilization loss identification
Workload Schedulers	Read + Advisory	Job queue, batch size parameters, GPU allocation	Workload placement and batch size advisory signals
Network Monitoring	Read	East-west traffic, InfiniBand fabric utilization	Network-aware workload placement recommendations

SECTION 5

The Economic Case

5.1 The Investment

Investment Component	Amount	Type	Notes
Edge hardware and installation	$20,000,000	One-time capex	Deployed to existing campus infrastructure; no new sensors required
Annual Synestra platform license	$15,000,000	Annual opex	All six modules included
Year 1 Total Investment	$35,000,000	Combined	0.14% of maximum possible annual campus revenue ($24.04B)
Year 2+ Annual Cost	$15,000,000	Annual opex only	Edge hardware is a one-time investment

5.2 The Return

EA recovery estimates are grounded in the validated production results documented in Section 3.2. Conservative recovery assumes 3 percentage points of EA improvement from coordination of cooling, power, and compute domains, consistent with and below the individual-domain recovery rates documented in published research.

Scenario	EA Improvement	Annual Revenue Recovery	Daily Recovery	Payback Period
Conservative	+3.0 ppt	$2,280,000,000	$6,246,575	5.6 days
Base Case	+4.0 ppt	$2,716,000,000	$7,441,096	4.7 days
Optimistic	+5.0 ppt	$3,150,000,000	$8,630,137	4.1 days

5.6 day conservative payback period on full Year 1 investment. Year 2+ annual return: $2.28B on $15M annual license. ROI ratio of 152:1 on ongoing opex.

5.3 The Scale Argument

The economic case for a single 2.5 GW campus is compelling. The systemic argument for EA as an industry-level imperative is more significant still. U.S. data center capacity is projected to grow from approximately 30 GW in 2025 to 90 GW or more by 2030. Global capex commitment over this period exceeds $1.7 trillion. The overwhelming majority is being built for AI workloads at

revenue densities of $10 to $15M per MW per year, by operators whose current operational architectures were designed for a general-purpose compute world that no longer describes their infrastructure.

Scale Scenario	Total AI Factory Capacity	Annual EA Loss (10 ppt)	Annual Loss Per ppt of EA
Current (2025)	~30 GW (US)	$37.5B/year	$3.75B
Near-term (2027)	~60 GW (US)	$75.0B/year	$7.50B
Projected (2030)	~90 GW (US)	$112.5B/year	$11.25B

The market for closing the EA gap is not a software opportunity. It is an infrastructure imperative. The industry is building $1.7 trillion of capacity over five years. It is simultaneously destroying hundreds of billions of dollars of that capacity's economic potential through coordination failures that are measurable, recoverable, and addressable today. Economic Availability is the metric that makes this destruction visible. Synestra is the architecture that stops it.

SECTION 6

Toward an Industry Standard

6.1 The Case for EA as a Reporting Standard

PUE was introduced by The Green Grid in 2007 as a simple, portable metric for comparing data center energy efficiency. It was adopted with remarkable speed. Within five years, PUE had become the universal reporting language for data center efficiency, cited in regulatory filings, investment prospectuses, RFPs, and sustainability reports worldwide. It succeeded because it was simple to calculate, comparable across facilities, and aligned with a metric that operators and owners actually cared about: energy cost.

DCiE followed. CUE and WUE extended the framework to environmental domains. Each metric was valuable. Each is now insufficient for the AI factory era because each is a facility-level metric that cannot capture workload-infrastructure coupling, cross-domain coordination losses, or economic output: the three defining characteristics of the performance challenge AI factory operators face.

Economic Availability is proposed as the next-generation governing metric for AI factory infrastructure: the successor to PUE in the same sense that revenue per available seat mile succeeded load factor as the governing metric for airline performance in the 1990s. Like PUE, EA is simple to define, comparable across facilities, and aligned with the metric that operators, investors, and landlords actually care about: economic output from deployed capital. Unlike PUE, EA is revenue-referenced, workload-sensitive, and inclusive of all five coordination loss categories.

6.2 What EA Reporting Would Require

Reporting Economic Availability requires three operational capabilities that no existing data center management tool combination currently provides in integrated form:

1. Live telemetry integration across power, cooling, and compute domains. The raw data for EA calculation is present in BMS, SCADA, DCIM, and workload scheduler outputs. EA reporting requires these streams to be integrated into a unified campus model updated at sub-minute frequency.
2. A revenue density model calibrated to the facility's specific workload mix. EA is revenue-referenced. Its calculation requires a mapping from infrastructure metrics to economic output at the specific workload mix operated by the facility's tenants. This model must be maintained dynamically as workload mix evolves.

3. A coordination layer capable of attributing economic losses to specific domains. EA reporting is only useful if it is actionable. Operators must be able to identify which of the five EA loss categories is responsible for the gap between actual EA and target EA, and at what magnitude.

Synestra provides all three capabilities in integrated form, enabling operators to report Economic Availability as a real-time operational metric, not a quarterly audit result. The platform produces hall-level, campus-level, and tenant-level EA scores continuously, with attribution to specific loss categories and coordination recommendations ranked by economic recovery value.

6.3 The Research Agenda

Economic Availability as a formal industry metric requires validation through production deployment and peer-reviewed academic study. The individual components of EA loss are documented. What remains to be established through production research is:

- The precise aggregate magnitude of EA loss at GW-scale AI factory campuses across multiple facility types, geographic regions, and workload mixes.
- The compounding economic benefit of implementing all five EA recovery vectors simultaneously through unified campus-scale coordination.
- The rate of EA improvement achievable through machine learning-based coordination over progressive deployment cycles as Lifecycle Analytics accumulates.
- The optimal EA reporting frequency, attribution methodology, and comparability standard for cross-facility benchmarking.
- The relationship between Economic Availability and long-term asset valuation in sale, refinancing, and REIT reporting contexts.

Synestra invites operators, hyperscaler infrastructure teams, academic researchers, and standards bodies including The Green Grid, Uptime Institute, ASHRAE, and DOE national laboratories to participate in establishing Economic Availability as the governing performance metric for AI factory infrastructure. We are actively seeking research partners to conduct the first production-scale EA study: a comprehensive measurement of the full EA gap and the economic impact of full-stack coordination at a GW-scale campus, under peer review, with results published in open-access form.

Interested parties: research@synestra.ai

The data center industry moved from technical availability to PUE because PUE captured something that availability could not: the efficiency of energy conversion. The industry must now move from PUE to Economic Availability because EA captures something that PUE cannot: the efficiency of converting physical infrastructure into economic output. That transition is not optional. It is demanded by the economics of AI factory infrastructure.

REFERENCES

Key Citations

[1] Latif, I. et al. (2025). Single-Node Power Demand During AI Training: Measurements on an 8-GPU NVIDIA H100 System. IEEE Access. DOI: 10.1109/ACCESS.2025.3554728. Also arXiv:2412.08602. Brookhaven National Laboratory and LBNL.

[2] Sakalkar, V. et al. (2020). Data Center Power Oversubscription with a Medium Voltage Power Plane and Priority-Aware Capping. ASPLOS 2020, ACM. DOI: 10.1145/3373376.3378533.

[3] Evans, R. and Gao, J. (2016). DeepMind AI Reduces Google Data Centre Cooling Bill by 40%. Google DeepMind Research Blog, July 20, 2016.

[4] Shehabi, A. et al. (2024). United States Data Center Energy Usage Report. Lawrence Berkeley National Laboratory, LBNL-2001637.

[5] Barroso, L.A. and Holzle, U. (2007). The Case for Energy-Proportional Computing. IEEE Computer, vol. 40, no. 12, pp. 33 to 37. DOI: 10.1109/MC.2007.443.

[6] Chen et al. (2025). Infrastructure Demand Profiles of AI Workload Stages. arXiv:2509.07218. Texas A&M; / Harvard, November 2025.

[7] McKinsey and Company (2025). Building Data Centers Bigger, Faster. McKinsey Capital Excellence Practice, August 2025. Source of $1.7T global capex projection and $250B recoverable opportunity estimate.

[8] Uptime Institute (2024). Global Data Center Survey. Source of industry average PUE 1.58.

[9] NVIDIA (2026). Scaling Token Factory Revenue and AI Efficiency by Maximizing Performance per Watt. NVIDIA Developer Blog, March 2026.

About Synestra: Synestra is developing a campus-scale coordination platform for AI factory infrastructure. Our mission is to close the Economic Availability gap at the GW scale, enabling AI factories to convert their full power envelope into productive intelligence. The platform integrates with existing BMS, SCADA, DCIM, EPMS, and workload schedulers to establish a unified campus model, model causal relationships between workload decisions and infrastructure outcomes, and deliver real-time coordination recommendations that recover economic output without new hardware, additional power, or workload disruption.

Investor and operator inquiries: contact@synestra.ai | Research partnerships: research@synestra.ai | synestra.ai