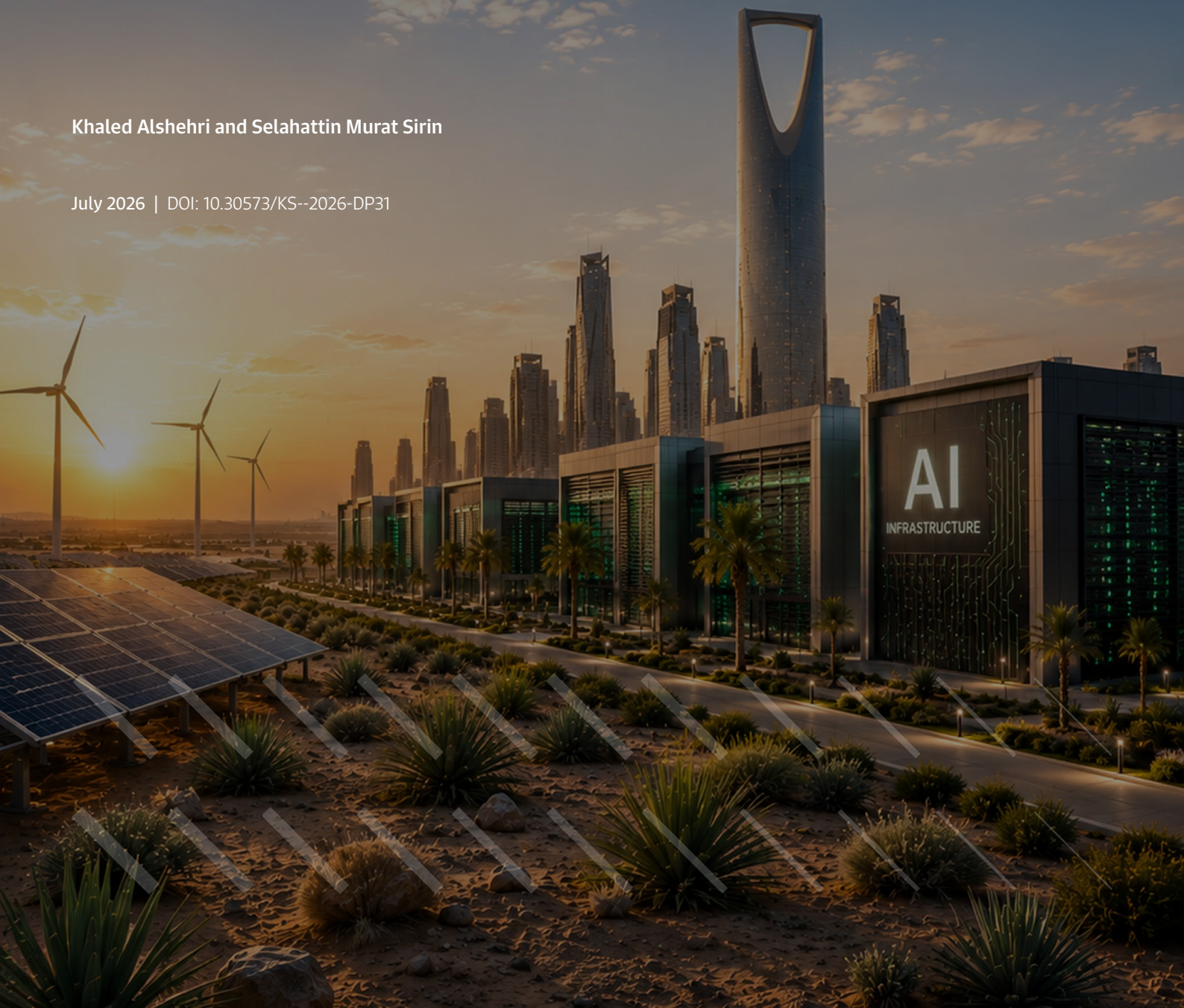
Discussion Paper

The Levelized Cost of AI Infrastructure in Saudi Arabia

Energy-AI Nexus Analysis of Competitiveness and Policy

Khaled Alshehri and Selahattin Murat Sirin

July 2026 | DOI: 10.30573/KS--2026-DP31



About KAPSARC

KAPSARC is an advisory think tank within global energy economics and sustainability providing advisory services to entities and authorities in the Saudi energy sector to advance Saudi Arabia's energy sector and inform global policies through evidence-based advice and applied research.

This publication is also available in Arabic.



Legal Notice

© Copyright 2026 King Abdullah Petroleum Studies and Research Center (“KAPSARC”). This Document (and any information, data or materials contained therein) (the “Document”) shall not be used without the proper attribution to KAPSARC. The Document shall not be reproduced, in whole or in part, without the written permission of KAPSARC. KAPSARC makes no warranty, representation or undertaking whether expressed or implied, nor does it assume any legal liability, whether direct or indirect, or responsibility for the accuracy, completeness, or usefulness of any information that is contained in the Document. Nothing in the Document constitutes or shall be implied to constitute advice, recommendation or opinion. The views and opinions expressed in this publication are those of the authors and do not necessarily reflect the official views or position of KAPSARC.

Key Points

- Saudi Arabia is rapidly scaling digital infrastructure amid intensifying global competition for data center investment.
- A four-pillar “Energy-AI Nexus” framework links energy governance and AI infrastructure strategy, with each pillar mapping directly to a parameter in the levelized cost of computing.
- Saudi Arabia’s competitive electricity tariff offsets higher cooling loads, sustaining cost competitiveness across 10 comparator markets with a baseline hosting cost of $0.0045-$0.0052 per million tokens for representative Llama 3 70B inference.
- Carbon and water externalities can increase computing costs by up to 14.7%, with grid decarbonization identified as a lever for reducing environmentally adjusted costs.
- AI hardware and utilization are the dominant drivers of long-run compute cost reduction, though electricity tariff design and facility efficiency standards remain the most policy-actionable levers for governments.

Abstract

Saudi Arabia is expanding digital infrastructure as global competition for AI-oriented compute industries accelerates. Yet the growth of data centers depends directly on the energy systems that power them, creating tradeoffs between competitiveness, resource use, and sustainability. This paper examines the "Energy-AI Nexus" in Saudi Arabia and explores how the Kingdom can convert energy advantage into durable digital advantage. We develop a four-pillar policy framework (Integrated Governance and Planning, Adaptive Policy and Regulatory Framework, Resource and Capability Development, and Transparency and Accountability) linking AI infrastructure expansion with energy governance, and we map each pillar to the economic parameters that drive compute costs. We then apply the levelized cost of computing (LCOC) to evaluate AI-oriented data center investments under varying tariff, efficiency, and utilization conditions, and extend the framework to incorporate carbon and water externalities. Anchored to H100-class hardware and Saudi Arabia's electricity tariffs, as low as $48/MWh (the SERA 2025 cloud computing tariff) and an assumed upper bound of $85/MWh, the baseline infrastructure hosting cost is estimated at $0.0045-$0.0052 per million tokens for Llama 3 70B inference, among the most cost-competitive under modeled assumptions across 10 comparator markets examined. Carbon and water externalities can raise compute costs by up to 14.7% depending on grid emission factor and water pricing assumptions, with grid decarbonization identified as a lever for reducing environmentally adjusted costs. The findings indicate that sustained digital competitiveness requires not only affordable energy but also efficient infrastructure, high facility utilization, and explicit integration of environmental constraints into policy design.

Keywords: AI Data Centers, Energy-AI Nexus, Levelized Cost of Computing, AI Infrastructure

01 Introduction

Digital technologies, particularly artificial intelligence (AI),[1] are reshaping the global economy, and data centers are critical infrastructure for digital transformation and economic competitiveness (Upham, Sovacool, and Monyei 2023). Unlike conventional information technology facilities built for general-purpose computing, new AI-oriented data centers are designed for high-performance workloads, combining accelerator-class chips, dense servers, and advanced power and cooling systems.

Globally, installed IT equipment power capacity (GW IT, excluding cooling and other facility auxiliaries) exceeded 111 GW in 2024, with more than 60% concentrated in the United States and China (Alshehri et al. 2026). At the same time, resource-rich countries are seeking to leverage energy advantage, climate conditions, and policy incentives to attract data center investment and capture digital value. Among energy-exporting economies, Norway and Saudi Arabia are emerging as regional hubs, with projections indicating rising shares of digital infrastructure investment over the coming decade (IEA 2025b).

However, expanding data center investments place additional pressure on energy systems. Data center capacities are projected to double by 2030, exceeding 224 GW IT (Alshehri et al. 2026). Given the fossil fuel intensity of many grids, the environmental footprint can be substantial and, in some cases, comparable to the annual emissions of mid-sized countries (OECD 2022).

Experiences in Chile and Uruguay demonstrate that misalignment between digital investment and regulatory capacity can create grid stress, community opposition, and investor uncertainty (PBS News 2024; Montgomery 2024). Without well-designed policies, rapid expansion raises concerns about grid adequacy, sustainability, and infrastructure readiness (Calvino, Dechezleprêtre, and Haerle 2025; Soares, Yarime, and Klemun 2024).

An emerging literature examines the interdependence between digital infrastructure and energy systems (OECD 2022; Calvino, Dechezleprêtre, and Haerle 2025). Three strands are directly relevant. First, technoeconomic analyses of data center costs – notably the levelized cost of computing (LCOC) (Kristiansen, Hjelmeland, and Korpas 2024) – remain confined to high-income economies. Second, digital infrastructure governance frameworks (Soares, Yarime, and Klemun 2024; Li et al. 2023) lack explicit integration of energy system constraints. Third, sustainability studies (OECD 2022; Huang et al. 2020) rarely connect environmental impacts to investment economics.

Together these leave a gap: no prior study has jointly evaluated the policy framework, cost competitiveness, and environmental externalities of AI data center investment in an energy-exporting economy.

1 For a list of the abbreviations used in the paper, see Appendix A.

This paper examines how Saudi Arabia can translate its comparative advantage in energy resources into durable digital advantage. By 2024, the Kingdom operated 58 facilities with 290 MW of IT capacity, with projections exceeding 4 GW by 2030 under high-growth scenarios (Alshehri et al. 2026). Rather than focusing solely on capacity expansion, we evaluate the economic and policy conditions that determine long-term competitiveness. We apply a four-pillar energy-digital nexus framework – Integrated Governance and Planning, Adaptive Policy and Regulatory Framework, Resource and Capability Development, and Transparency and Accountability – to assess policy alignment. We then conduct a technoeconomic analysis of AI-oriented data centers using the LCOC framework introduced by Kristiansen, Hjelmeland, and Korpas (2024) and extend it to incorporate carbon and water externalities.

Our findings show that Saudi Arabia has established an integrated digital policy regime that supports data center investment, though structural risks remain. While preferential electricity tariffs contribute to a cost advantage, competitiveness is highly sensitive to utilization rates, hardware efficiency, and environmental constraints. The analysis indicates that long-term digital advantage depends not only on energy pricing but also on operational efficiency and explicit integration of environmental externalities into policy design.

Building on these findings, this paper contributes to the literature in three ways. First, it introduces the "Energy-AI Nexus" framework linking AI infrastructure governance to energy system planning in resource-exporting economies. Second, it provides the first application of the LCOC to an energy-exporting economy, extending it to incorporate environmental externalities and translating results into a US$/token metric. Third, it delivers an integrated assessment of Saudi Arabia benchmarked against 10 comparator markets (the United States represented by Texas and Virginia), with full data in the Appendix.

The remainder of the paper proceeds as follows: Section 2 reviews global data center expansion and its strategic implications. Section 3 develops the Energy-AI Nexus policy framework (a detailed assessment of Saudi Arabia's institutional architecture against this framework is provided in Appendix D). Section 4 presents the LCOC framework, the baseline analysis, and an international comparison. Section 5 provides US$/token interpretation and practical insights. Section 6 integrates environmental impacts. Section 7 concludes. The Appendix provides detailed data tables.

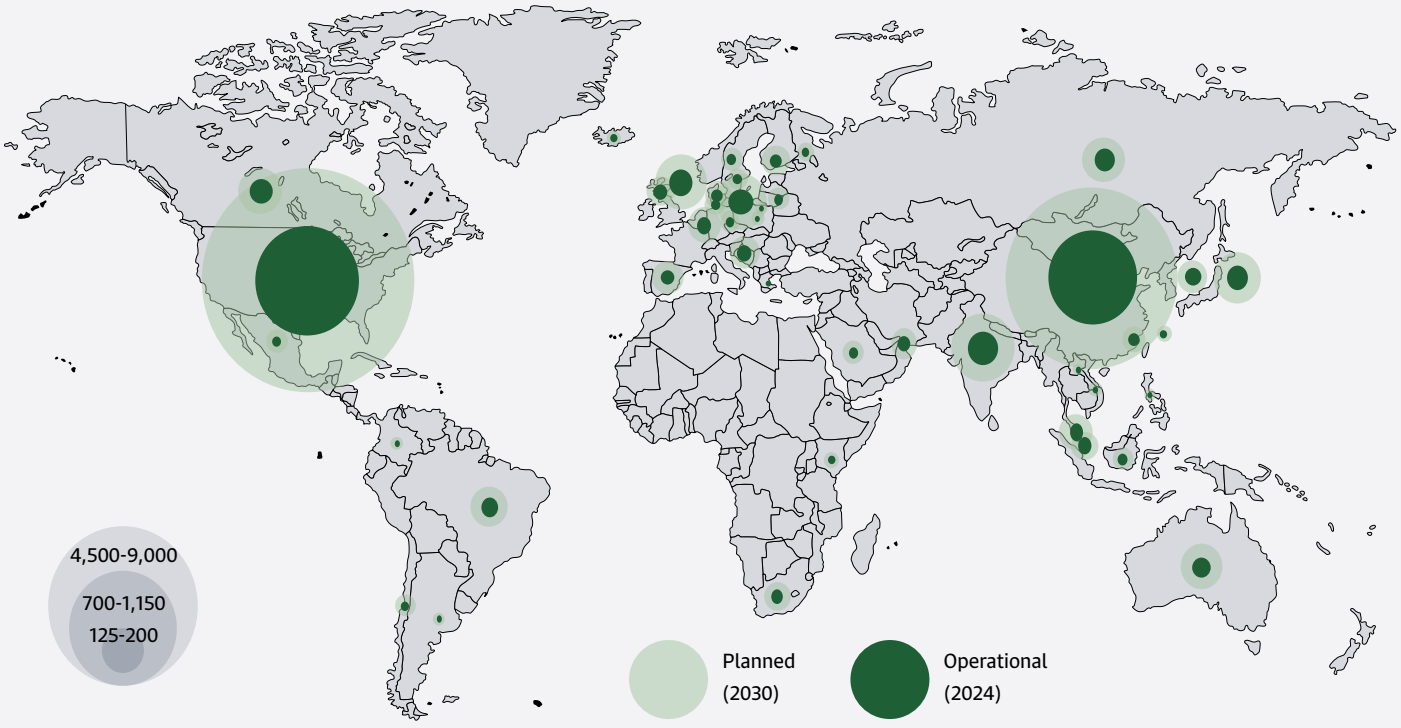
Global Data Center Expansion and Strategic Implications

02

2.1 Regional Expansion Dynamics

AI diffusion has become the primary driver of demand in the data center sector. Globally, more than 11,000 facilities operate today (S&P Global), with the United States hosting 5,426, followed by Germany (529), the United Kingdom (523), and China (449). Total installed capacity (excluding cryptomining unless otherwise noted) is projected to reach 224 GW IT by 2030, a CAGR of roughly 12% over 2024-2030 (Figure 1). Electricity consumption is projected to nearly double to 1,891 TWh (CAGR ~14%).

Figure 1. Overview of current and planned data center capacities globally.



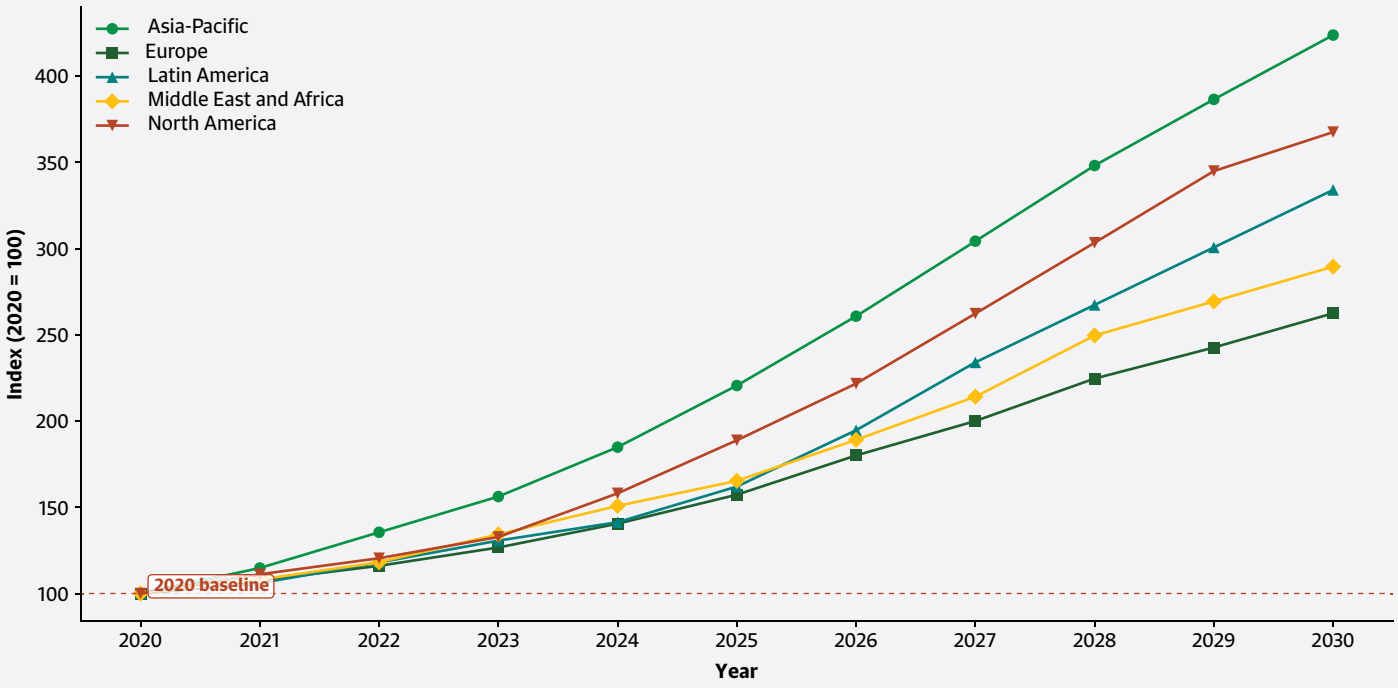
Source: Alshehri et al. (2026).

Within this trajectory, Saudi Arabia is projected to record the fastest growth in data center electricity demand in the Middle East and Africa region, reaching 10-41 TWh by 2030, approximately 3%-11% of national electricity demand (Alshehri et al. 2026; Mikayilov and Darandary 2024).

Figure 2 shows indexed growth in data center electricity demand by region (2020-2030). Asia-Pacific records the strongest relative expansion at more than four times its 2020 level, driven by rapid digitalization across China, India, and Southeast Asia. Latin America and North America each approximately triple, while the Middle East and Africa region (smaller in absolute terms) nearly triples its 2020 level. Europe expands at the slowest pace.

For Saudi Arabia, this trajectory positions the Kingdom to capture a disproportionate share of MEA demand growth.

Figure 2. Regional growth trajectories of data center electricity demand, indexed to 2020 levels.

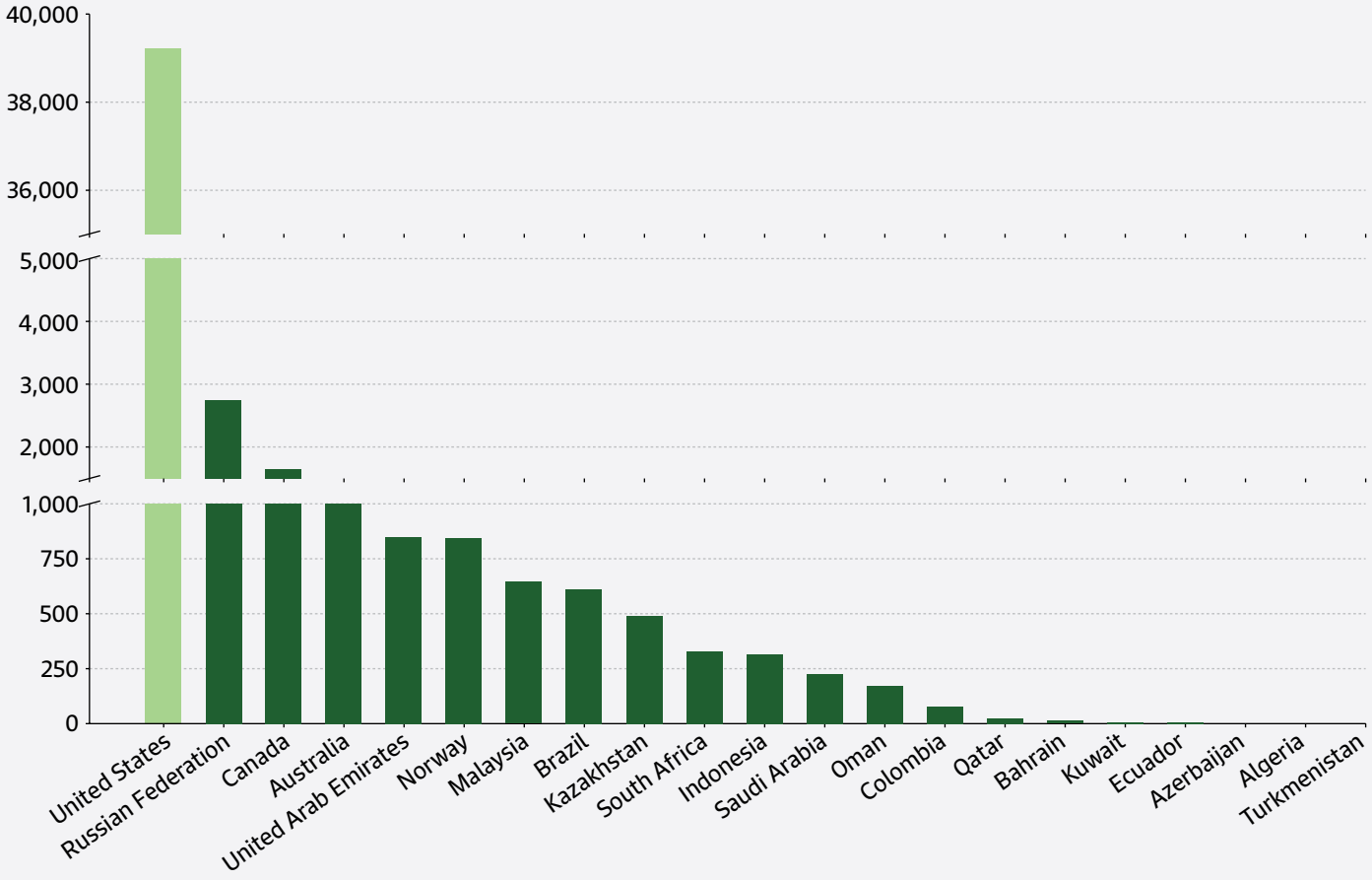


Source: Authors, using 451 Research from S&P Global Energy Horizons (2025) data.

Figure 3 presents operational data center capacity by country as of Q1 2025. The United States is a clear outlier at approximately 39 GW (an order of magnitude larger than any other country), followed by the Russian Federation at roughly 2.7 GW. U.S. capacity is highly concentrated at the subnational level: Northern Virginia alone hosts ~5 GW (the world's largest single cluster), with Texas, Illinois, and a handful of other states accounting for most of the remainder – placing Saudi Arabia's 2030 High scenario of roughly 4 GW on a comparable scale to a leading U.S. state market. A second tier includes Canada and Australia (~1–1.5 GW each). The United Arab Emirates leads the Middle East at ~0.8 GW. Norway, Malaysia, Brazil, Kazakhstan, South Africa, Indonesia, and Saudi Arabia represent emerging regional hubs.

These global and regional dynamics establish the structural backdrop against which Saudi Arabia's digital infrastructure ambitions could be evaluated.

Figure 3. Operational data center capacity by country (MW IT), mid-2025.



Note: Values represent operational facilities, including cryptomining, as of Q1 2025.
Source: Authors, using 451 Research from S&P Global Energy Horizons (2025) data.

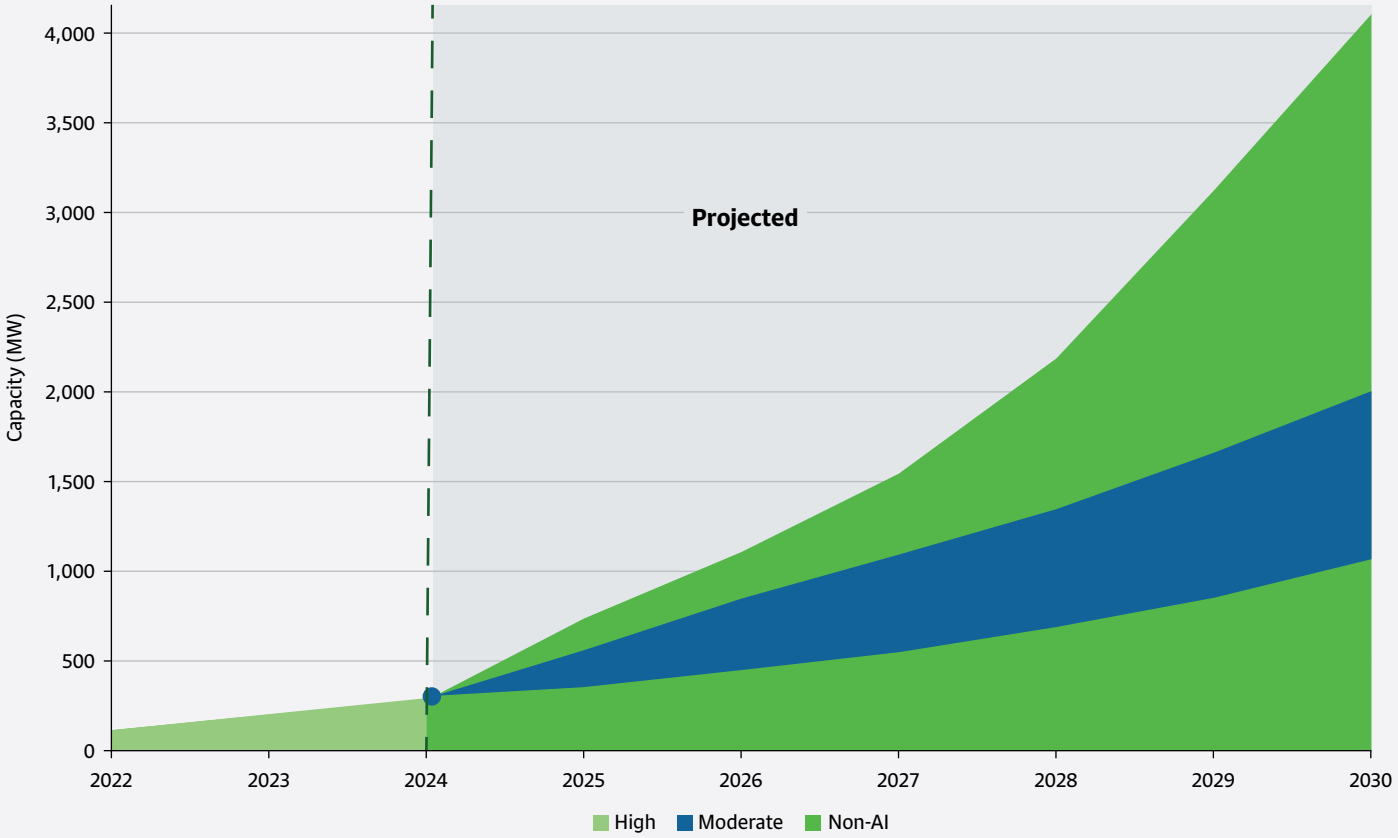
2.2 Saudi Capacity Scenarios

Figure 4 presents projected installed IT capacity in Saudi Arabia under three alternative post-2024 trajectories: a "Non-AI" baseline, a "Moderate" expansion scenario, and a "High AI"-driven scenario. From a base of approximately 290 MW in 2024, installed capacity could reach about 1,050 MW under the Non-AI path, 2,000 MW under the Moderate scenario, and up to 4,100 MW under the High scenario by 2030.

These scenarios imply sharply divergent growth dynamics: CAGRs of roughly 24% (Non-AI), 38% (Moderate), and over 55% (High AI-driven) over 2024-2030. Under the High trajectory, installed IT capacity would increase more than 14-fold within six years.

At roughly 4 GW of installed IT capacity, Saudi Arabia could move from a modest 2024 base to a potential top-10 global ranking by 2030. In sub-national terms, the 4 GW target is comparable in scale to Northern Virginia's operational footprint ~ 5 GW as of Q1 2025), the world's largest single data center cluster, and exceeds leading secondary U.S. markets such as Dallas-Fort Worth. Even the Moderate scenario positions Saudi Arabia as a leading secondary hub among energy-exporting economies.

Figure 4. Projected installed IT capacity in Saudi Arabia under alternative expansion scenarios.



Note: The High scenario reflects accelerated AI-driven investment, while the other scenarios represent conservative deployment.
Source: Alshehri et al. (2026).

2.3 Economic and Environmental Trade-Offs of Data Center Expansion

Data center investment can generate substantial economic benefits: digital infrastructure quality, foreign direct investment, human capital formation, data sovereignty, and renewable energy integration when supported by coherent regulatory and energy frameworks (IEA 2025a; Sovacool, Monyei, and Upham 2022; Gaglio, Kraemer-Mbula, and Lorenz 2022).

These gains are not automatic. Data centers are capital-intensive and generate relatively limited direct employment relative to their electricity footprint, and fiscal incentives or preferential tariffs may shift costs to other sectors if externalities are not properly internalized. Assessing economic value therefore requires moving beyond headline investment figures toward integrated evaluation of cost competitiveness, tax impacts, and opportunity costs.

Environmental trade-offs are equally significant. AI-centric facilities are energy-intensive and rely heavily on cooling systems that increase both electricity and water consumption (Huang et al. 2020; OECD 2022). Recent estimates suggest AI-driven data centers could account for up to 3.5% of global emissions by 2030, equivalent to approximately 700 million tons of CO_2 annually (Jamison et al. 2025); water intensity can reach 26 million liters per megawatt per year depending on climate and technology (Spindler, Hahn-Petersen, and Hosseini 2024), raising concerns in arid regions.

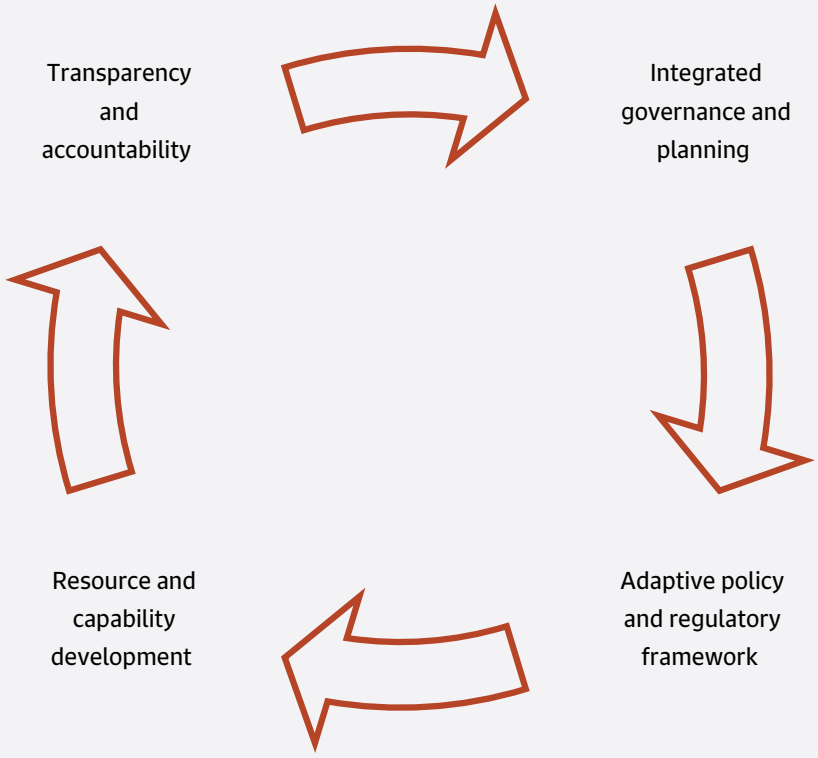
In energy-exporting economies, data centers are not merely ICT assets – they function as large industrial loads that directly interact with tariff design, grid planning, environmental regulation, and industrial policy. The following section develops a four-pillar Energy-AI Nexus framework that provides a structured basis for aligning digital expansion with these systems.

Policy Framework for the Energy-AI Nexus

03

Capturing the economic benefits of data center expansion requires a coherent regulatory and policy framework that aligns digital growth with sustainability, energy security, and economic transformation objectives (OECD 2022; IEA 2025a). We develop an Energy-AI Nexus framework spanning integrated planning, regulation, capability development, and accountability (Figure 5).

Figure 5. The Energy-AI Nexus policy framework.



Note: The four pillars are mutually reinforcing: governance creates conditions for effective regulation; regulation shapes the investment environment for capability development; capability development enables meaningful accountability; and accountability feeds back into governance design.
Source: Authors.

We discuss each pillar below:

- **Integrated Governance and Planning:** Effective digital infrastructure policy requires alignment between decarbonization objectives, energy system strategies, and the full life cycle of data center development (Oró et al. 2015; OECD 2022). This includes coordinating renewable energy deployment with data center siting, integrating AI compute demand forecasting into grid planning, and designing infrastructure around waste heat and resource recovery (Wang et al. 2025; Li et al. 2023). AI data centers exhibit fundamentally different load profiles from conventional digital infrastructure, with near-continuous high-density power draw for training and variable demand for inference, making dedicated demand-side planning essential. Multi-regional coordination is needed to balance demand against available grid capacity, locating facilities where electricity supply, connectivity, and demand centers converge (Soares, Yarime, and Klemun 2024). Governments therefore need to move beyond fragmented, market-driven approaches toward coordinated frameworks that integrate national, regional, and municipal priorities within broader sustainability objectives.

- **Adaptive Policy and Regulatory Framework:** An adaptive regulatory architecture mitigates environmental and system-level risks while preserving flexibility in a rapidly evolving landscape. Regulation must integrate efficiency standards, investment incentives, and accountability mechanisms that reflect the energy intensity of AI data center operations (Oró et al. 2015; Rong et al. 2016; Koronen, Åhman, and Nilsson 2020). Energy efficiency benchmarks, most directly power usage effectiveness (PUE) and water usage effectiveness (WUE), provide measurable performance standards and encourage continuous technological improvement. Progressive PUE thresholds, starting at achievable near-term targets such as 1.4 and tightening toward 1.2 over a defined policy horizon, can be embedded in licensing conditions without deterring investment, while WUE standards incentivize adoption of liquid and immersion cooling systems that reduce water intensity in arid climates. Several jurisdictions have progressively tightened PUE requirements (Li et al. 2023), and tariff structures and renewable procurement frameworks can further incentivize low-carbon electricity sourcing. Carbon pricing mechanisms and metrics such as the carbon usage effectiveness (CUE) align data center expansion with national decarbonization goals; where resources are constrained, governments may impose limits on water use, emissions, or environmentally intensive activities such as cryptocurrency mining. The objective is to direct scarce energy, water, and environmental resources toward socially and economically productive uses, with tradeoffs between investment attraction and environmental stewardship explicitly managed.

- **Resource and Capability Development:** Sustained digital competitiveness depends not only on infrastructure deployment but also on skilled labor, resilient supply chains, and innovation capacity. Human capital development is central: structured training programs in AI infrastructure engineering, machine learning operations, and graphics processing unit (GPU) cluster operations, certification pathways, and mid-career reskilling initiatives can help domestic labor markets keep pace with infrastructure expansion (Donnellan et al. 2024), since the skills required for AI-oriented facilities differ materially from those for conventional IT operations. Supply chain resilience is equally critical, as data center expansion depends on timely access to grid equipment, transformers, cables, cooling systems, and high-performance AI hardware; global bottlenecks and extended lead times can delay both digital and energy projects (IEA 2025a). Strategic localization, coordinated grid planning, and targeted industrial incentives may reduce vulnerability to external supply constraints. Finally, research and innovation policies, including investment in energy-efficient computing, advanced cooling, and AI integration within priority sectors, can deepen economic spillovers and limit dependency on foreign hardware, software, and talent.

- **Transparency and Accountability:** Transparency in electricity consumption, carbon intensity, water usage, compute utilization rates, and AI workload profiles enables policymakers to assess whether digital infrastructure aligns with national sustainability and energy objectives (Upham, Sovacool, and Monyei 2023; Li et al. 2023). For AI-oriented facilities, disclosure of training versus inference workload ratios is particularly valuable, since these workloads differ substantially in power demand, cooling requirements, and carbon intensity per unit of compute output, yet are rarely distinguished in current reporting frameworks. Standardized metrics and disclosure requirements reduce information asymmetries and improve regulatory oversight, particularly where preferential tariffs or fiscal incentives are provided. In resource-constrained systems, life-cycle assessment approaches allow evaluation of both direct and indirect impacts, extending beyond operational electricity use to embodied emissions, water stress, and supply chain effects.

Without systematic reporting, rapid digital expansion risks generating localized externalities that undermine broader economic gains. Accountability mechanisms therefore serve environmental goals, competitiveness objectives, and governance simultaneously, with transparent cost and performance metrics providing the foundation for integrated technoeconomic evaluation.

The four pillars of the Energy-AI Nexus framework operate through concrete economic levers shaping the unit cost of compute production. “Integrated Governance and Planning” determines electricity tariff design and grid coordination. “Adaptive Regulation” defines facility-level efficiency standards and the environmental pricing mechanisms through which carbon and water externalities are internalized. “Resource and Capability Development” influences hardware computing efficiency and utilization rates achievable through workload aggregation and skilled operations. “Transparency and Accountability” create the institutional conditions under which environmental costs are measured and disclosed, enabling the environmental cost extension developed in Section 6.

Together, these linkages allow the LCOC to function as a quantitative instrument for translating governance choices into measurable impacts on compute cost competitiveness. (Saudi Arabia’s institutional architecture is assessed against this framework in Appendix D.)

The LCOC Framework

04

Building on the four-pillar framework developed in Section 3, this section operationalizes the linkage between policy variables and compute economics through a technoeconomic analysis of Saudi Arabian data centers. The approach translates policy instruments (electricity tariffs, efficiency standards, utilization rates, and environmental pricing) into impacts on the unit cost of computing, with sensitivity analysis exposing trade-offs among cost, efficiency, and environmental factors.

We adopt a levelized cost framework, widely used in energy systems analysis for consistent comparison across technologies (IEA 2020; Ueckerdt et al. 2013; Aldersey-Williams and Rubert 2019; IEA 2025b). The standard LCOE does not transfer directly because data center output is computing, not energy.

Kristiansen, Hjelmeland, and Korpas (2024) introduce the LCOC to address this:

$$\text{LCOC} = \frac{c + \sum_{i=0}^{n-1} \frac{\alpha \cdot p \cdot k \cdot t + d \cdot c}{(1+r)^i}}{\sum_{i=0}^{n-1} \frac{\gamma \cdot k \cdot t}{(1+r)^i}} \quad (1)$$

The numerator represents discounted life-cycle costs per kW of IT load; the denominator captures discounted compute output in PFLOPs/kW. The resulting unit ($/PFLOP, or equivalently $/EFLOP at 10^{18} scale) represents the net-present cost of delivering computing services over the project lifetime.

The parameters are: c (CAPEX per kW of IT load), p (electricity tariff), α (PUE), k (load factor), t (annual operating hours), d (non-energy OPEX as a share of CAPEX), r (WACC), and γ (computing efficiency in PFLOPs/kW). Electricity price and PUE affect LCOC linearly through operating costs, while computing efficiency and load factor enter nonlinearly through their joint influence on cost allocation and compute output. This structure allows systematic evaluation of trade-offs among infrastructure design, operational practices, and policy variables.

The PUE measures facility-level energy efficiency and is defined as total data center energy divided by IT equipment energy:

$$\text{PUE}\ (\alpha) = \frac{\text{Total Data Center Facility Energy}}{\text{IT Equipment Energy}} = 1 + \frac{\text{Other Facilities Energy}}{\text{IT Equipment Energy}} \quad (2)$$

By definition, PUE ≥ 1; in practice, values exceed 1 due to cooling and auxiliary loads. Leading facilities report values near 1.1 (Google 2025; Van Zandt 2023), while regional averages in warmer climates are higher: ICEF (2025) reports a global

average of 1.41 in 2024 and a Middle East average above 1.9. S&P Global Data (2024) estimates Saudi data centers average approximately 1.53. In hot climates, achieving lower PUE requires advanced cooling technologies such as liquid or immersion cooling, alongside evaporative options whose efficiency depends on local humidity and water availability. While these systems increase upfront CAPEX, efficiency gains may reduce life-cycle costs, an interaction explicitly captured in the LCOC framework. As regional developers pursue PUE targets near 1.3 (TONOMUS 2025), the trade-off between capital intensity and operating efficiency becomes central to long-term competitiveness.

Load factor k captures infrastructure utilization. For AI-oriented facilities, higher utilization spreads fixed capital costs over greater compute output, substantially reducing LCOC. Reported industry load factors span 60%-95% depending on workload type: dedicated training clusters typically operate at 85%-95%, while mixed inference and development environments run closer to 60%-70%.

Computing efficiency γ(PFLOPs/kW) measures compute output per unit of power. Recent AI hardware advances have raised state-of-the-art chips above 3 PFLOPs/kW, with emerging architectures exceeding 7 PFLOPs/kW (Salvator, Bhargava, and Devleker 2025); see Table B1 in Appendix B for the full hardware comparison.

The remaining economic parameters, electricity price, CAPEX, and WACC, determine operating costs, capital recovery, and financing conditions, respectively. Together with the technological parameters above, they define the economic viability of alternative data center configurations.

We use the LCOC in Equation (1) and conduct a case study of Saudi Arabia to quantify how different parameters shape compute costs. Table 1 shows the parameters and assumptions used for the baseline model. Then, we use sensitivity analysis to show the trade-offs between inputs and derive policy-relevant insights.

Table 1. Parameters used for the baseline LCOC model.

Factor	Default value	Unit	Notes
PUE (α)	1.53	–	Average for existing data centers in Saudi Arabia; S&P Global (2025)
Load factor (k)	80	%	Reflects realistic utilization range from inference-heavy workloads (~60%) to dedicated training clusters (~90%-95%); Aterio (2025)
Computing efficiency (γ)	3.137	PFLOPs/kW	NVIDIA DGX H100, representative of the workload profile in Saudi Arabia; see Table B1
Electricity price (p)	48-85	$/MWh	Baseline = $48/MWh (SERA 2025 cloud computing tariff); upper bound is an assumed ceiling for tariffs outside the cloud computing tariff regime.
CAPEX (c)	10,000	$/kW	Kristiansen, Hjelmeland, and Korpas (2024); also consistent with Equinix planned hyperscale AI data center; Malik (2025)
WACC (r)	7	%	Reflecting the investment risk profile of sovereign-backed infrastructure in Saudi Arabia, consistent with discount rates applied in regional energy system analyses; IEA (2025b)
Annual non-energy OPEX/CAPEX (d)	10	%	Baseline analysis by Kristiansen, Hjelmeland, and Korpas (2024)
Project lifetime (n)	15	Years	Assumed project duration
Average time per year (t)	8,760	h/year	Hours in a calendar year

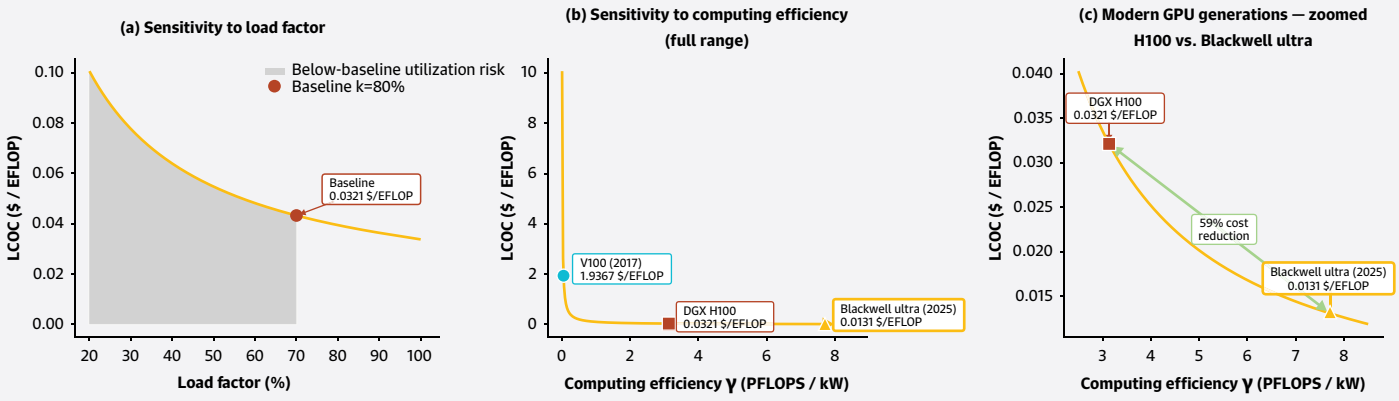
Note: All parameters reflect 2024-2025 conditions unless otherwise stated. Sensitivity analysis conducted later explores the impact of deviations from these baseline values.

For the baseline parameters, anchored to the NVIDIA H100 computing efficiency representative of existing operational large-scale AI deployments in Saudi Arabia, the LCOC is estimated at 0.032 $/EFLOP.

4.1 Sensitivity Analysis

Figure 6 examines the primary operational and hardware drivers of compute cost. Panel (a) confirms that LCOC is highly responsive to utilization, particularly below 60%-70% where fixed capital costs are spread over limited output. Panel (b) illustrates hardware-driven cost reduction across GPU generations: from the V100 (0.052 PFLOPs/kW, 1.9367 $/EFLOP) to the H100 (3.137 PFLOPs/kW, 0.032 $/EFLOP) and the Blackwell Ultra (7.71 PFLOPs/kW, 0.013 $/EFLOP), which is approximately a 150-fold reduction over eight years. Panel (c) isolates the H100-to-Blackwell Ultra transition, showing a 59% cost reduction within three years.

Figure 6. LCOC sensitivity to load factor (a), computing efficiency across GPU generations (b), and zoomed H100 vs. Blackwell Ultra (c).

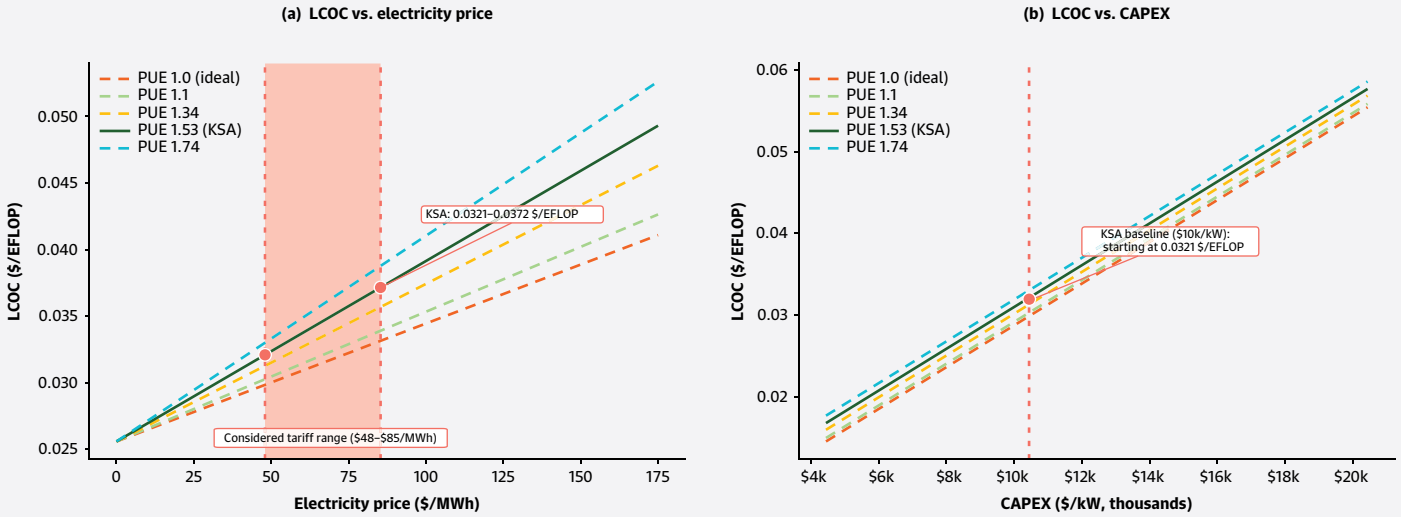


Source: Authors, using S&P Global (2024) data.

Computing efficiency and load factor are largely shaped by hardware choices and operational practices, leaving electricity pricing and facility design as the more actionable policy levers. In Saudi Arabia, the hot climate elevates baseline PUE relative to cooler regions, but low cloud computing tariffs offset this, making the PUE/electricity-price interaction central to long-term competitiveness.

Figure 7 shows LCOC sensitivity to electricity price (panel a) and construction CAPEX (panel b) across five PUE scenarios spanning 1.0 (ideal) to 1.74 (high cooling load). In panel (a), the lines fan out from a common origin at zero electricity cost, with steeper slopes for higher PUE values. At Saudi Arabia's tariff range, the difference in LCOC between PUE 1.0 and 1.74 remains modest, confirming that low tariffs compress the cost penalty of thermally intensive operating conditions. Panel (b) shows that PUE has limited differentiating power for CAPEX sensitivity; the five curves are nearly parallel because the capital recovery term is independent of PUE. CAPEX level dominates: at the baseline $10,000/kW, KSA's LCOC is 0.0321 $/EFLOP; the $8,000/kW low and $15,000/kW high scenarios yield 0.0270 and 0.0449 $/EFLOP, respectively. Full scenario analysis is in Appendix C.

Figure 7. Five scenarios range from ideal efficiency (PUE = 1.0) to high cooling load conditions (PUE = 1.74), with KSA's average of 1.53 highlighted as the baseline case.



Source: Authors, using 451 Research from S&P Global Energy Horizons (2025)z data.

Figure 8 shows a contour map of LCOC as a function of electricity price and PUE, with 10 comparator markets overlaid as reference points. The United States is represented by its two leading data center markets — Texas and Virginia — at forecast 2026 data center tariffs; Malaysia is shown at its Regulatory Period 4 (RP4) dedicated AI tariff. The remaining comparators use national industrial electricity tariffs. We compute facility-level PUE averages from S&P Global (2025) for each market. Each contour line represents a constant LCOC value.

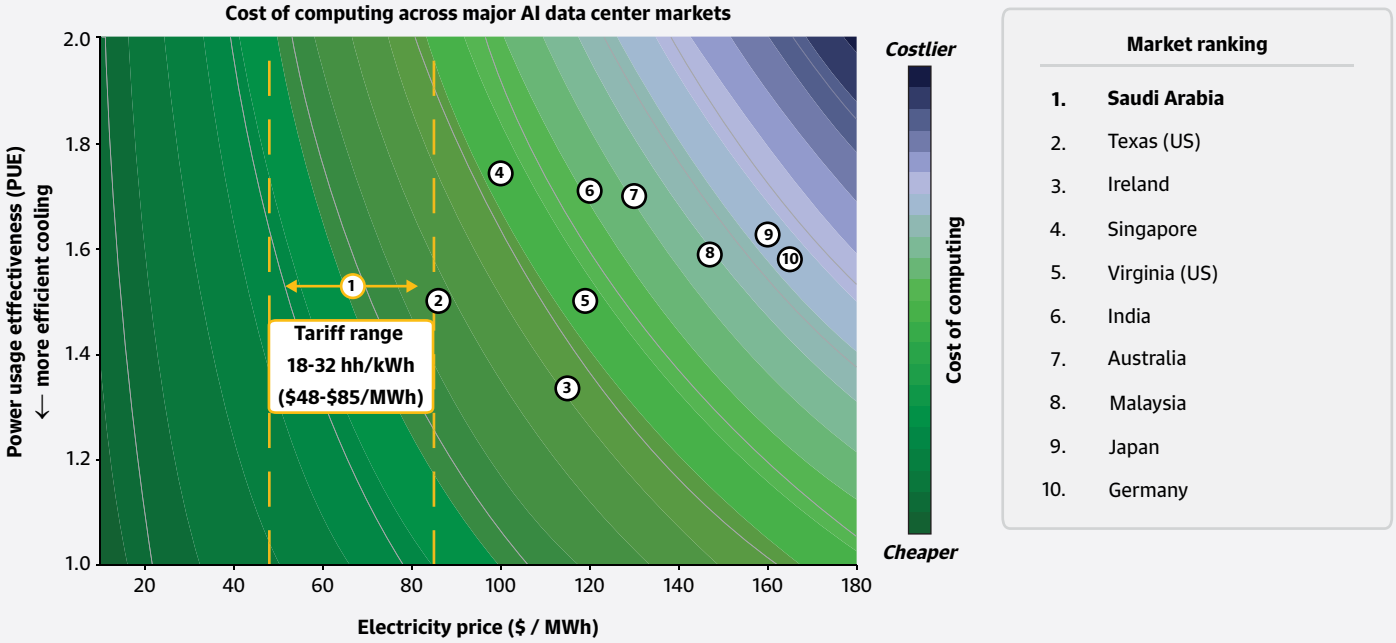
The interaction between PUE and electricity tariff is asymmetric: at lower tariffs, PUE exerts limited influence; the cost difference between a PUE of 1.5 and 2.0 widens from roughly 6% at $48/MWh to approximately 11% at $85/MWh. Across Saudi Arabia's considered tariff range ($48-$85/MWh) and at the baseline PUE of 1.53, LCOC spans 0.0321-0.0372 $/EFLOP. At 2024-2025 baseline comparator tariffs, Saudi Arabia leads the comparator set across its entire tariff range; even at the upper end (~$85/MWh, 0.0372 $/EFLOP) its cost only converges on the Texas market (0.0370 $/EFLOP), the next-lowest-cost comparator.

Each comparator is benchmarked at the most representative published tariff for large or data center loads, with the per-market source listed in Appendix B (Table B2). Where a jurisdiction has established a dedicated or forecast data center tariff, we apply it directly - Malaysia at its RP4 dedicated-AI rate, and the United States at the forecast 2026 tariffs for Texas and Virginia; elsewhere we apply the prevailing large-industrial rate, as no dedicated data center schedule is yet published. The comparison therefore reflects current best-available pricing for each market, with each value justified by its source in Table B2.

Recent regulatory developments show a clear upward trajectory in dedicated AI data center tariffs across several comparator markets, which is why these rates are incorporated directly where published. In Malaysia, Regulatory Period 4 (RP4, effective July 1, 2025) raised the average base tariff to 45.40 sen/kWh and placed data centers in a new ultra-high voltage rate class. Once voltage-based capacity and network charges are layered onto the base tariff, effective data center tariffs average about 60 sen/kWh, approximately $147/MWh, or nearly three times Saudi Arabia's cloud computing rate (Tenaga Nasional Berhad 2024; Onag 2025). In the United States, retail electricity prices rose 6.9% in 2025, more than double headline inflation, driven in part by surging data center demand (Goldman Sachs Research 2026). Virginia's State Corporation Commission approved a dedicated GS-5 large-load rate class for data centers (25 MW and above) in November 2025, effective January 2027 (Virginia SCC 2025). At forecast 2026 tariffs, Texas (~$86/MWh) stays close to Saudi Arabia's upper bound while Virginia (~$119/MWh) rises to fifth.

These dedicated and forecast tariffs are applied directly for Malaysia, Texas, and Virginia; the remaining markets are shown at prevailing industrial rates, since no dedicated data center schedule is yet published for them. As such schedules are introduced, those markets' compute costs would rise relative to Saudi Arabia's. Full country data, including tariffs, PUE values, and sources, are provided in Appendix B.

Figure 8. LCOC contour as a joint function of electricity price and PUE, with 10 comparator markets overlaid (the United States represented by its two leading data center markets, Texas and Virginia).



Note: Markets ranked 1-10 by LCOC (low to high). Saudi Arabia (rank 1, highlighted) occupies the lowest-cost region. (Full country data is in Appendix B.)

Source: Authors, using 451 Research from S&P Global Energy Horizons (2025) data.

From $/EFLOP to $/Token: Practical Interpretations and Insights

05

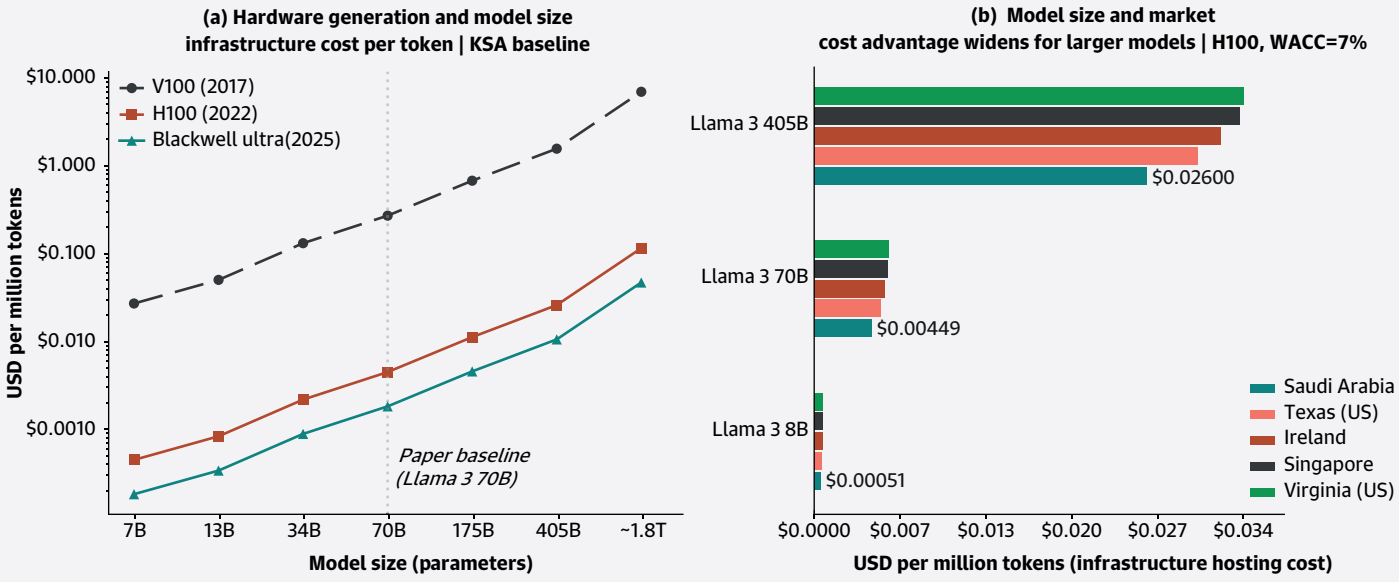
Building on the cost framework developed in Section 4, this section translates the LCOC into a metric directly relevant to operators, policymakers, and investors: the cost of generating a million output tokens for representative AI workloads. This conversion turns engineering parameters into an actionable signal for tariff design, capacity planning, and competitive positioning.

The LCOC expressed in dollars per EFLOP provides a rigorous basis for comparing infrastructure costs across locations and hardware configurations. To make this metric more accessible for policymakers and investors, we convert the LCOC to a cost per million tokens: the standard consumption unit in large language model inference. The conversion follows from the computational requirements of transformer-based models, where the number of floating-point operations per token is approximately twice the parameter count (Kaplan et al. 2020; Grattafiori et al. 2024). For a 70-billion-parameter model, specifically Llama 3 70B, an open-weight model benchmarked on H100-class hardware, the implied compute requirement is approximately 140 GFLOPs per token, yielding a conversion factor of approximately 7.14 billion tokens per EFLOP.

Applying this conversion to the baseline LCOC yields an infrastructure hosting cost of approximately $0.0045 per million tokens for AI inference workloads in Saudi Arabia. Figure 9 extends this across model sizes and GPU generations (panel a) and compares hosting costs across the five lowest-cost locations for three representative model sizes (panel b). The unit $/million tokens reflects infrastructure hosting cost only, derived from the LCOC using the Llama 3 70B inference profile (140 GFLOPs/token; Kaplan et al. 2020; Grattafiori et al. 2024).

Commercial API providers serving Llama 3 70B inference charged between $0.33 and $2.86 per million tokens in 2024 (Artificial Analysis 2025) depending on provider and tier, reflecting model development amortization, software infrastructure, reliability engineering, and commercial margins that sit above the infrastructure layer.

Figure 9. Infrastructure hosting cost per million tokens by AI hardware generation and model size (panel a) and across the five lowest-cost locations for three representative models (panel b).



Note: All costs reflect infrastructure hosting only.

Source: Authors, using 451 Research from S&P Global Energy Horizons (2025) data.

Table 2. LCOC sensitivity to key operational scenarios.

Scenario	$/EFLOP	$/M tokens
KSA baseline (H100, WACC=7%)	**0.0321**	**0.0045**
High utilization (k=95%)	0.0281	0.0039
Low utilization (k=50%)	0.0475	0.0067
Double the cloud computing tariff ($96/MWh)	0.0386	0.0054

Note: Full scenario analysis including AI hardware efficiency and CAPEX sensitivity in Appendix C.

Three findings stand out:

(1) As shown in Figure 9 panel (a), AI hardware efficiency has a far larger effect on token-level cost than electricity pricing: upgrading from V100 to H100 reduces compute cost by approximately 98%, while doubling the electricity tariff increases costs by only 20%. However, AI hardware upgrades entail substantially higher upfront capital expenditure, and the decision to adopt next-generation accelerators ultimately rests with individual operators rather than policymakers. From a policy perspective, the more directly actionable levers remain electricity tariff design and facility efficiency standards.

(2) As demonstrated in Table 2, the impact of underutilization is significant: token costs increase sharply below approximately 60%-70% utilization, ranging from $0.0039 per million tokens at 95% to $0.0067 at 50%.

(3) Saudi Arabia's infrastructure hosting cost of $0.0045 per million tokens sits well below commercial API pricing for equivalent models (Artificial Analysis 2025). This gap reflects the value-chain components that sit above the compute layer (including network and data transfer, software and inference optimization, model development amortization, reliability engineering, security, and commercial margins) and represents the broader digital value that energy-exporting economies can seek to progressively capture through sustained investment in AI infrastructure and services.

Integrating Environmental Impacts

06

We employ two environmental efficiency metrics, CUE and WUE, and integrate them into the LCOC framework to assess environmentally adjusted compute costs.

6.1 Carbon Efficiency and the CUE

CUE is defined as the ratio of carbon emissions from powering an entire data center to the electricity consumed by the IT equipment:

$$\text{CUE} = \frac{\text{Total } CO_2 \text{ Emissions (kg)}}{\text{IT Equipment Energy Use (kWh)}} \tag{3a}$$

The numerator captures all emissions from electricity generation, including cooling and auxiliary systems; the denominator reflects electricity used by the IT load. For Saudi Arabia, data center electricity demand is projected to reach 26.8 TWh by 2030 under the moderate scenario (2,000 MW IT capacity at 100% load factor; Alshehri et al. 2026). Consistent with the literature (Climatiq 2021; Rowaihy et al. 2025), we assume a current grid emission factor of 0.51 kg CO_2/kWh, yielding approximately 13.7 $MtCO_2$ of annual data center emissions. Assuming an average PUE of 1.53 (S&P Global 2025), the implied IT energy consumption is 17.5 TWh. Substituting these values into the CUE formula gives:

$$\text{CUE} = \frac{13.7 \cdot 10^9 \text{ kg } CO_2}{17.5 \cdot 10^9 \text{ kWh}} \approx 0.78 \text{ kg } CO_2/kWh. \tag{3b}$$

Under Saudi Arabia's goal of sourcing 50% of electricity from renewables and 50% from natural gas (MoE 2025), the effective grid emission factor would decline to roughly 0.3 kg CO_2/kWh, lowering the CUE to about 0.459 kg CO_2/kWh. This represents a reduction of approximately 41%. The current CUE calculation uses an average grid emission factor. Future work could test marginal and long-run system emission factors, because incremental data center load may be served by a different generation mix than the average grid.

6.2 Water Use and the WUE

WUE measures the volume of water consumed for cooling relative to the electricity delivered to IT equipment:

$$\text{WUE} = \frac{\text{Annual Water Use (L)}}{\text{IT Equipment Energy Use (kWh)}} \tag{4a}$$

The numerator represents annual facility water consumption; the denominator isolates IT-equipment electricity. WUE is particularly critical in water-scarce regions such as Saudi Arabia, which relies heavily on desalination (IEA 2025b; Alwazir and Dichter 2022). Cooling needs for hyperscale data centers can reach about 25 million liters of water per megawatt per year (Spindler, Hahn-Petersen, and Hosseini 2024); applied to an average IT load of 2,000 MW (Alshehri et al. 2026, moderate scenario), this implies total water use of approximately 50 billion liters, giving:

$$\text{WUE} = \frac{50 \cdot 10^9 \text{ L}}{17.5 \cdot 10^9 \text{ kWh}} \approx 2.85 \text{ L/kWh} \tag{4b}$$

Saudi Arabia is the world's largest producer of desalinated water, which is energy- and capital-intensive, and subsidies create a fiscal burden (Alwazir and Dichter 2022). WUE therefore represents both a local environmental constraint and a long-term economic externality for the digital economy. Construction-phase and supply-chain water are not included in this metric and could become material as deployment scales. Data center cooling also creates a circular dependency in arid economies: higher compute demand increases electricity consumption, which raises desalination energy requirements. The true energy footprint of AI infrastructure therefore extends beyond facility boundaries in ways not captured by PUE or WUE alone, but directly relevant to integrated energy system planning.

6.3 Environmental Valuation Scenarios: Shadow Carbon and Water Costs

To reflect environmental externalities, we adjust the LCOC by modifying the OPEX term. Both CUE and WUE already include PUE effects and are defined per unit of IT energy, so we incorporate them directly into the per-kW OPEX, multiplying by t (annual operating hours) for dimensional consistency:

$$\text{OPEX}^* = p \cdot \alpha \cdot k \cdot t + \lambda \cdot \text{CUE} \cdot k \cdot t + \mu \cdot \text{WUE} \cdot k \cdot t \quad (5)$$

where, p is the electricity price ($/MWh), α the power usage PUE, k the load factor, and t is the annual operating hours. The additional terms internalize environmental externalities: λ represents the shadow carbon price ($/kg CO_2) applied to the shadow carbon usage effectiveness, while μ is the water price ($/L) applied to the water usage effectiveness. Multiplying each by t ensures dimensional consistency with the other OPEX components, which are expressed on a per-kW-year basis. We then use this adjusted OPEX* in the standard LCOC formulation:

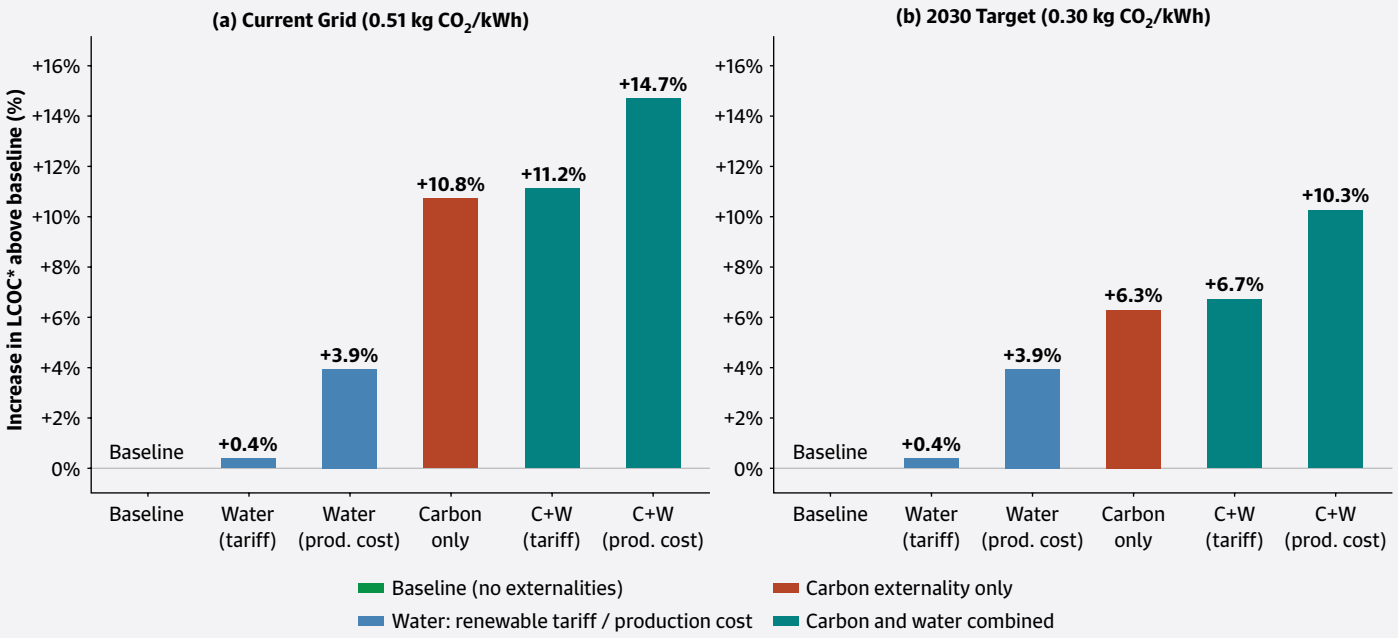
$$\text{LCOC}^* = \frac{c + \sum_{i=0}^{n-1} \frac{\text{OPEX}^* + dc}{(1+r)^i}}{\sum_{i=0}^{n-1} \frac{\gamma k t}{(1+r)^i}} \quad (6)$$

The baseline parameters for LCOC calculations are used here. Here, we state the additional assumptions needed to calculate the impacts of the CUE and WUE on the levelized cost. In this study, carbon pricing is valued at $50/t$CO_2$ (equivalent to $0.05/kg CO_2), a level consistent with many compliance instruments and social cost projections (World Bank 2025). Saudi Arabia's current carbon-market architecture is better represented as a voluntary or project-based crediting and offsetting framework, not as an economy-wide carbon price imposed on emitters. Therefore, the $50/t$CO_2$ value should be treated as a scenario parameter rather than an observed or mandated Saudi carbon price. It is used as a stylized carbon valuation parameter to test the sensitivity of AI infrastructure economics to future carbon constraints, or social-cost accounting.

For water, two cost assumptions are used: the renewable water tariff of $0.0005/L reflecting what data center operators actually pay under current industrial pricing (MODON 2025), and the production cost of $0.005/L reflecting what water actually costs to supply (consistent with Alwazir and Dichter (2022), who document that tariffs were maintained at 5% to 10% of actual production cost). The gap between the two figures, a factor of 10, represents the scale of Saudi Arabia's water subsidy regime, a cost currently borne by the state rather than operators but material to any integrated assessment of the digital economy's resource footprint. Future work could enhance the assumptions on the water prices, however, the current analysis provides an analytical approach to calculate cost implications of data center water consumption.

Figure 10 decomposes LCOC* by externality source under current and future grid conditions. At the current grid emission factor (0.51 kg CO_2/kWh), carbon pricing dominates: the CUE adjustment raises LCOC* by 10.8%, while water at the operator tariff ($0.0005/L) adds less than 1%. Water at production cost ($0.005/L) contributes an additional 3.9% but remains secondary to carbon. Combined, the production scenario raises LCOC* by 14.7% over baseline. Under the future grid target (0.30 kg CO_2/kWh), the carbon penalty falls to 6.3%, reducing the combined production adjustment to 10.3%. Full decomposition across all scenarios is in Appendix C. The results in this paper should not be interpreted to conclude that water is less important than carbon for policy, as the current model restricts the attention to specific exploratory scenarios and the contribution here is mainly methodological.

Figure 10. Decomposition of LCOC* above the unadjusted baseline by externality type and grid scenario.



Note: Each bar shows the % age increase in compute cost attributable to carbon pricing (orange) and water pricing (blue).

Source: Authors, using 451 Research from S&P Global Energy Horizons (2025) data.

Policy Implications and Conclusion

07

In this paper, we have provided an integrated policy and technoeconomic analysis of Saudi Arabia as a major energy-exporting economy and an emerging regional hub for AI-centric data center investment. The Kingdom has assembled a layered digital policy regime and developed an integrated framework with a national, phased roadmap for upstream infrastructure and downstream AI and data services. This roadmap also targets workforce development and an innovation ecosystem in data and AI – important for long-term operations and value creation.

To assess the long-term economics of AI data centers, we apply the LCOC framework and extend it to incorporate carbon and water externalities. The LCOC for Saudi Arabia, anchored to H100-class hardware, is estimated across the considered tariff range ($48-$85/MWh, with $48/MWh corresponding to the SERA 2025 cloud computing tariff and the upper bound representing an assumed ceiling for tariffs outside that regime) at 0.0321-0.0372 $/EFLOP – equivalent to $0.0045-$0.0052 per million tokens for Llama 3 70B inference. This positions Saudi Arabia as the lowest-cost location among the 10 comparator markets examined, reflecting the combined effect of competitive electricity pricing and facility-level efficiency broadly consistent with regional peers.

AI hardware efficiency is the dominant driver of long-run cost reduction, though generational upgrades entail higher capital expenditure and remain decisions for individual operators. From a policy perspective, the more actionable levers are electricity tariff design and facility efficiency standards. When environmental externalities are incorporated, carbon intensity emerges as the primary cost adjustment factor under the studied scenarios: at the current grid emission factor, full internalization of carbon and water costs raises LCOC* by up to 14.7%, falling to 10.3% under the 2030 renewable energy target.

Based on these findings, we propose five policy recommendations for Saudi Arabia and other resource-rich economies seeking to translate energy advantage into digital competitiveness.

Immediate (Years 1-2):

- **Pair competitive tariffs with minimum efficiency standards:** Maintaining cost-reflective data center tariffs while phasing in binding efficiency thresholds (progressive PUE targets and WUE benchmarks) creates a sustainable and investment-friendly environment without eroding the tariff advantage.
- **Prioritize high utilization in early deployment stages:** Because LCOC is highly sensitive to load factor, particularly below 60%-70% where fixed capital costs are spread

over limited compute output, workload aggregation and procurement structures that guarantee steady front-loaded utilization are an economic imperative, not merely an operational preference.

- **Standardize transparency and life-cycle accounting:** Mandatory disclosure of electricity use, carbon intensity, water consumption, and compute utilization rates (distinguishing training from inference workloads) improves regulatory oversight and enables rigorous project-level assessment, including supply-chain impacts.

Medium-Term (Years 3-5):

- **Accelerate renewable energy deployment:** Grid decarbonization is a lever for reducing environmental-adjusted compute costs, delivering a larger improvement than any facility-level efficiency measure. Dedicated renewable procurement frameworks for data center operators can accelerate this transition.
- **Link hub incentives to environmental performance:** Data center investment incentives conditioned on measurable environmental performance indicators (water use, waste heat recovery, and carbon intensity) align commercial expansion with national sustainability objectives.

Finally, our analysis has limitations that future research should address. The LCOC does not capture system-level effects such as grid flexibility, network constraints, or the opportunity cost of energy and water at full economic prices, adjustments that would increase the environmental cost estimates reported here. As more empirical data become available, future research can provide deeper insights into the Energy-AI Nexus and its implications for sustainable development in Saudi Arabia and other energy-exporting economies. Ultimately, sustained digital advantage requires aligning competitive tariffs, progressive infrastructure efficiency standards, and timely access to next-generation AI hardware.

References

Aldersey-Williams, J., and T. Rubert. 2019. "Levelised Cost of Energy – A Theoretical Justification and Critical Assessment." *Energy Policy* 124: 169–179. https://doi.org/10.1016/j.enpol.2018.10.004.

Alshehri, Khaled, Marwa Mahmoud AlFattani, Laila Bashmal, and Ghalia Alshmmari. 2026. "AI and Energy: The Future of Data Centers in Saudi Arabia." Riyadh: King Abdullah Petroleum Studies and Research Center (KAPSARC). https://www.kapsarc.org/our-offerings/publications/ai-and-energy-the-future-of-data-centers-in-saudi-arabia/.

Alwazir, Albara'a, and Jacob Dichter. 2022. *Saudi Arabia's Water Sector*. Economic Brief. U.S.-Saudi Business Council. February. https://ussaudi.org/wp-content/uploads/2022/02/Water-2022-Economic-Brief.pdf.

Artificial Analysis. 2025. "Llama 3 Instruct 70B: API Provider Benchmarking & Analysis." https://artificialanalysis.ai/models/llama-3-instruct-70b/providers.

Aterio. 2025. "Power Capacity Estimation ML Model." https://knowledge.aterio.io/data-products/us-data-centers/our-modeling-approach/power-capacity-estimation-ml-model.

Australian Energy Market Operator (AEMO). 2025. *Quarterly Energy Dynamics Q4 2024*. https://www.aemo.com.au/-/media/files/major-publications/qed/2024/qed-q4-2024.pdf?la=en.

Australian Energy Regulator (AER). 2024. *State of the Energy Market 2024*. https://www.aer.gov.au/system/files/2025-07/State%20of%20the%20energy%20market%202024.pdf.

Bundesverband der Energie- und Wasserwirtschaft (BDEW). 2024. *BDEW – Strompreisanalyse: December 2024*. https://www.bdew.de/media/documents/BDEW-Strompreisanalyse_12-2024_Q796OxD.pdf.

Calvino, Flavio, Antoine Dechezleprêtre, and Daniel Haerle. 2025. "How Can the Digital Transformation Affect the Net-Zero Transition? A Conceptual Framework for the 'Twin' Transition." OECD Science, Technology, and Industry Working Papers 2025/17. https://doi.org/10.1787/4ee71f6d-en.

Central Electricity Regulatory Commission (CERC). 2024. *Annual Report 2023–24*. New Delhi: CERC. https://www.cercind.gov.in/2024/market_monitoring/Annual%20Report%202023-24.pdf.

Climatiq. 2021. "Electricity Supplied from Grid." [Saudi Arabia: 2020 Grid Mix]. https://www.climatiq.io/data/emission-factor/d642f6c8-32d7-4749-ac48-6463f30e0ae7.

Communications, Space & Technology Commission (CST). 2022. *Telecommunications and Information Technology Act*. Issued under the Council of Ministers Resolution No. (592), dated 01/11/1443 AH. Riyadh: Government of Saudi Arabia. https://share.google/K0NZZw4cRYYeml0m2.

Communications, Space & Technology Commission (CST). 2023. *Data Center Services Regulations*. Riyadh: Government of Saudi Arabia. RS40. First Version. https://www.cst.gov.sa/en/regulations-and-licenses/regulations/Document-1546.

Communications, Space & Technology Commission (CST). 2025. *Global AI Hub Law (Draft)*. Riyadh: Government of Saudi Arabia. https://www.cst.gov.sa/en/media-center/news/N2025041401.

Digital Government Authority (DGA). 2023. *Digital Government Strategy*. Riyadh: Government of Saudi Arabia. https://my.gov.sa/en/content/digital-strategy.

Donnellan, Douglas, Andy Lawrence, Daniel Bizo, Peter Judge, John O'Brien, Jacqueline Davis, Max Smolaks, Jabari Williams-George, and Rose Weinschenk. 2024. *Uptime Institute Global Data Center Survey 2024*. Uptime Intelligence. https://datacenter.uptimeinstitute.com/rs/711-RIA-145/images/2024.GlobalDataCenterSurvey.Report.pdf.

Energy Information Administration (EIA). 2024. "Electric Power Monthly, Table 5.6: Average Retail Price of Electricity." https://www.eia.gov/opendata/.

Energy Market Authority (EMA). 2024. *Singapore Energy Statistics 2024*. https://www.ema.gov.sg/resources/singapore-energy-statistics.

U.S. Energy Information Administration (EIA). 2026. Electric Power Monthly, Table 5.6.A: Average Price of Electricity to Ultimate Customers by End-Use Sector, by State. Washington, DC: U.S. EIA. https://www.eia.gov/electricity/monthly/ (accessed 2026).

Eurostat. 2024. *Electricity Prices for Non-Household Consumers – Second Half 2023*. Luxembourg: European Commission. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Electricity_price_statistics.

Federation of Electric Power Companies of Japan (FEPC). 2024. *The Electric Power Industry in Japan 2024*. Tokyo: FEPC.

Gaglio, Cyrielle, Erika Kraemer-Mbula, and Edward Lorenz. 2022. "The Effects of Digital Transformation on Innovation and Productivity: Firm-Level Evidence of South African Manufacturing Micro and Small Enterprises." *Technological Forecasting and Social Change* 182: 121785. https://doi.org/10.1016/j.techfore.2022.121785.

Google. 2025. "Power Usage Effectiveness." Google Data Centers. https://www.google.com/about/datacenters/efficiency/.

Goldman Sachs Research. 2026. "The Macroeconomic Spillovers from AI Electricity Demand." February 11. https://www.gspublishing.com/content/research/en/reports/2026/02/11/cefdda9f-3b7c-4a53-9950-dfcfa3c6bbc2.html.

Grattafiori, Aaron, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, et al. 2024. "The Llama 3 Herd of Models." arXiv preprint arXiv:2407.21783. https://arxiv.org/abs/2407.21783.

Growth Lab. 2024. *Atlas of Economic Complexity*. Harvard Kennedy School. https://atlas.hks.harvard.edu.

Huang, Pei, Benedetta Copertaro, Xingxing Zhang, Jingchun Shen, Isabelle Löfgren, Mats Rönnelid, Jan Fahlen, Dan Andersson, Mikael Svanfeldt. 2020. "A Review of Data Centers as Prosumers in District Energy Systems: Renewable Energy Integration and Waste Heat Reuse for District Heating." *Applied Energy* 258: 114109. https://doi.org/10.1016/j.apenergy.2019.114109.

Innovation for Cool Earth Forum (ICEF). 2025. *Sustainable Data Centers Roadmap*. October. https://www.icef.go.jp/roadmap/.

International Energy Agency (IEA). 2020. *Projected Costs of Generating Electricity 2020 Edition*. Paris: International Energy Agency. https://www.iea.org/reports/projected-costs-of-generating-electricity-2020.

International Energy Agency (IEA). 2025a. *Energy and AI*. Paris: International Energy Agency. https://www.iea.org/reports/energy-and-ai/.

International Energy Agency (IEA). 2025b. *The Future of Electricity in the Middle East and North Africa*. Paris: International Energy Agency. https://www.iea.org/reports/the-future-of-electricity-in-the-middle-east-and-north-africa.

Jamison, Stephanie, Sanjay Podder, Adam Burden, Bhaskhar Ghosh, Senthil Ramani, Shalabh Kumar Singh, and Matthew Robinson. 2025. *Powering Sustainable AI: Balancing Growth with Environmental Responsibility*. Accenture. https://www.accenture.com/content/dam/accenture/final/corporate/corporate-initiatives/sustainability/document/Powering-Sustainable-AI-CanvasV10.pdf.

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. "Scaling Laws for Neural Language Models." arXiv preprint arXiv:2001.08361. https://doi.org/10.48550/arXiv.2001.08361.

Koronen, C., M. Åhman, and L. J. Nilsson. 2020. "Data Centres in Future European Energy Systems – Energy Efficiency, Integration and Policy." *Energy Efficiency* 13 (1): 129–144. https://doi.org/10.1007/s12053-019-09833-8.

Kristiansen Nøland, Jonas, Martin Hjelmeland, and Magnus Korpas. 2024. "Will Energy-Hungry AI Create a Baseload Power Demand Boom?" *IEEE Access* 12: 110353–110360. https://doi.org/10.1109/ACCESS.2024.3440217.

Li, Guozhu, Zixuan Sun, Qingqin Wang, Shuai Wang, Kailiang Huang, Naini Zhao, Yanqiang Di, Xudong Zhao, and Zishang Zhu. 2023. "China's Green Data Center Development: Policies and Carbon Reduction Technology Path." *Environmental Research* 231: 116248. https://doi.org/10.1016/j.envres.2023.116248.

Malik, Saf. 2025. "What is Driving Saudi Arabia's $21 bn Data Centre Investment Surge?" *Capacity Media*, March 10. https://www.capacitymedia.com/article/2eir2o1uu11xqfdsdwrgg/news/what-is-driving-saudi-arabias-21bn-data-centre-investment-surge.

Mikayilov, Jeyhun I., and Abdullah Darandary. 2024. "Modelling and Projecting Regional Electricity Demand for Saudi Arabia." *OPEC Energy Review* 48: 261–277. https://doi.org/10.1111/opec.12312.

Ministry of Communications and Information Technology (MCIT). 2019. *KSA Cloud First Policy*. Riyadh: Government of Saudi Arabia. https://mcit.gov.sa//sites/default/files/cloud_policy_en.pdf.

Ministry of Energy (MoE). 2025. "Saudi Arabia Aims for 50% Power Generation from Renewables by 2030." Riyadh: Government of Saudi Arabia. https://www.arabnews.com/node/1795406/saudi-arabia.

MODON. 2025. "Cost of Industry > Water" https://modon.gov.sa/en/Systems/IndustryCost/Pages/Water.aspx.

Montgomery, Blake. 2024. "Datacenters Meet Resistance Over Environmental Concerns as AI Boom Spreads in Latin America." *The Guardian*, November 10. https://www.theguardian.com/technology/2025/nov/10/data-centers-latin-america.

National Cybersecurity Authority (NCA). 2020. *National Cybersecurity Strategy*. Riyadh: Government of Saudi Arabia. https://cdn.nca.gov.sa/ar/national_cybersecurity_strategy-en.pdf.

OECD. 2022. *Measuring the Environmental Impacts of Artificial Intelligence Compute and Applications: The AI Footprint*. OECD Digital Economy Papers No. 341. https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/11/measuring-the-environmental-impacts-of-artificial-intelligence-compute-and-applications_3dddded5/7babf571-en.pdf.

Onag, Gigi. 2025. "Electricity Tariffs Spike Puts Malaysia's Data Center Hub Status at Risk." Light Reading, July 3. https://www.lightreading.com/data-centers/electricity-tariffs-spike-puts-malaysia-s-data-center-hub-status-at-risk.

Oró, Eduard, Victor Depoorter, Albert Garcia, and Jaume Salom. 2015. "Energy Efficiency and Renewable Energy Integration in Data Centres: Strategies and Modelling Review." *Renewable and Sustainable Energy Reviews* 42: 429–445. https://doi.org/10.1016/j.rser.2014.10.035.

PBS News. 2024. "Google to Pause Plans for Big Data Center in Chile Over Water Worries." September 17. https://to.pbs.org/4nksF0T

Rong, Huigui, Haomin Zhang, Sheng Xiao, Canbing Li, and Chunhua Hu. 2016. "Optimizing Energy Consumption for Data Centers." *Renewable and Sustainable Energy Reviews* 58: 674–691. https://doi.org/10.1016/j.rser.2015.12.283.

Rowaihy, Feras, Ali Hamieh, Naser Odeh, Mohamad Hejazi, Mohammed Al-Juaied, Abdulkader M. Afifi, and Hussein Hoteit. 2025. "Decarbonizing Saudi Arabia Energy and Industrial Sectors: Assessment of Carbon Capture Cost." *Carbon Capture Science & Technology* 14: 100375. https://doi.org/10.1016/j.ccst.2025.100375.

Salvator, Dave, Rohil Bhargava, and Kirthi Devleker. 2025. "NVIDIA Blackwell Ultra for the Era of AI reasoning." NVIDIA Technical Blog, March 19. https://developer.nvidia.com/blog/nvidia-blackwell-ultra-for-the-era-of-ai-reasoning/.

S&P Global. 2025. *Data Center Intelligence: Global Capacity & Pipeline Database*.

Saudi Data & AI Authority (SDAIA). 2020. "National Strategy for Data and Artificial Intelligence (NSDAI)." Riyadh: Government of Saudi Arabia. https://sdaia.gov.sa/en/SDAIA/SdaiaStrategies/Pages/NationalStrategyForDataAndAI.aspx.

Saudi Data & AI Authority (SDAIA). 2021. *Personal Data Protection Law (PDPL)*. Royal Decree No. M/19, 17 September 2021. Amended March 2023. Riyadh: Government of Saudi Arabia. https://sdaia.gov.sa/en/SDAIA/about/Documents/Personal%20Data%20English%20V2-23April2023-%20Reviewed-.pdf.

Saudi Electricity Regulatory Authority (SERA). 2025. "Consumption Tariff." https://www.sera.gov.sa/en/consumer/electric-tariff/electric-tariff-categories/consumption-tariff.

Soares, Ian V., Masaru Yarime, and Magdalena Klemun. 2024. "Balancing the Trade-Off Between Data Center Development and its Environmental Impacts: A Comparative Analysis of Data Center Policymaking in Singapore, Netherlands, Ireland, Germany, USA, and the UK." *Environmental Science & Policy* 157: 103769. https://doi.org/10.1016/j.envsci.2024.103769.

Sovacool, B. K., C. G. Monyei, and P. Upham. 2022. "Making the Internet Globally Sustainable: Technical and Policy Options for Improved Energy Management, Governance and Community Acceptance of Nordic Datacenters." *Renewable and Sustainable Energy Reviews* 154: 111793. https://doi.org/10.1016/j.rser.2021.111793.

Spindler, Wesley, Luna A. Hahn-Petersen, and Sadaf Hosseini. 2024. "Why Circular Water Solutions are Key to Sustainable Data Centres." World Economic Forum, November 7. https://www.weforum.org/stories/2024/11/circular-water-solutions-sustainable-data-centres/.

Tenaga Nasional Berhad. 2024. "Regulatory Period 4 (RP4) Base Electricity Tariff 2025-2027." Filing with the Energy Commission, Malaysia. https://www.tnb.com.my/.

TONOMUS. 2025. "TONOMUS Launches NEOM's First Digital Communications Facility." July 18. https://www.neom.com/en-us/newsroom/tonomus-communication-facility.

Ueckerdt, Falko, Lion Hirth, Gunnar Luderer, and Ottmar Edenhofer. 2013. "System LCOE: What are the Costs of Variable Renewables?" *Energy* 63: 61–75. http://dx.doi.org/10.1016/j.energy.2013.10.072.

Upham, Paul, Benjamin K. Sovacool, and Chukwuka G. Monyei. 2023. "Imaginaries on Ice: Sociotechnical Futures of Data Centre Development in Norway and Iceland." *Environment and Planning E: Nature and Space* 6 (3): 1905–1922. https://doi.org/10.1177/25148486221126619.

Van Zandt, Brooke. 2023. "A Decade of Greener Computing Blooms Inside NREL's Data Center." National Laboratory of the Rockies, August 2. https://www.nlr.gov/news/detail/program/2023/a-decade-of-greener-computing-blooms-inside-nrels-data-center.

Virginia State Corporation Commission. 2025. Final Order, Case No. PUR-2025-00058 — Application of Virginia Electric and Power Company, 2025 Biennial Review (GS-5 large-load rate class for customers ≥25 MW and ≥75% load factor, effective January 1, 2027; final order issued November 25, 2025). Richmond, VA.

Wang, Yongzhen, Jiayu Lin, Yibo Han, Kai Han, Juntao Han, Te Han, Yiming Wei. 2025. "Comprehensive Evaluation of All-Element Flexibility Resources in Data Centers: Considering Synergistic Benefits of Computing, Electricity, and Heat." *Applied Energy* 399: 126442. https://doi.org/10.1016/j.apenergy.2025.126442.

World Bank. 2025. *State and Trends of Carbon Pricing 2025*. Washington, DC: World Bank. https://openknowledge.worldbank.org/bitstreams/152de0c2-e2be-49d6-aec1-3be8ebad4f74/download.

Appendix

Appendix A. Nomenclature

- α PUE, ratio of total facility energy to IT load energy (unitless).
- γ Computing efficiency measured as delivered performance in PFLOPs per kW (PFLOPs/kW).
- λ Carbon price applied in the CUE-based adjustment (\$/kg CO_2).
- μ Water price applied in the WUE-based adjustment (\$/L).
- *c* Overnight construction cost (capital expenditure, CAPEX) in \$/kW.
- *d* Annual non-energy operating expenditure (OPEX) as a fraction of CAPEX (%).
- *k* Load factor (utilization rate of IT equipment), ranging from 0 to 1 (unitless).
- *n* Project lifetime (years).
- *p* Average electricity price (\$/MWh).
- *r* Weighted Average Cost of Capital (WACC, %).
- *t* Average operating hours per year, typically 8,760 h.
- CUE Carbon usage effectiveness, emissions intensity per kWh of IT electricity use ($kgCO_2$/kWh).
- LCOC Levelized cost of computing, life-cycle cost per unit of compute (\$/PFLOP or \$/EFLOP).
- LCOC* Environmental-adjusted LCOC, incorporating carbon and water costs (\$/PFLOP or \$/EFLOP).
- WUE Water usage effectiveness, water consumption per kWh of IT electricity use (L/kWh).

Appendix B: AI Hardware Benchmarks and International Cost Comparison

Table B1. Representative computing efficiency for various AI hardware.

Provider	Platform	Year released	γ(PFLOPs/kW)
Intel	GPU Flex 140	2022	0.107
	GPU Flex 170	2022	0.107
	GPU Max 1100	2022	0.048
	GPU Max 1550	2023	0.049
AMD	Instinct MI300X	2023	0.209
Cerebras	CS-3	2024	5.430
NVIDIA	Tesla T4	2018	0.116
	Tesla V100	2017	0.052
	DGX A100	2022	0.769
	DGX H100	2022	**3.137**
	DGX B200	2024	5.035
	GB200 NVL72	2024	6.000
	Blackwell Ultra	2025	**~7.710**

Notes: γ= computing efficiency in PFLOPs per kilowatt of IT power. Orange rows indicate hardware used in the paper's baseline (DGX H100) and next-generation benchmark (Blackwell Ultra).

Source: Authors' compilation; Kristiansen, Hjelmeland, and Korpas (2024).

Table B2. International cost comparison of the levelized cost of AI computing.

	Market	Electricity price ($/MWh)	PUE	LCOC ($/EFLOP)	Cost ($/M tokens)
1	Saudi Arabia	48-85	1.530	0.0321-0.0372	0.0045-0.0052
2	Texas (US)	86	1.502	0.0370	0.0052
3	Ireland	115	1.336	0.0392	0.0055
4	Singapore	100	1.744	0.0410	0.0057
5	Virginia (US)	119	1.502	0.0414	0.0058
6	India	120	1.711	0.0438	0.0061
7	Australia	130	1.701	0.0452	0.0063
8	Malaysia	147	1.590	0.0463	0.0065
9	Japan	160	1.628	0.0487	0.0068
10	Germany	165	1.581	0.0487	0.0068

Note: Tariffs reflect the most representative published large-load or data center rate for each market. Texas and Virginia use forecast 2026 data center tariffs (EIA 2026; Virginia SCC 2025); Malaysia uses its RP4 dedicated-AI tariff; Saudi Arabia uses the SERA 2025 cloud computing tariff; all other markets use prevailing large-industrial rates.

Sources: SERA (2025); Tenaga Nasional Berhad (2024); EIA (2024); Eurostat (2024); EMA (2024); BDEW (2024); FEPC (2024); AER (2024); AEMO (2025); CERC (2024); S&P Global (2025).

Appendix C: Full LCOC Scenario Analysis

Table C1. LCOC scenario analysis.

Scenario	LCOC ($/EFLOP)	$/M tokens
KSA baseline (H100, WACC=7%)	**0.0321**	**0.0045**
High utilization (k = 95%)	0.0281	0.0039
Low utilization (k = 50%)	0.0475	0.0067
Low CAPEX ($8,000/kW)	0.0270	0.0038
High CAPEX ($15,000/kW)	0.0449	0.0063
Double the cloud-computing tariff ($96/MWh)	0.0386	0.0054
Legacy hardware – V100	1.9367	0.2711
Next-gen – Blackwell Ultra	0.0131	0.0018

Table C2. LCOC* decomposition by externality scenario.

Scenario	OPEX*	LCOC*	ΔOPEX%	ΔLCOC%
Baseline (no env. adj.)	**1,515**	**0.0321**	—	—
WUE only – tariff ($0.0005/L)	1,525	0.0322	+0.7%	+0.4%
WUE only – production cost ($0.0050/L)	1,615	0.0334	+6.6%	+3.9%
CUE only – current grid	1,788	0.0356	+18.1%	+10.8%
WUE + CUE – current, tariff	1,798	0.0357	+18.7%	+11.2%
WUE + CUE – current, production cost	1,888	0.0368	+24.7%	+14.7%
CUE only – 2030 grid	1,676	0.0341	+10.6%	+6.3%
WUE + CUE – 2030, tariff	1,686	0.0343	+11.3%	+6.7%
WUE + CUE – 2030, production cost	1,776	0.0354	+17.2%	+10.3%

Notes: Baseline LCOC = 0.0321 $/EFLOP (OPEX* = $1,515/kW). OPEX* = annual operating cost per kW IT load ($/kW/year). Carbon price: $50/t$CO_2$. Water tariff: $0.0005/L (MODON 2025 renewable water rate). Water production cost: $0.0050/L (derived from Alwazir and Dichter 2022: consumer tariff = 10% of production cost). WUE = 2.854 L/kWh (Spindler, Hahn-Petersen, and Hosseini 2024). Current grid EF: 0.51 kg CO_2/kWh; 2030 target: 0.30 kg CO_2/kWh.

Appendix D: Assessment of Saudi Arabia's Current Policies for the Energy-AI Nexus

Saudi Arabia's energy system remains heavily dependent on fossil fuels, with crude oil and natural gas dominating electricity generation. Under Vision 2030 and the National Renewable Energy Program, the Kingdom aims to increase the share of renewables to 50% of electricity generation by 2030 while expanding natural gas capacity to replace oil-fired generation. Longer-term climate commitments include a 2060 net-zero target under the Saudi Green Initiative and the Circular Carbon Economy framework. These energy system objectives form the structural context within which digital infrastructure expansion needs to be evaluated.

Hydrocarbon rents continue to account for a substantial share of Saudi Arabia's GDP, reinforcing the strategic imperative to diversify economic activity. Within Vision 2030, digital infrastructure is positioned as a foundational capability for sectoral transformation rather than a standalone ICT initiative. The Kingdom has therefore pursued an accelerated expansion of cloud and data center infrastructure, supported by a layered policy architecture encompassing digital government strategy, data protection law, cloud regulation, and cybersecurity frameworks. Together, these measures aim to create both demand and institutional readiness for large-scale computing deployment.

Saudi Arabia's digital and AI ecosystem is governed through a centralized institutional architecture that distributes responsibilities across ministries and specialized authorities. Core ministries oversee ICT strategy, investment promotion, industrial development, and energy policy, while sector-specific agencies regulate telecommunications, cybersecurity, digital government, and data governance. This structure reflects a deliberate attempt to coordinate digital transformation across economic and regulatory domains.

In recent years, the Kingdom has developed an increasingly comprehensive legal and regulatory framework to support digital infrastructure expansion. These measures signal a shift from ad hoc ICT development toward a structured governance model aimed

Table D1. Core ministries and agencies involved in Saudi Arabia's Energy-AI Nexus.

Entity	Role	Key relevant Nexus pillar
Ministry of Energy (MoE)	Oversees national energy policy across oil, gas, electricity, and renewables; supports digital expansion through tariff design and infrastructure planning.	Integrated Governance and Planning
Ministry of Communications and Information Technology (MCIT)	Leads national ICT policy and strategy to expand digital infrastructure.	Integrated Governance and Planning
Ministry of Investment	Designs foreign direct investment policy and incentives to attract technology and AI capital.	Resource and Capability Development
Ministry of Industry and Mineral Resources (MIMR)	Oversees industrial policy and promotes strategic investment in technology sectors.	Resource and Capability Development
Saudi Data & AI Authority (SDAIA)	National authority for data and AI governance.	Transparency and Accountability
Digital Government Authority (DGA)	Regulates digital government services and cloud adoption in public entities.	Adaptive Policy and Regulatory Framework
Communications, Space & Technology Commission (CST)	Regulates telecommunications, spectrum, and technology markets.	Adaptive Policy and Regulatory Framework
National Cybersecurity Authority (NCA)	Oversees national cybersecurity policy and compliance.	Transparency and Accountability
Research, Development & Innovation Authority (RDIA)	Funds and coordinates research and innovation in AI and emerging technologies.	Resource and Capability Development

Note: Pillar colors correspond to the four-pillar framework in Figure 5. Entities may contribute to multiple pillars; primary nexus role shown.

at positioning Saudi Arabia as a regional digital hub. The effectiveness of this architecture ultimately depends on how well digital policy is integrated with energy system planning and infrastructure constraints. Table D1 summarizes the principal digital and AI policies in Saudi Arabia and maps them to the four pillars of the Energy-AI Nexus framework.

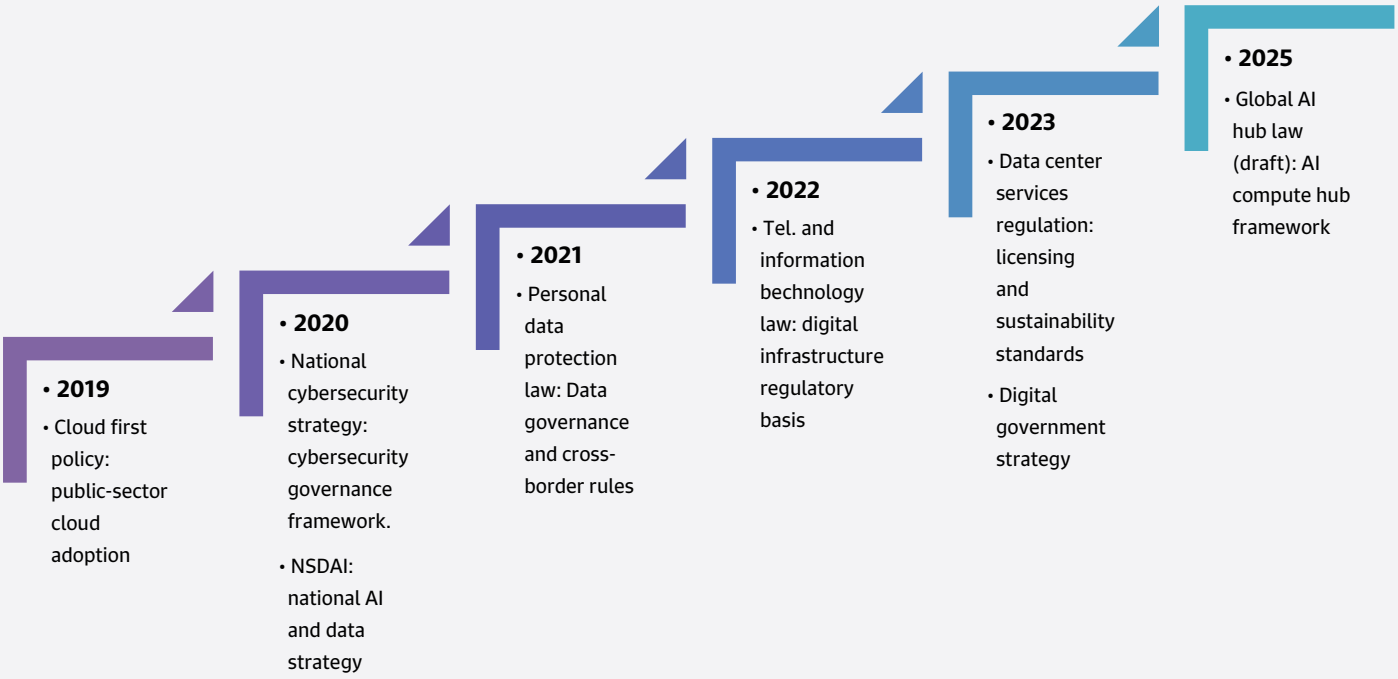
Since 2019, Saudi Arabia has progressively developed a layered regulatory framework to support digital infrastructure expansion. MCIT's Cloud First Policy (2019) initiated this transition by mandating cloud adoption across public agencies and accelerating migration from legacy IT systems to scalable digital platforms (MCIT 2019).

In 2020, two foundational strategies reinforced this trajectory. The National Cybersecurity Strategy, led by the NCA, established a structured framework for risk management, incident response, and information security, embedding cybersecurity within national development priorities (NCA 2020). In parallel, the NSDAI, led by the SDAIA, positioned data and AI as central pillars of economic diversification, outlining priorities in skills development, regulatory reform, investment attraction, and innovation (SDAIA 2020).

Subsequent legislation consolidated the regulatory architecture. The Personal Data Protection Law (2021, amended 2023) formalized data governance and cross-border transfer rules under the SDAIA's supervision (SDAIA 2021). The Telecommunications and Information Technology Act (2022) expanded regulatory oversight of digital platforms and telecommunications markets under the CST (2022).

In 2023, the Data Center Services Regulations introduced licensing and operational standards for data center providers, strengthening investor confidence and service quality. The DGA further advanced public-sector digital transformation with targets toward 2030 (DGA 2023).

Figure D1. Timeline of major digital, AI, and data governance policies in Saudi Arabia (2018-2025) (sequential development of these key national efforts).



Source: Authors' compilation.

Most recently, the proposed Global AI Hub Law (2025, draft) seeks to institutionalize Saudi Arabia's ambition to serve as a regional AI compute hub. The draft framework introduces specialized hub structures – Private, Extended, and Virtual – allowing foreign entities to host data and services under defined jurisdictional arrangements (CST 2025). By combining sovereign data governance with investment flexibility, the proposed law aims to enhance digital sovereignty, attract capital, and leverage Saudi Arabia's geographic and energy advantages.

We assess Saudi Arabia's digital governance architecture using the Energy-AI Nexus framework introduced above. Overall, the Kingdom demonstrates strong institutional coherence, strategic ambition, and sustained capability-building efforts. As digital infrastructure scales, the next phase of policy development will increasingly require deeper integration with energy system planning and environmental performance metrics, as reflected in the following dimensions:

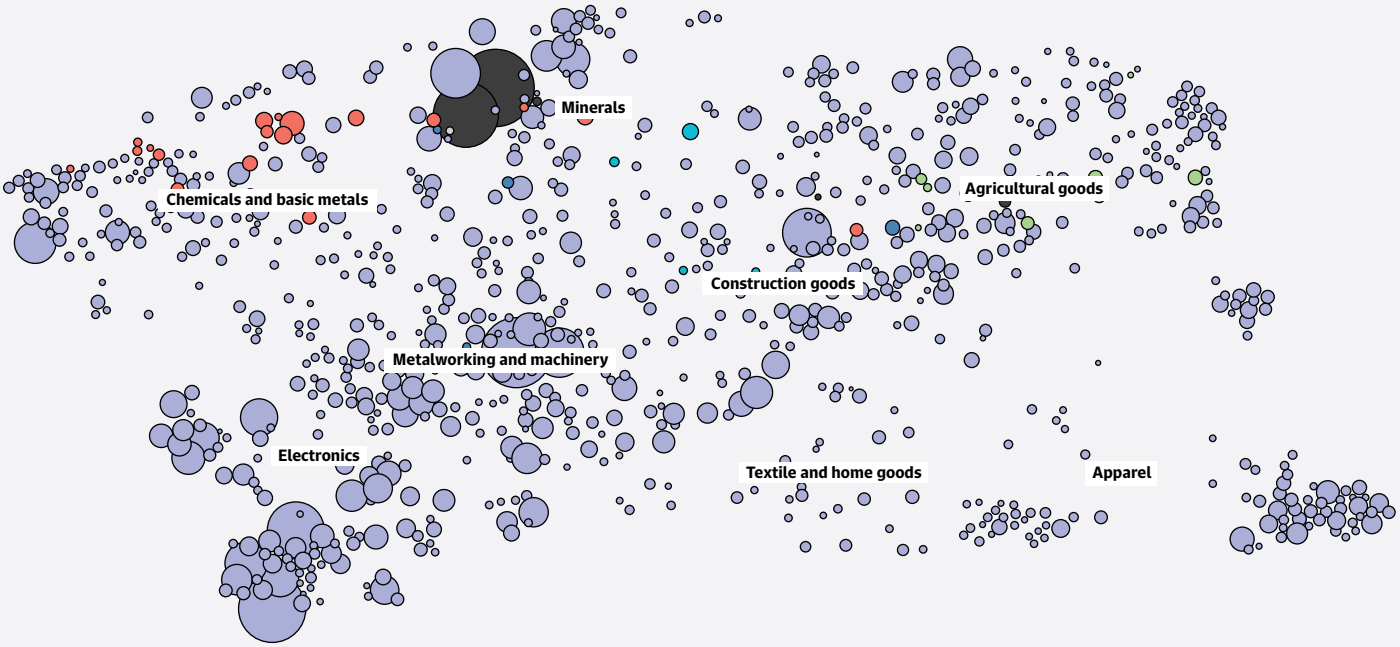
- **Integrated Planning and Governance:** The MoE plays a foundational role through its oversight of transmission and distribution infrastructure, renewable energy investment, and electricity tariff design, each of which directly shapes the economic parameters driving data center costs. Saudi Arabia's governance architecture reflects a centralized and coordinated approach to digital transformation. The establishment of SDAIA and the alignment of digital, industrial, and investment strategies provide a structured national roadmap for infrastructure expansion and AI ecosystem development. The proposed Global AI Hub framework further reinforces this integrated model by positioning the Kingdom as a regional compute platform through specialized hub structures and cross-border arrangements. The integration of economic, technological, and investment strategies represents a clear institutional strength. As data center capacity expands, continued coordination between digital authorities and energy planners will help ensure that compute growth aligns with grid modernization and renewable deployment pathways.

- **Adaptive Policy and Regulatory Framework:** Saudi Arabia has established a coherent and investment-oriented regulatory environment covering data protection, cybersecurity, telecommunications oversight, and licensing of data center services. These measures reduce legal uncertainty and provide a stable foundation for long-term capital deployment. The Data Center Services Regulations and related frameworks enhance operational clarity and support quality standards for infrastructure providers. As AI-centric facilities scale, regulatory refinement may further incorporate performance-based mechanisms (such as efficiency benchmarks and environmental reporting standards) to complement existing governance structures. This evolution would build on the Kingdom's current framework rather than require a structural overhaul. The Data Center Services Regulations of the CST (2023) and the proposed Global AI Hub Law together establish the licensing framework, jurisdiction rules, and sustainability reporting requirements for data center operators, representing a significant step toward adaptive regulation.

- **Resource and Capability Development:** Capability development represents one of the strongest pillars of Saudi Arabia's digital strategy. National initiatives prioritize workforce training, AI specialization, research and innovation capacity, and ecosystem development. Investment strategies aim to attract multinational firms while increasing domestic participation and knowledge transfer. Supply-chain development remains a strategic priority. As illustrated in Figure D2, Saudi Arabia's revealed comparative advantage is concentrated in minerals, chemicals, and basic metals, while advanced electronics and machinery remain limited. Ongoing industrial and localization strategies are designed to address these gaps over time. Strengthening domestic manufacturing capabilities in critical digital infrastructure components will further enhance resilience and reduce exposure to global supply constraints, reinforcing the Kingdom's ambition to serve as a regional AI hub. The MIMR, through the National Industrial Strategy, and RDIA contribute to this dimension by targeting supply-chain development and AI research capacity, respectively.

  Transparency and Accountability: Institutional foundations for transparency and accountability are well developed in the domains of data governance and cybersecurity. Regulatory supervision, compliance monitoring, and structured oversight mechanisms are clearly articulated within the digital policy architecture. As digital infrastructure expands, the incorporation of standardized environmental reporting, covering electricity use, carbon intensity, and water consumption, would further strengthen the Kingdom's governance model. Embedding life-cycle assessment approaches and measurable sustainability indicators within digital policy instruments would reinforce alignment between rapid infrastructure growth and long-term environmental objectives. These measures would complement existing national sustainability initiatives and enhance policy credibility in global markets.

Figure D2. Saudi Arabia in the product space (2023).



Note: Each node represents a product traded internationally; edges connect products that are frequently co-exported by countries, indicating shared productive capabilities (skills, infrastructure, institutions). Node size reflects the product's share of total world trade. Node color denotes product category (sector). Filled nodes indicate products in which Saudi Arabia has revealed a comparative advantage (exporting more than the product's share in total world trade) while unfilled nodes represent products outside Saudi Arabia's current export specialization. The position of the country within the network reveals nearby diversification opportunities: products closer to existing strengths typically require fewer new capabilities to develop competitively.

Source: Growth Lab (2024).

About the Authors



Khaled Alshehri

Dr. Khaled Alshehri is a Fellow at KAPSARC, with research focused on integrating digital technologies into the power sector to optimize grid operations, enhance economic efficiency, and support energy transition. Drawing on extensive experience across government, academia, and industry, Dr. Alshehri leverages advanced tools from game theory, control, and optimization to address complex challenges in energy systems. In 2024, Dr. Alshehri was honored with the Questrom-CEMA Best Paper Award for his co-authored work on the efficient aggregation of distributed energy resources – a recognition of significant contributions to the field. He earned his B.S. in Control and Instrumentation Engineering from King Fahd University of Petroleum and Minerals (KFUPM) and his M.S. and Ph.D. in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign (UIUC).



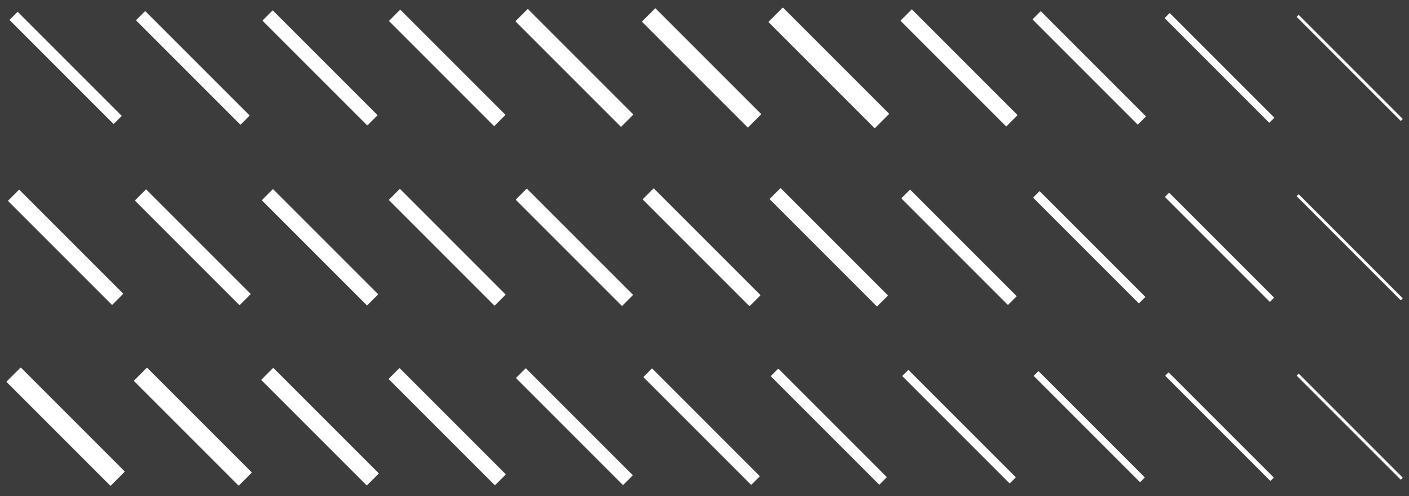
Selahattin Murat Sirin

Dr. Selahattin Murat Sirin is a Research Fellow at KAPSARC specializing in regulation, business economics, electricity markets, and sustainable energy. He earned his Ph.D. from Ivey Business School at Western University and holds an M.P.A. in Public Administration from Harvard University, an M.S. in European Studies from the Middle East Technical University, and a B.A. in Economics from Ankara University. He has published academic papers and reports on subjects such as renewable energy policies and electricity markets.

About the Project

This paper is part of the KAPSARC project “Technoeconomic Analysis of AI Data Center Investments in Saudi Arabia.” The project investigates how large-scale digital infrastructure investments, particularly in data centers, interact with energy systems, regulatory frameworks, and national economic strategy. As countries compete to attract compute intensive industries, understanding how policy levers such as electricity pricing, infrastructure performance, and investment incentives shape cost competitiveness have become increasingly important. The project provides a structured basis for evaluating tradeoffs, assessing externalities, and guiding policy choices that support long-term economic value in resource exporting economies.





kapsarc.org info@kapsarc.org /kapsarc