

The Energy Cost of Artificial Intelligence

-Modeling AI-Driven Data Center Electricity Demand with LLM Predictions

Danbo Chen • The Ohio State University • AAEA conference 2026
Zijun Zhou • Meta Platform

Background & Motivation

- Generative AI is driving unprecedented growth in global computational demand
- AI GPU racks consume up to 6× more power than conventional server racks
- Energy demand scales super-linearly with model size; deployment is globally uneven
- Conventional forecasts fail to capture spatial concentration and firm-level siting strategies
- The compute–energy nexus is poorly quantified in current energy-planning frameworks

118 TWh 6-firm baseline (2024)	239–295 TWh by 2030 (modeled)	13–17 % CAGR (2024–2030)
-----------------------------------	----------------------------------	-----------------------------

Research Questions

- Where will AI data centers be built, and what drives siting decisions?
- How fast will their electricity demand grow from 2025 to 2030?
- Which regions face grid stress from concentrated AI loads?

Data & Inputs

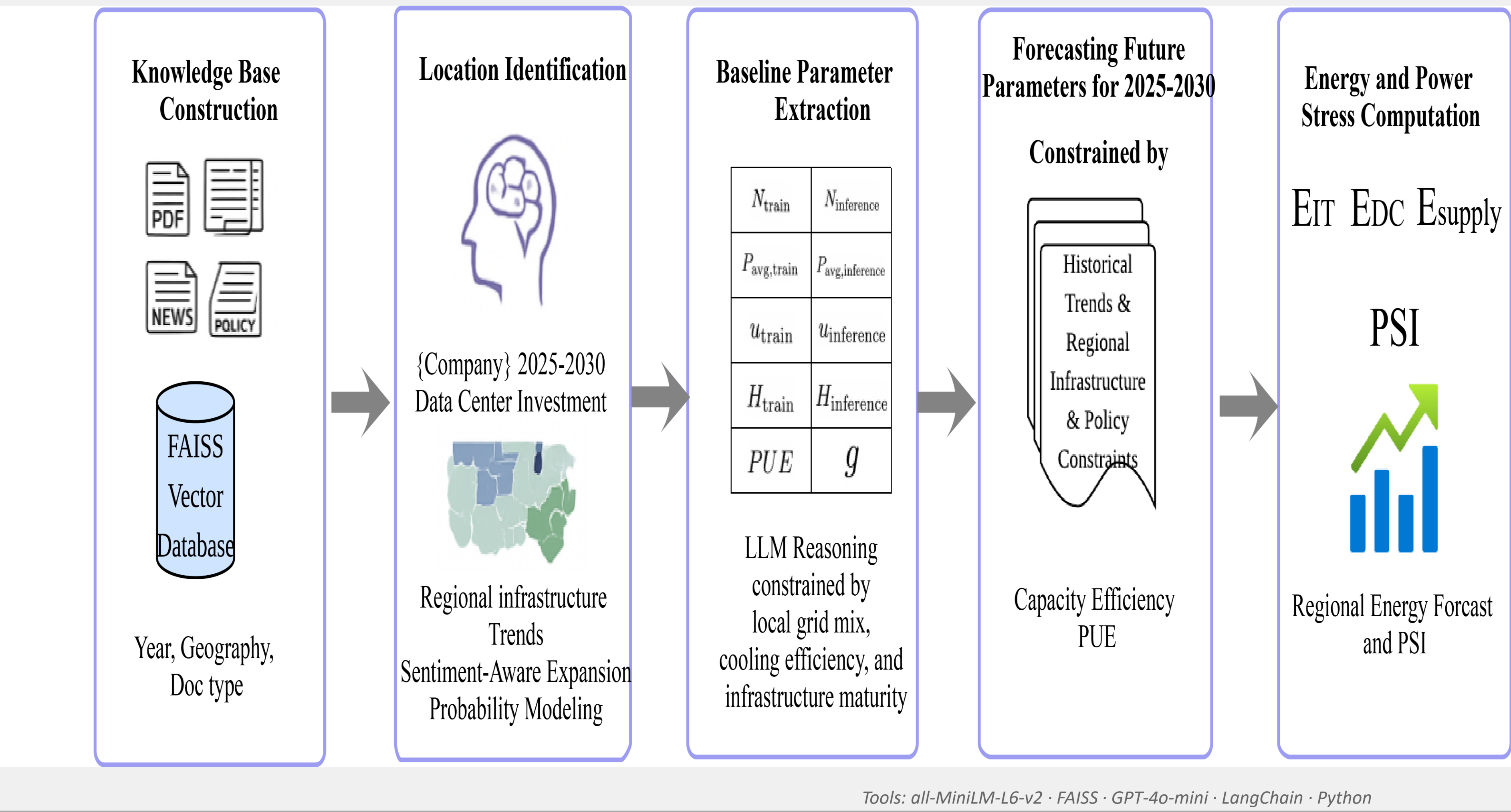
- Corporate disclosures: SEC EDGAR filings, ESG reports, earnings calls
- Infrastructure registry: S&P Capital IQ — facility counts 2015–2024
- Energy data: IEA, U.S. EIA (state-level), World Bank, OECD Energy Statistics
- Geospatial: Landsat 8 / Sentinel-2 via Google Earth Engine, OpenStreetMap
- Policy documents: U.S. DOE, European Commission DG-ENER, national grid operators

Firms covered: Amazon • Microsoft • Google • Meta • Oracle • Apple

Result 1 — Global Siting Patterns

- >90 % of AI compute concentrated in North America, Western Europe, and Asia–Pacific
- Siting driven by renewables, policy corridors, and connectivity — not user proximity
- Five strategic archetypes identified:
 - Renewable-anchored — Google, Meta
 - Policy-responsive — Microsoft, Oracle
 - Market-expanding — Amazon
 - Vertically integrated — Apple
- Top hubs: Virginia, Texas, Oregon, Iowa (US) · Netherlands, Nordic (EU) · Singapore, Taiwan, Japan (APAC)

Framework: AI–Energy Coupling (RAG + LLM Pipeline)



Result 1 — Global AI Data-Center Distribution & Siting Determinants

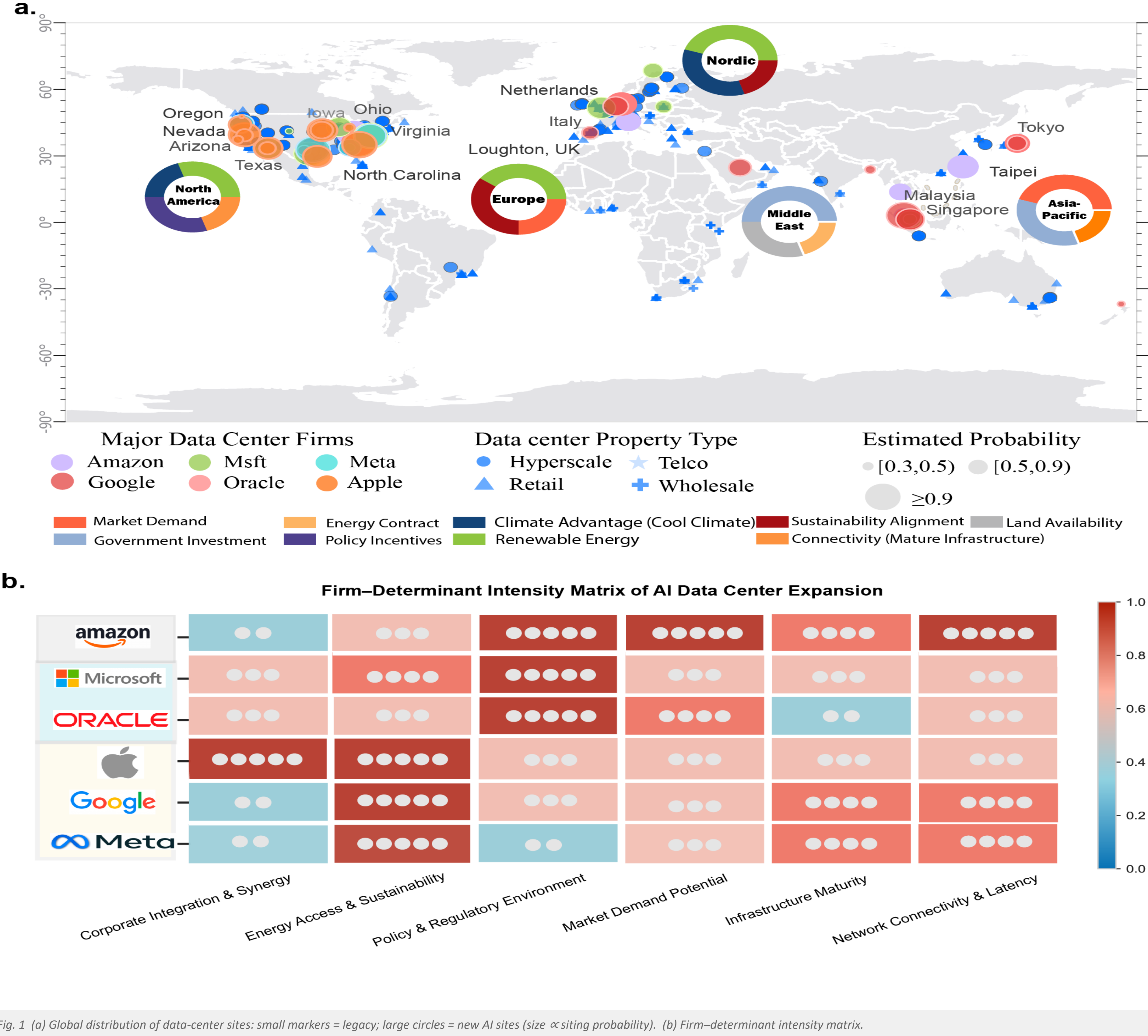


Fig. 1 (a) Global distribution of data-center sites: small markers = legacy; large circles = new AI sites (size ∝ siting probability). (b) Firm–determinant intensity matrix.

Result 2 — Electricity Demand Forecast (2025–2030)

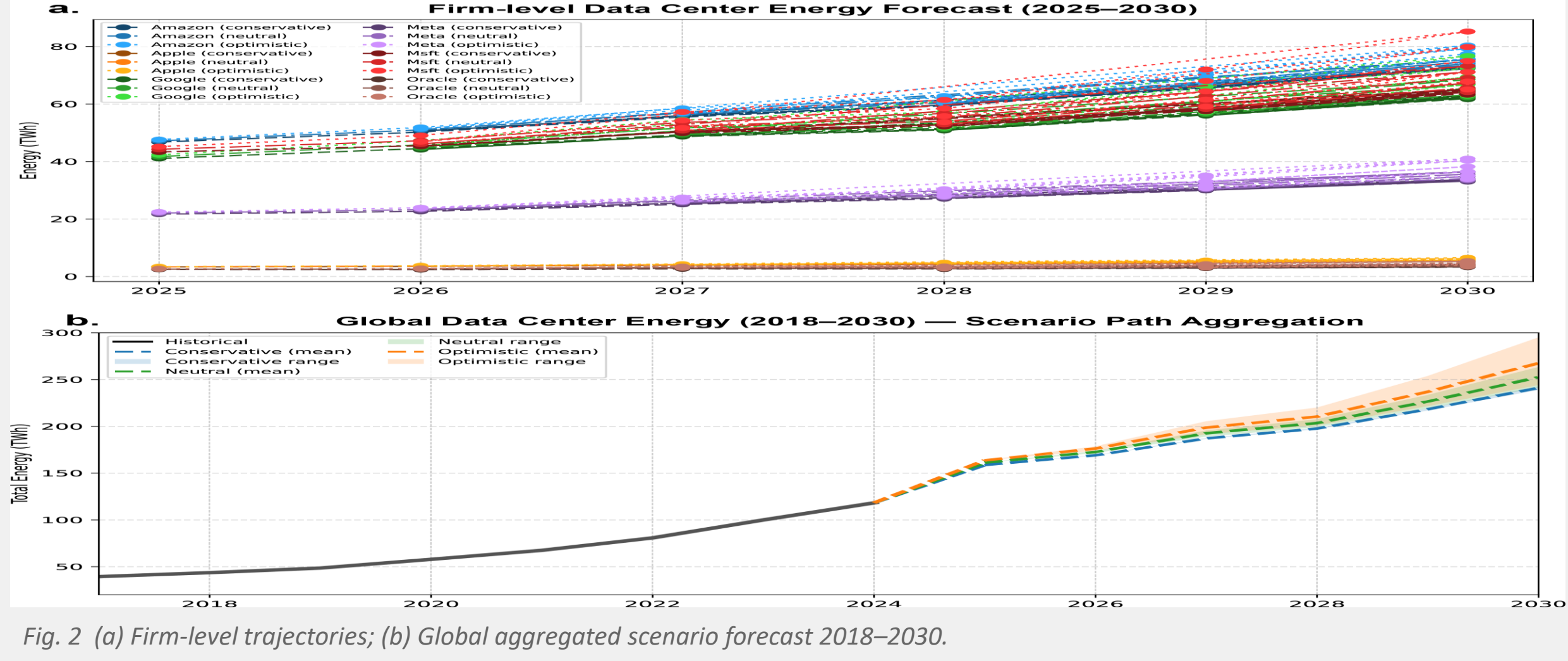


Fig. 2 (a) Firm-level trajectories; (b) Global aggregated scenario forecast 2018–2030.

Metric	Value
Baseline 2024	~118 TWh
Conservative 2030	~239 TWh
Optimistic 2030	~295 TWh
CAGR	13–17 %
IEA cross-check	~265 TWh
Equiv. load	~27 M US HH/yr

- Amazon, Microsoft, Google dominate absolute demand
- Growth accelerates post-2027 with large-scale generative-AI deployment

Result 3 — Regional Power Stress Index (PSI) >265 TWh implied

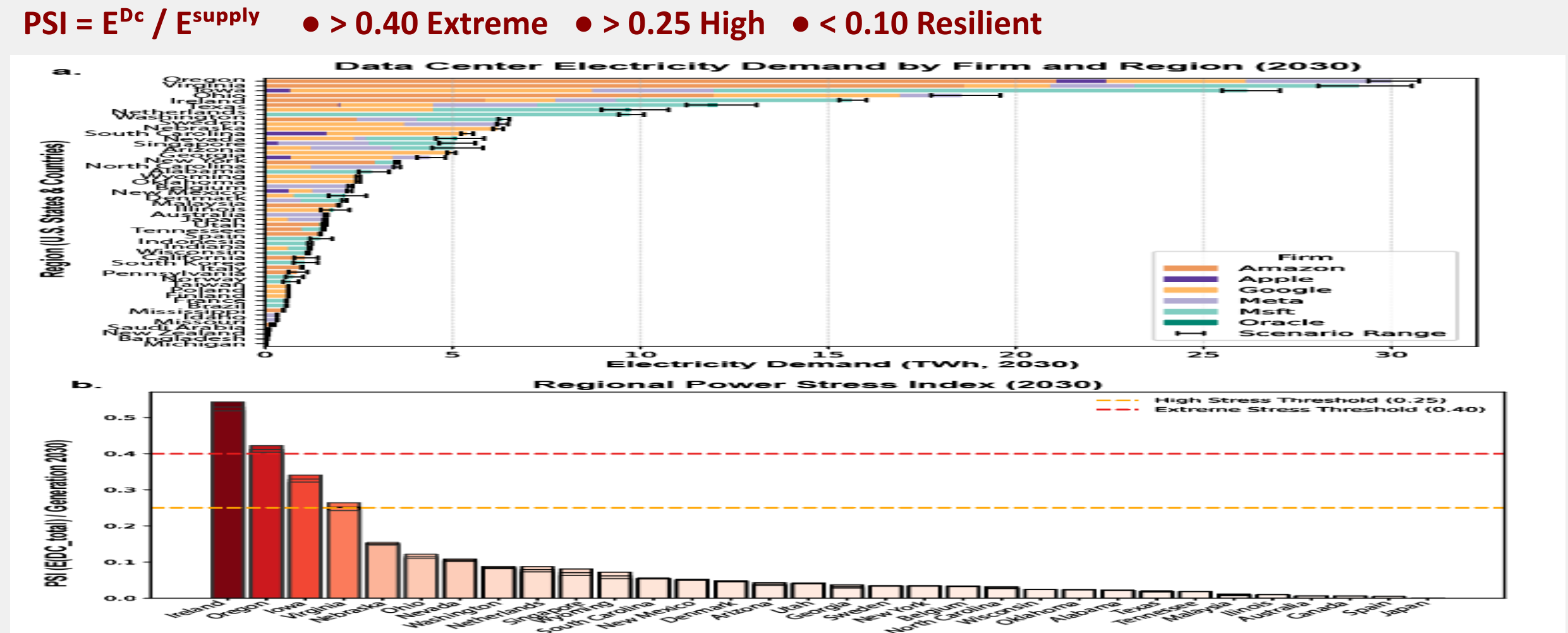


Fig. 3 (a) DC demand by firm & region 2030. (b) PSI bar chart; dashed lines at 0.25 and 0.40 thresholds.

Region	PSI (2030)	Status
Ireland	~0.50	Extreme ⚠
Oregon	~0.40	High ⚠
Iowa / Virginia	0.20–0.30	Elevated
Texas / Japan	<0.10	Resilient ✓

- <10 regions account for ≈two-thirds of total projected AI electricity load
- PSI > 0.25 signals need for dedicated transmission expansion
- Texas advantage: independent grid, renewable diversity, efficient interconnection

Conclusions & Policy Implications

Key Findings

- AI infrastructure is evolving from a marginal load into a structural grid component
- Spatial clustering creates a “digital energy divide” between stressed and resilient regions
- Absent coordination, AI demand could offset planned renewable additions by 2030

Policy Recommendations

- Incorporate AI load projections into national capacity-expansion plans
- Pair fiscal incentives with mandatory renewable procurement requirements
- Deploy data centers as anchor loads for PPAs and demand response
- Strengthen grid interties in high-stress corridors (Oregon, Ireland, Virginia)
- Direct new investment toward renewable-rich, resilient regions (Texas, Nordic)