Supplementary Information

Table S1 Summary of the six major hyperscale operators, their estimated market shares, and AI-driven data center expansion strategies

firm	**Infrastructure scale (2025)**	**Approx. standing**	**Est. market share (2025)**	**AI expansion strategy**
Amazon Web Services (AWS)	32 cloud regions, 102 availability zones (+4 planned)	Largest operator globally	30–31% (global cloud infrastructure)	Rapid build-out of AI-optimized infrastructure (e.g., Trainium and Inferentia chips); launch of Bedrock and Titan foundation models; dedicated AI clusters for customers.
Microsoft Azure	62 regions, 120 availability zones, 200+ data centers	Second largest globally	∼20%	Massive AI investment via OpenAI partnership; deployment of NVIDIA H100 GPU clusters and liquid-cooled servers; expansion in Iowa, Virginia, and Sweden.
Google Cloud Platform (GCP)	39 regions, 118 availability zones, 35 owned centers across 10+ countries	Third largest globally	∼13%	Integration of TPUv5 AI accelerators; target of 24/7 carbon-free energy by 2030; expanding global AI training hubs.
Meta Platforms	24 data-center campuses (>53 million sq ft)	Leading social-AI infrastructure provider	∼4–6% (hyperscale capacity share)	Refit of compute sites into AI-specific clusters for LLaMA and generative-AI workloads; major CapEx on GPU superclusters; in-house energy-efficient design.
Oracle Cloud Infrastructure (OCI)	46 regions, 56 availability zones	Fifth largest cloud provider	∼3–4%	Partnership with NVIDIA (2024–2025) for regional AI infrastructure; focus on enterprise-grade AI workloads and secure environments.
Apple	˜20 data centers (U.S., Europe)	Mid-scale, vertically integrated	<2% (internal capacity)	Emphasis on on-device/private AI; investment in internal "AI Compute Fabric" for LLM training and edge inference; limited public-cloud exposure.

Note: Market share values are indicative and derived from cloud-infrastructure and hyperscale-capacity data (Statista 2025; Brightlio 2024; Synergy Research 2024). Exact operator-level shares for total data-center capacity are not publicly disclosed, so these proxies reflect each firm's relative scale within the global data-center ecosystem.

The analysis focuses on six technology firms—Amazon, Microsoft, Google, Meta, Oracle, and Apple—selected based on three criteria. First, market relevance: these firms collectively account for approximately 70–75% of global hyperscale and cloud-linked data-center capacity, as reported by industry analyses (e.g., Synergy Research Group, 2024). Second, data availability: each firm provides sufficiently detailed and consistent disclosures, including sustainability reports, SEC filings, and capital-expenditure announcements, enabling the extraction of infrastructure and energy-related parameters. Third, AI-infrastructure materiality: all selected firms have made publicly documented, large-scale investments in AI-specific computing infrastructure (e.g., training clusters, AI accelerators, and hyperscale AI campuses) during the study period (2015–2025), making their expansion trajectories consequential for regional electricity systems.

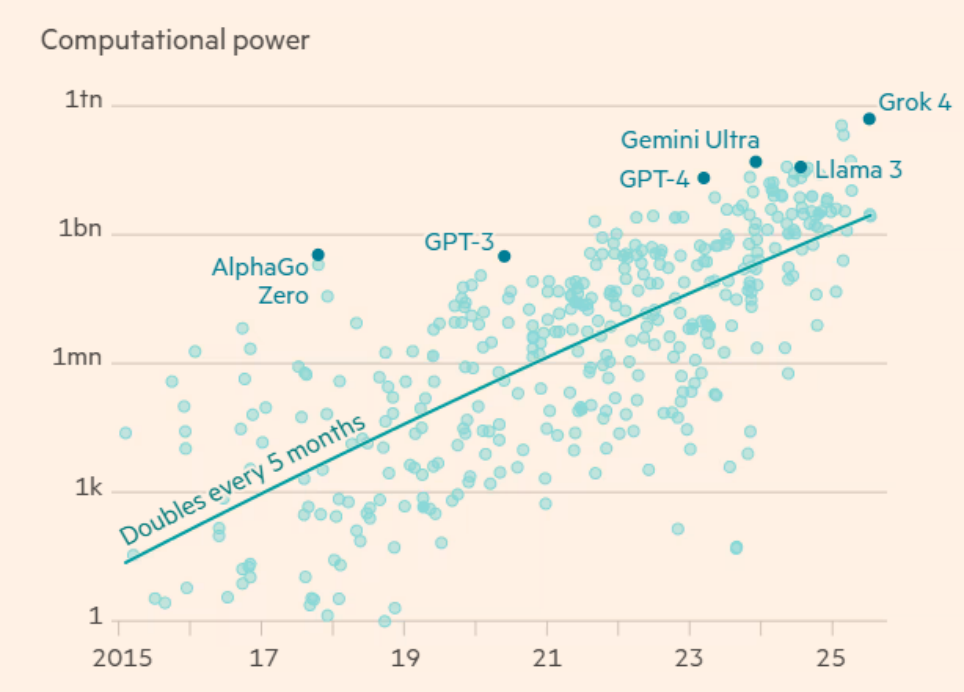


Fig. S1 Data includes selected AI models and is shown in log scale. Estimates are based on publicly reported numbers and calculations by Epoch.ai. Source: Epoch.ai, adapted from *Financial Times* (https://ig.ft.com/ai-data-centers/).

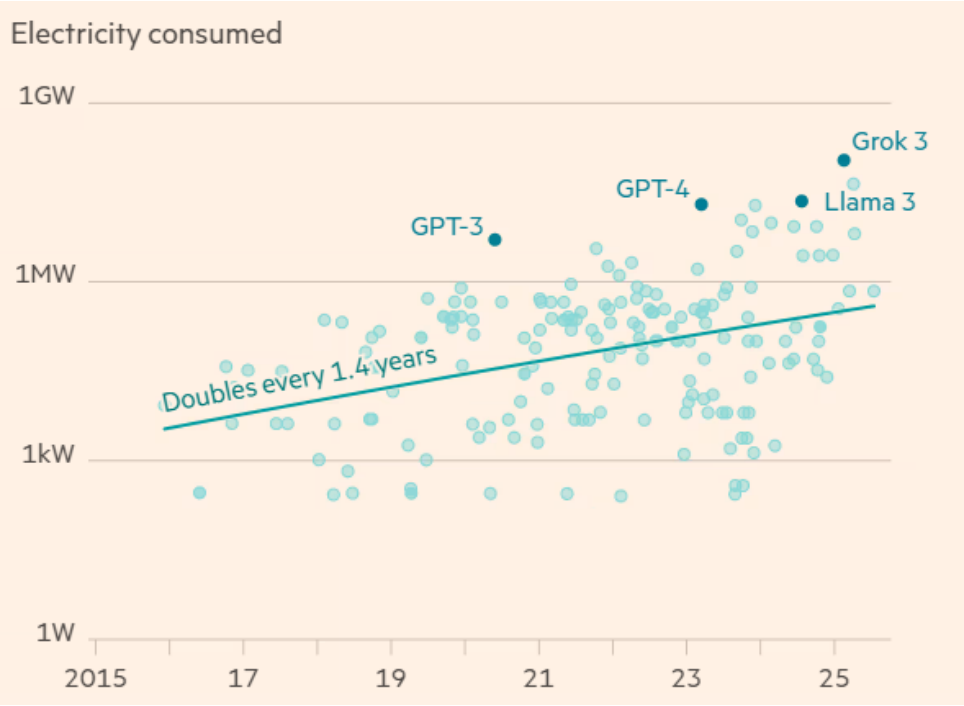


Fig. S2 Computation power of leading AI systems, measured in petaflops (10^{15} floating-point operations per second). Source: Epoch.ai, adapted from *Financial Times* (https://ig.ft.com/ai-data-centers/).

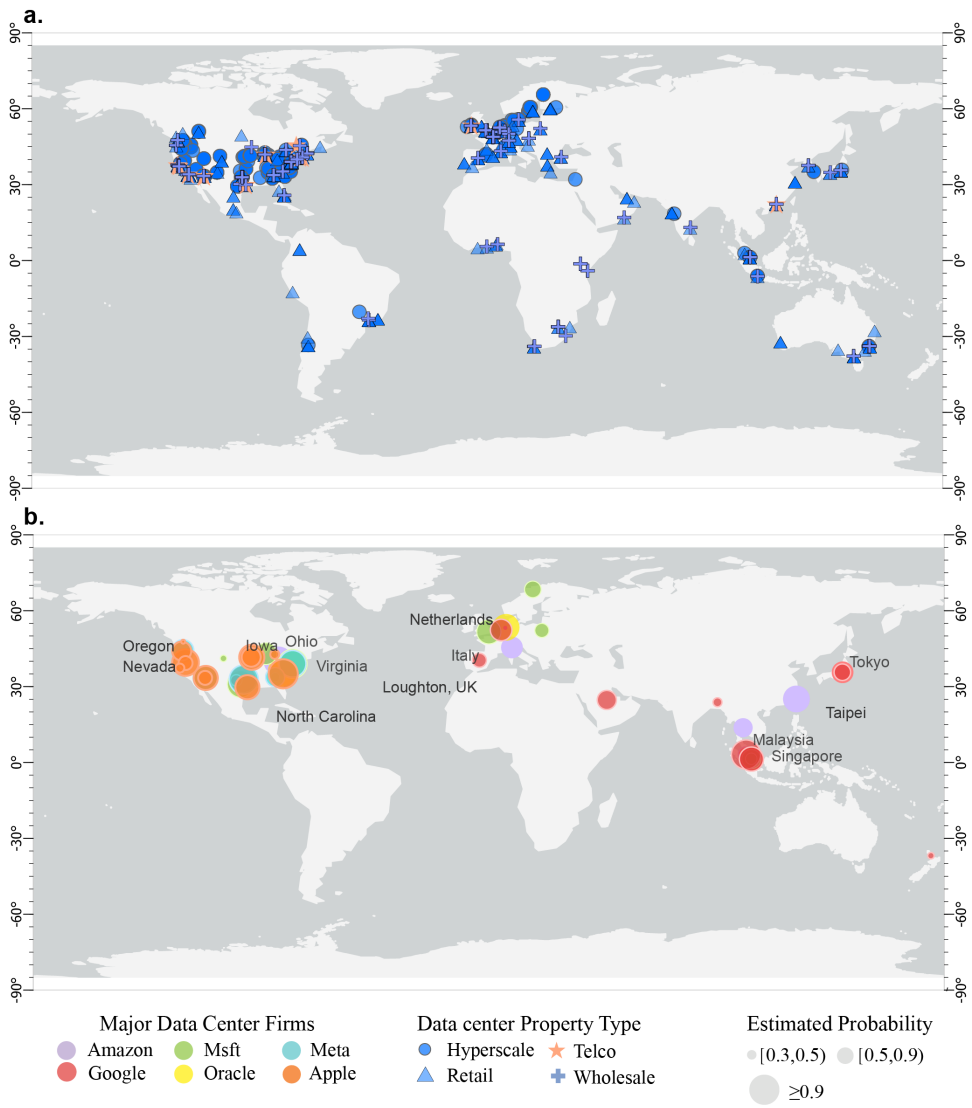


Fig. S3 Decomposition of AI data-center siting layers. (a) Legacy data-center infrastructure. (b) Spatial distribution of AI-specific data-center sites inferred from corporate disclosures and public announcements.

Figure S3b reveals a pronounced concentration of AI-specific data-center deployments in North America, Europe, and selected Asia-Pacific hubs, with limited representation in Africa and Australia. This pattern should not be interpreted as the absence of conventional data-center infrastructure in these regions. Rather, the LLM/RAG framework identified relatively few large-scale AI-oriented deployments based on corporate disclosures, policy documents, and public announcements. In Africa, this may reflect comparatively limited hyperscale AI investment, lower concentrations of cloud-computing infrastructure, and fewer publicly disclosed AI-campus projects. In Australia, although substantial cloud and colocation infrastructure exists, fewer AI-specific expansion announcements and large-scale AI-training deployments were identified during the study period relative to North America, Europe, and major Asia-Pacific hubs. Consequently, these regions receive lower representation within the inferred AI-deployment layer.