Emerald AI:

ERCOT Large Flexible Load Task Force discussion

March 28, 2025

emerald ai

Data centers and AI *can* be flexible power consumers

Around the country, utilities and grid operators are seeking flexibility from data centers—and many tech companies would accept limited flexibility and curtailment to achieve faster interconnection

Emerald AI has developed technology to control the computational power demand from data centers running cloud and AI workloads, **delivering acceptable compute service quality while reliably meeting power grid curtailment, emergency response, and demand response needs.**
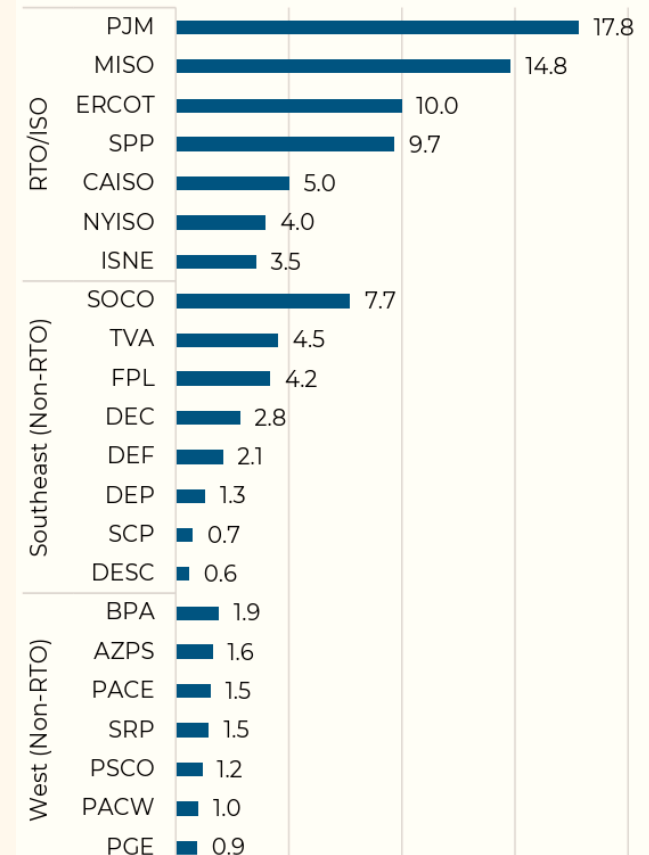
Power utilities see clear value of flexibility—with the right assurances
- If data centers can *reliably, enforceably, and verifiably* reduce load during system peak events, there is 100GW of interconnection headroom across US (Fig. 1)
- Utilities across North Carolina, Arizona, California, Oregon and elsewhere aim to either offer priority interconnection for flexibility or plan to require flexibility by mandate.
- **Emerald AI will execute a demo at a commercial, production-grade data center in Phoenix** with local utilities, tech companies, and EPRI's DCFlex to prove AI flexibility (see Appendix for details on demo design)

Data centers can solve these problems, despite historic resistance
- Major hyperscalers are amenable to curtailing up to 25% for up to 200 hours in return for priority interconnection of 1GW
- Data centers are actually the most controllable large loads in history. Emerald AI's technology provides responsive demand from seconds to hours.

Fig. 1: Capacity headroom with 0.5% Load Flexibility (GW)

	Region	GW
RTO/ISO	PJM	17.8
	MISO	14.8
	ERCOT	10.0
	SPP	9.7
	CAISO	5.0
	NYISO	4.0
	ISNE	3.5
Southeast (Non-RTO)	SOCO	7.7
	TVA	4.5
	FPL	4.2
	DEC	2.8
	DEF	2.1
	DEP	1.3
	SCP	0.7
	DESC	0.6
West (Non-RTO)	BPA	1.9
	AZPS	1.6
	PACE	1.5
	SRP	1.5
	PSCO	1.2
	PACW	1.0
	PGE	0.9

emerald ai

New solutions needed for new ERCOT regulations

As Texas Senate Bill 6 and ERCOT policy advance, new technologies can enable enforceable curtailment of large loads while supporting Texas economic growth and electric system reliability

Lt. Gov. Dan Patrick: Statement on the Unanimous Passage of Senate Bill 6 — Increasing Texas' Electric Grid Reliability

Fort Bend Herald AND TEXAS COASTER

Mar 24, 2025   💬 0

Emerald recognizes that large load flexibility has had mixed historical record. Price-responsive loads can fail to respond when the system needs capacity during emergencies. ERCOT system planners rightfully need to count on firmly available resources to meet grid operational needs and protect reliability.

Senate Bill 6 and large load policy formulation and rulemakings present a critical window of opportunity. As ERCOT develops rules for large load interconnection and performance, AI data centers must prove more than theoretical or price-following capabilities—they must be reliable resources.

Emerald AI seeks to build on its Phoenix demonstration: its next demo will focus on ERCOT enforceable curtailment. To reduce reliance on expensive capital equipment like new backup power infra or risking grid access shut-off, AI data centers must demonstrate that they can comply with SB 6, including:

- **Emergency Curtailment Capability Aligned with SB 6 §39.170:** Show deterministic response to ERCOT-issued curtailment signal and sustain over a defined period.

- **Contract-Enforceable Load Reduction:** Test automated execution in response to grid emergency event to prove to ERCOT a binding mechanism.

emerald ai

For Discussion: Emerald AI seeks guidance on ERCOT needs

After the Phoenix demonstration, Emerald AI will further scale up and bring partners to an ERCOT facility. What functionality would be most useful to demonstrate for the LFLTF?

Context and Objectives

Emerald AI will collaborate with its major strategic partners to execute demo that is useful to ERCOT's needs:

- Simulate ERCOT-curtailed load response
- Prove reliability with auditable results
- Show full telemetry path & duration compliance

Emerald AI Partnerships: Energy and Tech

Key Questions to ERCOT about Emerald AI Demo Design

1. **What data and telemetry will enable the most compelling demo?** For example, Emerald AI might:
   - Use real, historical EEA events as synthetic signals to test load curtailment performance
   - Demonstrate telemetry traceable to ERCOT or utility control centers, in direct alignment with the upcoming mandatory load-shed coordination requirements.

2. **Which energy-sector entities are ideal partners?**
   - For this demo, Emerald AI will bring data center, cloud, and AI partners.
   - Are there utility and TSP partners that have expressed interest in innovative demonstrations to ERCOT?

3. **How can this demo best demonstrate SB 6 compliance?** In addition to its core technology of controlling AI workloads and power demand, Emerald AI seeks to partner to demonstrate compliance tooling, such as:
   - Interconnection milestone management.
   - Readiness dashboard usable by TSPs.

emerald ai

Contact Us:

Varun Sivaram, *Founder & CEO:* varun@emeraldai.co

Ayse Coskun, *Chief Scientist:* ayse.coskun@emeraldai.co

Jack Megrue, *Business Operations Associate:* jack.megrue@emeraldai.co

emerald ai

Appendix:

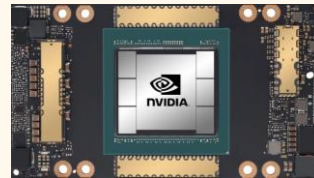# Emerald AI technical results and Arizona demo

Emerald AI's first commercial demo is in Phoenix (April '25)

Emerald AI, along with tech and energy partners, will modulate AI workloads in a commercial data center (production-grade, real cloud zone facility) to exactly match utility demand response goal

Commercial Phoenix Data Center
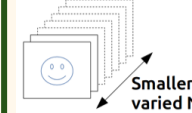
Live AI GPU Cluster

Several hundred NVIDIA GPUs running AI workloads
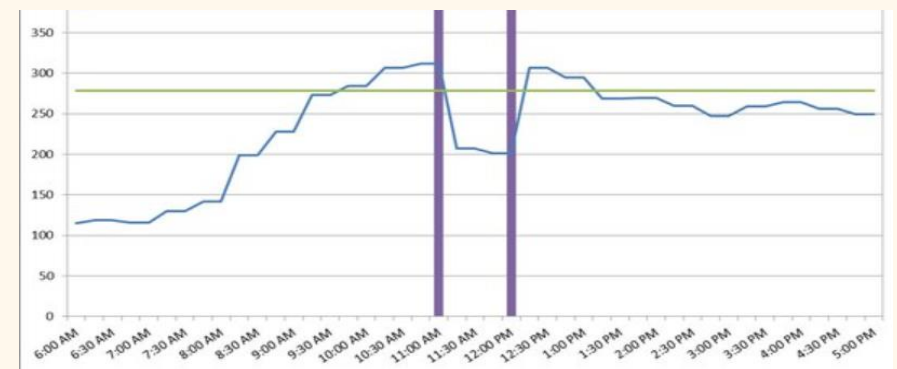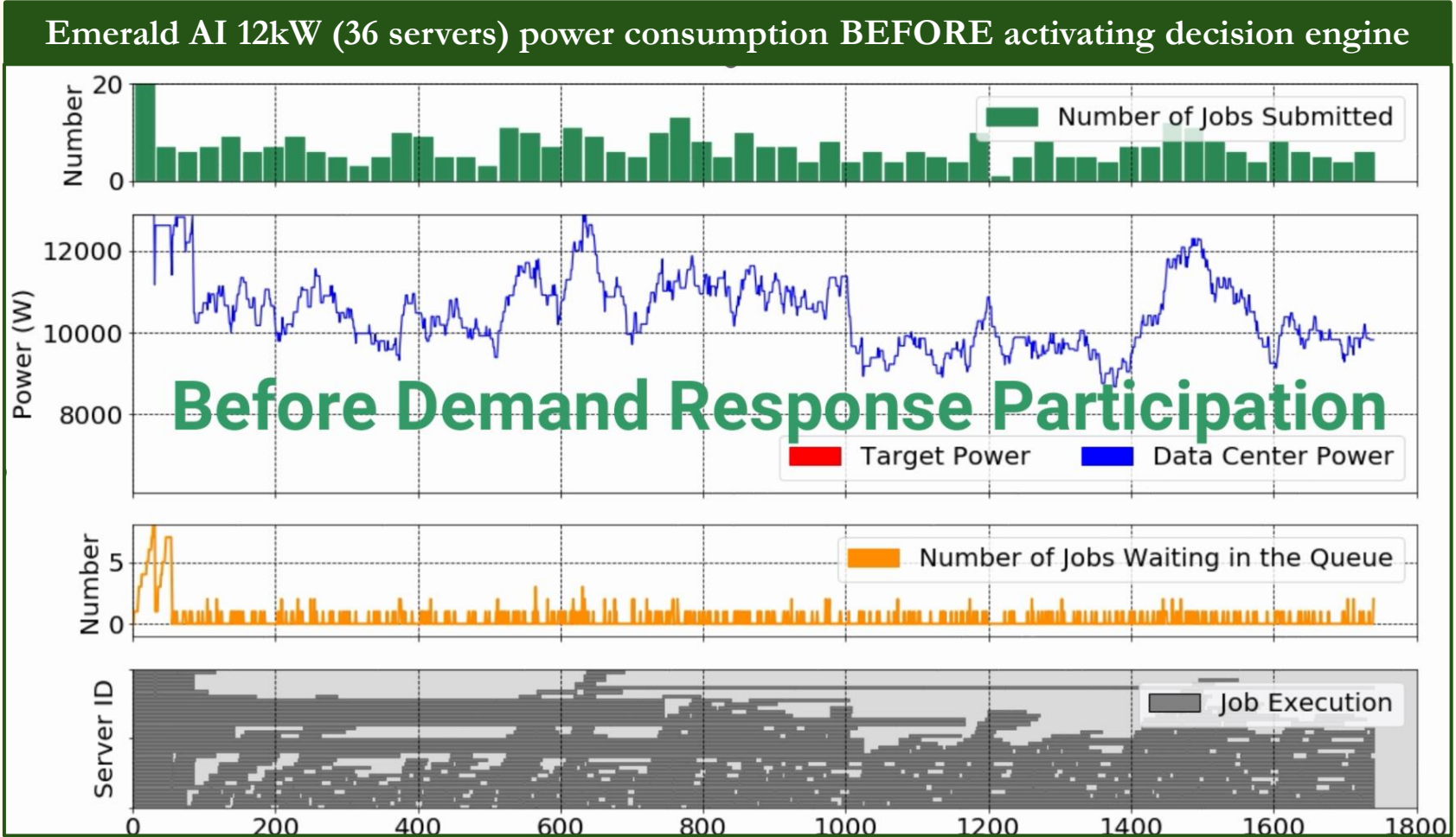
Emerald AI will orchestrate AI workloads...

Ensure acceptable AI user compute Quality of Service (QoS)
- Runtime
- Throughput

...and meet local utility demand response goals

emerald ai

Emerald AI prototype proved that real-world servers…
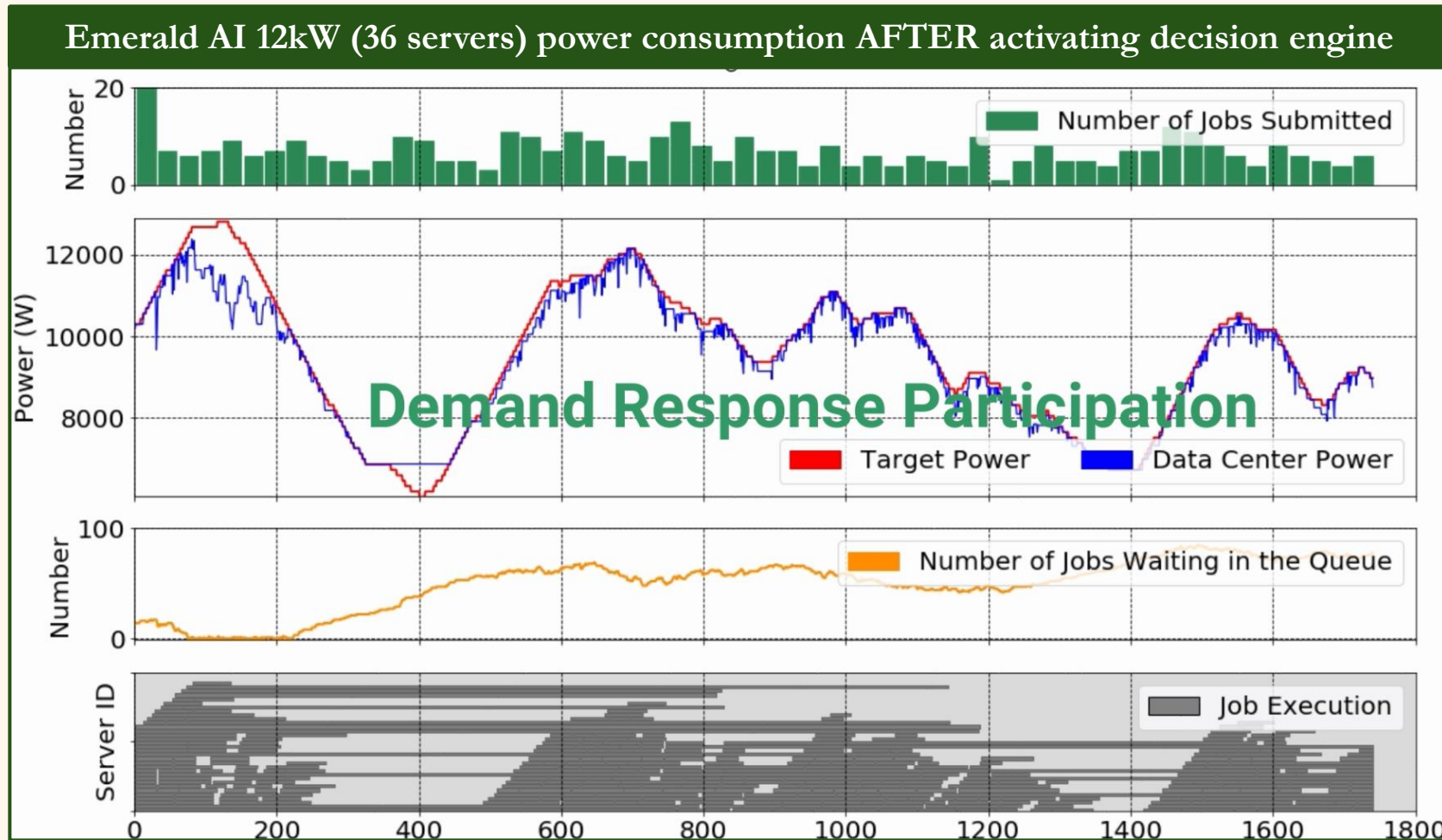

Emerald AI 12kW (36 servers) power consumption BEFORE activating decision engine

…can operate flexibly in response to PJM ISO 4-second regulation reserve signal *and meet user compute quality of service requirements*

Emerald AI 12kW (36 servers) power consumption AFTER activating decision engine

Source: Coskun et al., 2024

![emeraldai logo]

Powering the AI Revolution

Intelligence to control the interface between the world's two most valuable infrastructure networks: power grids and data centers

Emerald AI is building a software solution to enable AI data centers to flexibly adjust their power consumption from the electricity grid on demand. By orchestrating the computing loads of training, fine-tuning, and querying AI models across a network of data centers, Emerald's platform aims to give rise to "AI Virtual Power Plants" that transform data centers from liabilities to assets on the power grid. As a result, utilities and system operators can more rapidly interconnect new data centers to the power grid and compensate data centers for their flexibility, AI compute providers can democratize access to data center capacity at more attractive rates, and surging AI power demand can be met by existing and new renewable energy sources that are the cheapest and fastest to build but require flexibility in their use.

Leadership

Dr. Varun Sivaram
Founder & CEO

One of the world's foremost authorities on energy technologies, Dr. Sivaram was formerly the Chief Strategy and Innovation Officer at Orsted A/S (NASDAQ: DOGEF), a $20 billion Fortune 500 Global clean energy firm, and previously served as Chief Technology Officer of India's largest clean energy company, ReNew Power (NASDAQ: RNW). He served in the White House under Special Presidential Envoy for Climate John Kerry as Managing Director for Clean Energy and created the First Movers Coalition. Dr. Sivaram is a senior fellow for energy and climate at the Council on Foreign Relations and a senior advisor at the Boston Consulting Group, Aventurine Partners, and Everview Partners. He is the author of three books, including the bestselling "Taming the Sun" on the future of solar energy, and holds a Ph.D. in condensed matter physics from Oxford University, where he was a Rhodes Scholar, and undergraduate degrees from Stanford University.

Prof. Ayse Coskun
Chief Scientist

Professor Coskun is full professor in the Electrical and Computer Engineering Department at Boston University, the Associate Dean of Research at the College of Engineering, and the Director of the Center for Information and Systems Engineering. Her research group has published path-breaking research on orchestrating data centers to provide electric demand response and applying AI and machine learning to enhance cloud and high-performance computing. She received the Ernest Kuh Award for contributions to energy-efficient system-level design and an IBM Faculty Award for applying AI-based methods for DevSecOps. Earlier, Prof. Coskun worked in industry at Sun Microsystems (now Oracle). She serves on the Steering Committee of the Hariri Institute of Computing and as Deputy Editor-in-Chief of the IEEE Transactions on CAD. She holds a Ph.D. in Computer Science & Engineering from UCSD.