NVIDIA

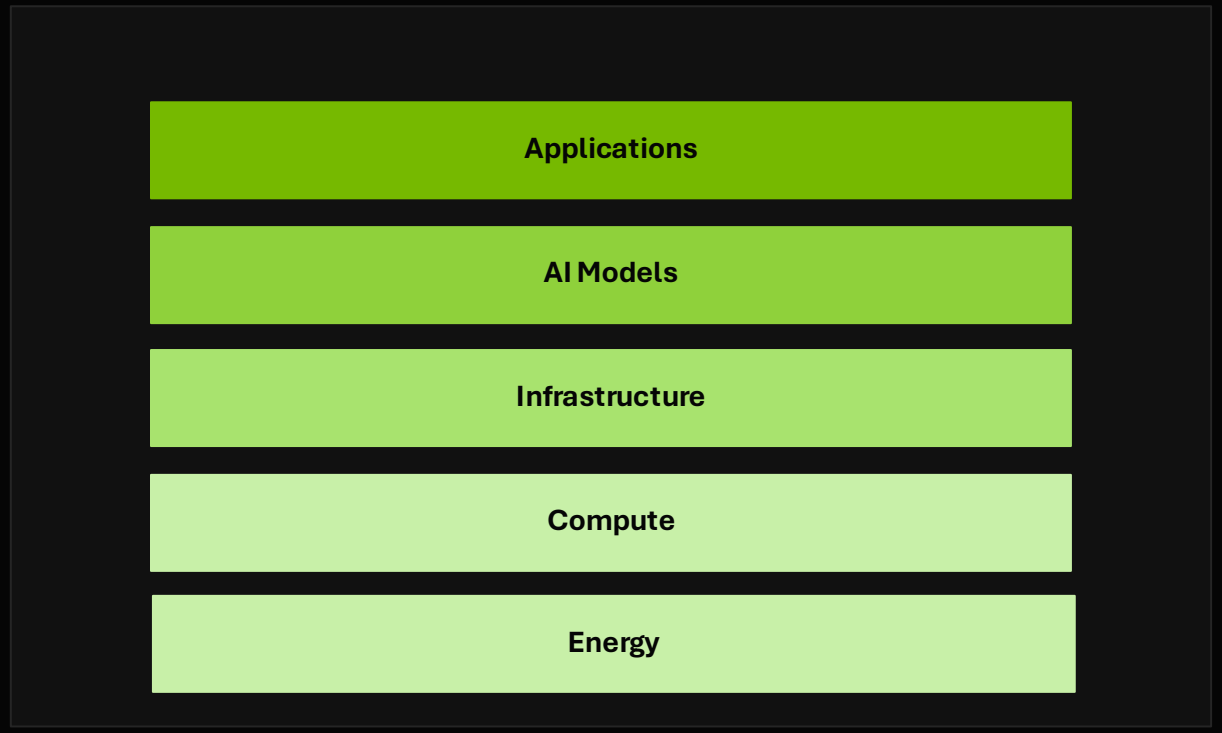
Increasing Compute Density, Agentic and Physical AI

Flexible AI factories for resiliency, affordability, and speed to power

ERCOT Board

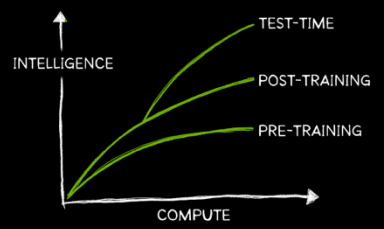
AI is Five Layer Cake

Energy is no more a commodity- it is the foundation of the AI technology stack.



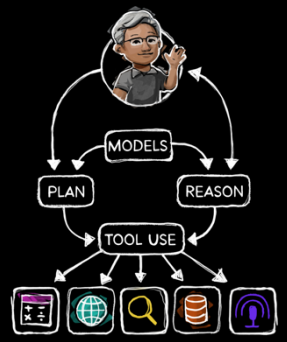
AI Scales Beyond LLMs

Growth in Reasoning model, and Physical AI supercharging Inferencing



3 SCALING LAWS

COMPUTE IS DATA



AI BECOMES AGENTIC

PHYSICAL AI TAKES LEAP

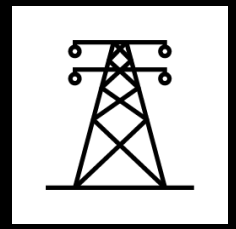
AI LEARNS LAWS OF NATURE

OPEN MODELS REACH FRONTIER

AI Growth Is Not a Bubble
ChatGPT to Claude Code (AI Agents)- 10,000X Inference Compute

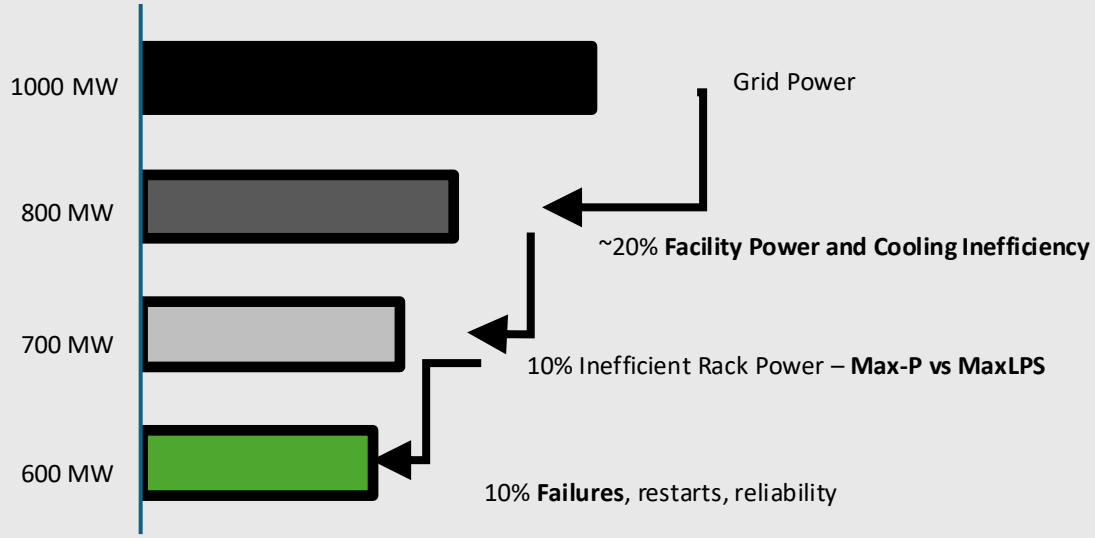
Challenge is to maximize AI tokens per watt of available energy

Facility, rack, downtime inefficiencies cause only a fraction of grid power to contribute to AI output



Power in

1 GW AI Factory



60-70% for AI Output

- Stranded Power
- Fixed and Static, Conservative Power / Cooling Configurations
- Overprovisioned for Failures and Extremes

NVIDIA DSX – Open Reference Architecture for GW AI Factory

Bringing over 100 Ecosystem Partners adopt standard design to maximize efficiency

Reference Designs, Best Practices, Methodologies

Rubin DSX Reference Designs

DSX Ecosystem

Omniverse DSX Blueprint

DSX Libraries, APIs, Software

DSX Max-Q

DSX Flex

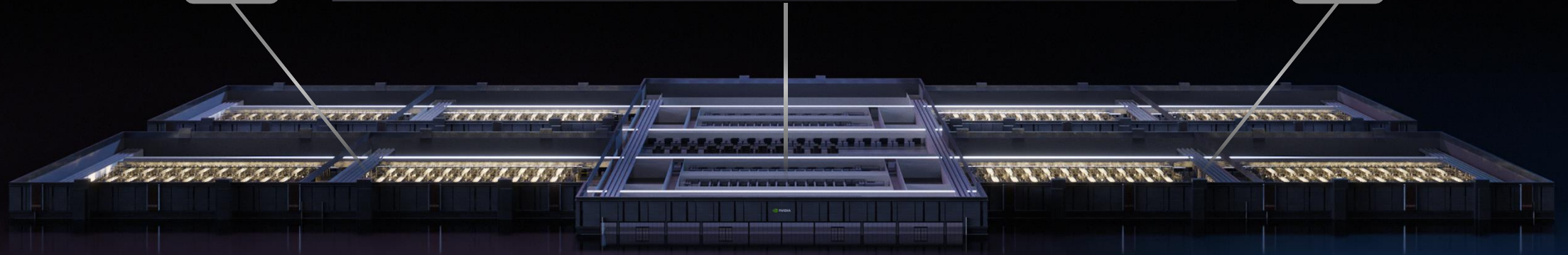
DSX Exchange

DSX Sim

Chips, Systems, Facilities

Power

Liquid Cooling



NVIDIA DSX AI Factory Platform

Accelerates Scalable, Energy-Efficient AI Factory Deployment

DSX Flex

CoreWeave · Crusoe · Digital Realty · emeraldai · NEBIUS · NSCALE · switch

DSX Max-Q

CoreWeave · Crusoe · Digital Realty · emeraldai · NEBIUS · NSCALE · Red Hat · run:ai · phaidra · switch

DSX Sim

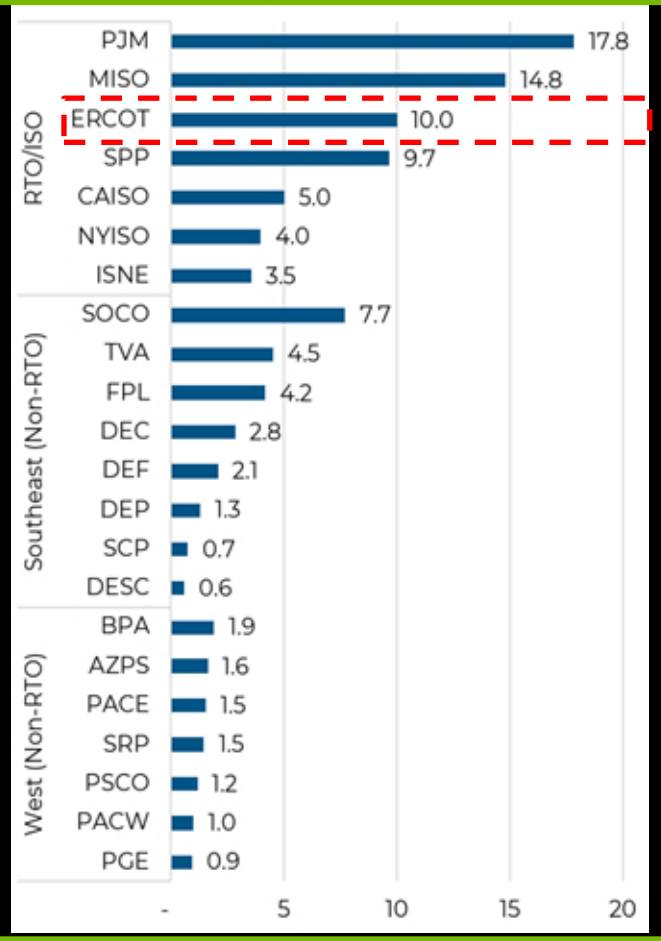
ABB · ARMADA · AVEVA · aviz · BECHTEL · networks · cadence · CHECK POINT · CLOUDIAN · Dassault Systèmes · DELL Technologies · EATON · etap · GE VERNOVA · hedgehog · HITACHI · Jacobs · KEYSIGHT · metalsoft · MIRANTIS · MITSUBISHI ELECTRIC · qetris · NOKIA · Open Nebula · paloalto · ptc · RAFAY · Red Hat · run:ai · Schneider Electric · SIEMENS · SIEMENS energy · slurm · Spectro Cloud · TRANE TECHNOLOGIES · TrendAI · VAST · vCluster · VERTIV · WEKA

DSX Exchange

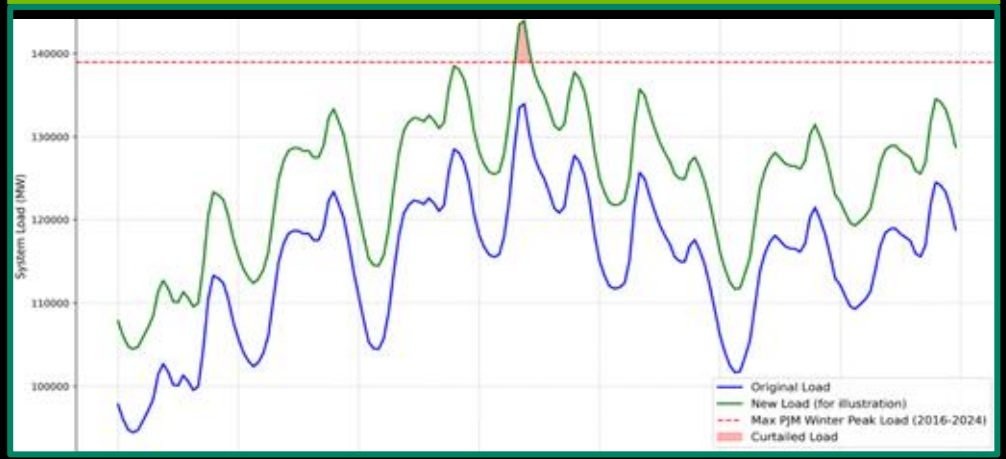
ABB · ARMADA · AVEVA · aviz · networks · cadence · CHECK POINT · CLOUDIAN · CoreWeave · Crusoe · Digital Realty · DELL Technologies · EATON · emeraldai · GE VERNOVA · hedgehog · HITACHI · KEYSIGHT · metalsoft · MIRANTIS · MITSUBISHI ELECTRIC · NEBIUS · qetris · NOKIA · NSCALE · Open Nebula · paloalto · phaidra · RAFAY · Red Hat · run:ai · Schneider Electric · SIEMENS energy · slurm · Spectro Cloud · switch · TRANE TECHNOLOGIES · TrendAI · VAST · vCluster · VERTIV · WEKA

Flexible AI Factories can unlock 10GW or more in ERCOT alone to power the AI revolution, protect affordability, and best utilize the existing power system

New capacity with 0.5% Load Flexibility (GW)



How flexibility enables new loads to utilize existing power system



First Mover Technology Companies (Duke, 2025)

Category	Examples
Operational flexibility	• Google deployed a "carbon-aware" temporal workload–shifting algorithm and is now seeking to develop geographic distribution capabilities (Radovanović 2020). • Google data centers have participated in demand response by reducing non-urgent compute tasks during grid stress events in Oregon, Nebraska, the US Southeast, Europe, and Taiwan (Mehra and Hasegawa 2023).
	• Startup companies like Emerald AI are developing software to enable large-scale demand response from data centers through recent advances in computational resource management to precisely deliver grid services while preserving acceptable quality of service for compute users

Source: Norris et al., Duke University, 2025

emeraldai NVIDIA

EPRI DCFlex- Grid-integrated Flexible Data Centers

Developers

COMPASS datacenters

QTS

Hyperscalers

Google

Meta

Microsoft

IPP's

CALPINE

Constellation.

nrg

RWE

VISTRA

ISO/RTO

California ISO

CENACE

CENTRO NACIONAL DE CONTROL DE ENERGÍA

ercot

ieso

smart KPX

MISO

NESO National Energy System Operator

pjm

SPP Southwest Power Pool

Technology Providers

CATERPILLAR

JENBACHER

INNIO

KRAKEN

NVIDIA

PADO

Schneider Electric

Advisory & Finance

EY

ING

KPMG

Engineering & Construction

CRIEPI Central Research Institute of Electric Power Industry

QUANTA

Ulteig

wesco

Utilities

aes

Alliant Energy

aps

Bonneville POWER ADMINISTRATION

AEH

Dominion Energy

DTE

DUKE ENERGY.

EAST KENTUCKY POWER COOPERATIVE

ENOWA. NEOM

entergy

exelon

Georgia Transmission

KEPCO

LES

LCRA

nationalgrid

Nebraska Public Power District

NEW YORK STATE OF OPPORTUNITY. NY Power Authority

OglethorpePower

PGE

ppl

PSEG

REC RayburnElectric

Rte

SRP

Southern Company

TVA TENNESSEE VALLEY AUTHORITY

UNITED POWER Your Touchstone Energy® Cooperative

DCFlex- NVIDIA and Emerald AI have completed a series of successful demos and deployments ahead of a 96 MW commercial deployment in Virginia

Demo Schedule

Phoenix Demo 1 (complete)

256 A100 GPUs

Proof of Concept: Reduce load by 40% while maintaining SLAs

Chicago Demo 2 (complete)

80 H100 GPUs

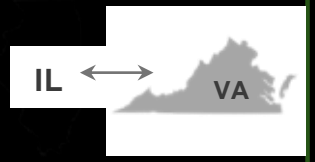
Enhanced realism (e.g., scheduler integrations, unknown workloads)

UK National Grid Demo 3 (complete)

96 B300 GPUs

Responded to real-time direct utility dispatch

AI Geo-shifting Demo 4 (complete)



160 H100 GPUs

Intelligently orchestrate workloads across geos

Oregon Demo 5 (complete)

144 GB300 GPUs

Real time direct utility dispatch with real workloads

96 MW Flexible AI Factory (H2 2026)

~25K Vera Rubin GPUs

First Power Flexible AI Factory

Each subsequent Demo has demonstrated our growing capabilities under more challenging circumstances

NVIDIA and Emerald AI have demonstrated AI Factory flexibility globally, including by relieving grid strain at halftime of a football match

Demonstration Goal

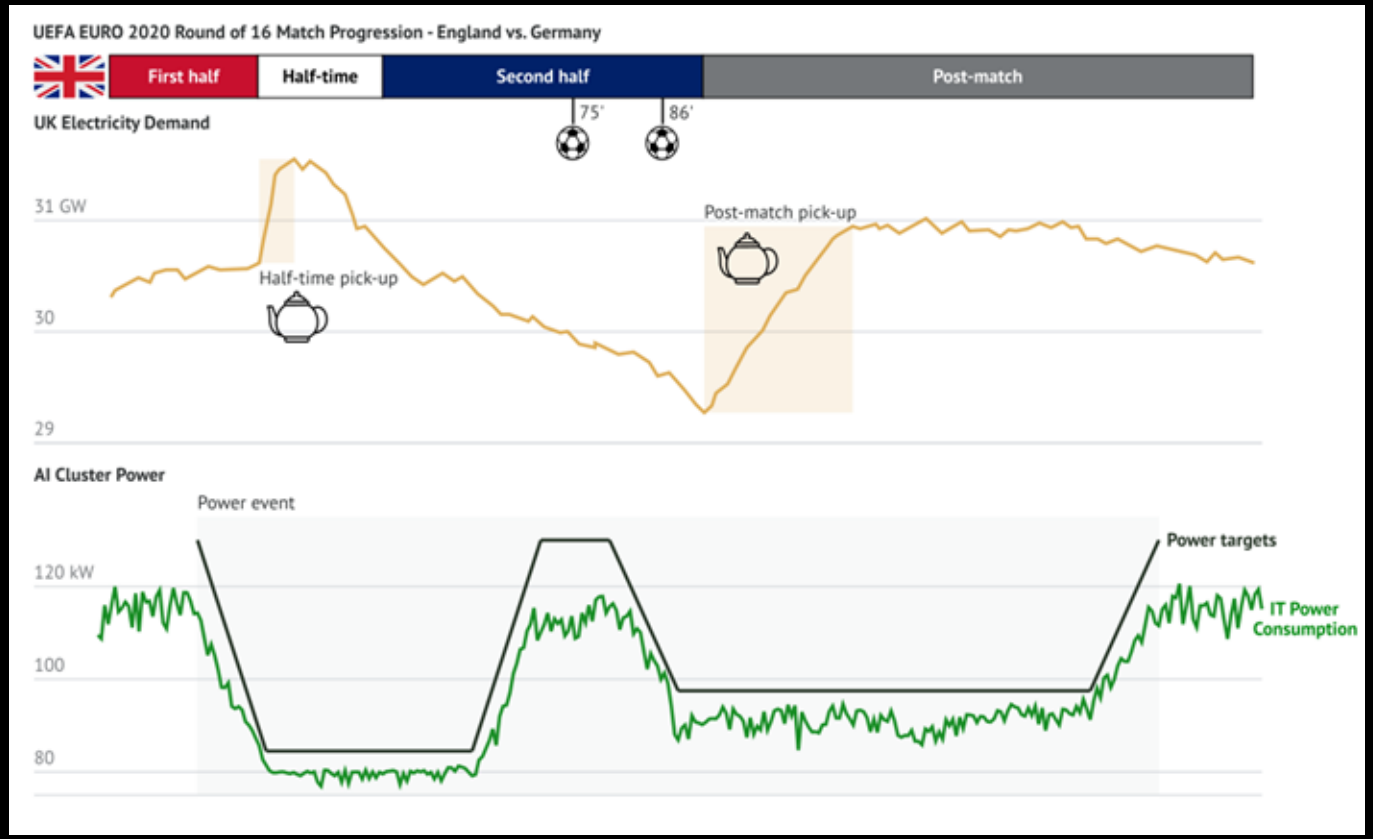
Demonstrate that an AI cluster can meet a designated power reduction in response to high periods of electricity use during major TV breaks in the UK

Results

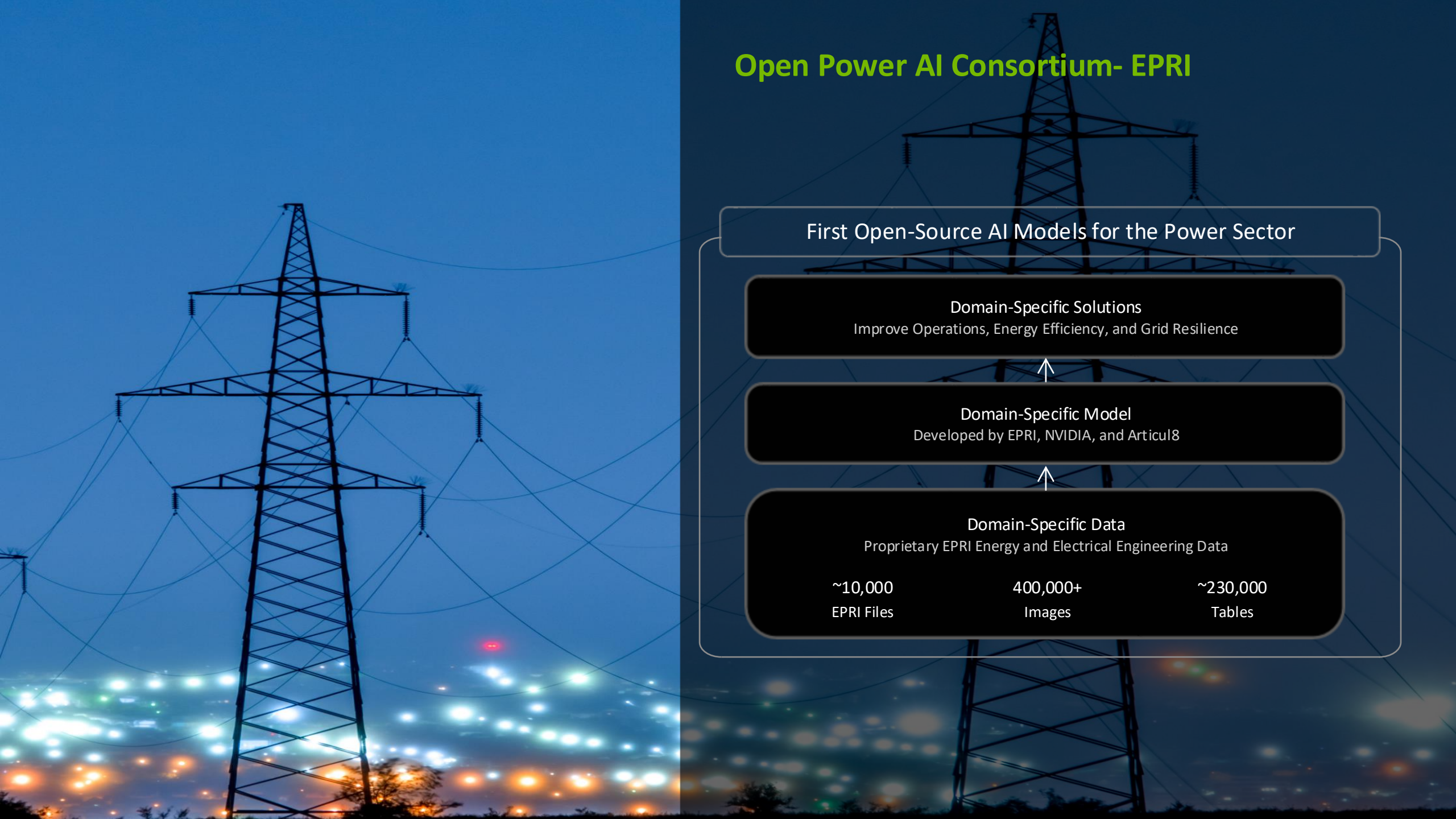
Successfully reduced cluster power by **35% for 30 minutes surrounding halftime** and for **25% for one hour at the end** of a football match, helping combat the '**TV Pickup**' phenomenon as viewers fire up electric tea kettles during breaks in TV viewing

Helped ease burden from a National Grid ESO-wide spike of **1 GW at half-time** and **1.6 GW at full-time**

Cluster Power Trace during soccer match (35% and 25% Reductions)



Open Power AI Consortium- EPRI



Texas leading the AI buildout

Building Grid Friendly Flexible AI Factories/Data Centers

Supercharge AI Growth

Reasoning Models, Agents, and Physical AI

Grid Resiliency

Flexible AI Factory, providing Peak load management

Affordability

Increase utilization of existing grid assets