From megawatts to margin:

AI in the Gulf enters its second phase in 2027

Deloitte | A Middle East Point of View - Issue 50 - 2026 | From megawatts to margin

The capital is committed and the sites are secured. What decides returns now is the pace of the build, the depth of domestic demand, and how far up the value chain the region chooses to go.

Saudi Arabia's operational data center capacity stood at 467 megawatts in the first quarter of 2026, having risen sevenfold since 2021.[1] A single Saudi company, HUMAIN, has published a roadmap targeting 1.9 gigawatts by 2030.[1] One firm's medium-term plan is therefore roughly four times the national installed base as it stands today.

That distance is the most useful way to understand the Gulf technology sector entering 2027. It reflects neither a shortage of ambition nor a shortage of capital, both of which the region has demonstrated at a scale few others could match. It reflects the gap between capacity that has been announced and capacity that is built, commissioned and carrying customer load—a gap every market building at this pace is working through, and one the Gulf is closing faster than most.

Three forces will determine how quickly that gap closes. Value in the AI value chain is migrating towards its two ends: the infrastructure beneath it and the applications above it. Access to advanced compute has become an instrument of statecraft, not simply a matter of commercial procurement. And the region is confronting a sequencing question every large infrastructure industry has faced: having committed the capital, secured the land, and negotiated chip allocations, how and when should the demand and operating capability that sit above the asset be built?

Value in AI is concentrating at the two ends of the value chain

General-purpose AI models—the class of systems behind ChatGPT and its competitors—are commoditizing quickly as they tend to converge fast and capability gaps close in months.

Margin is concentrating at the two ends of the chain: beneath the models, in chips, power, and data centers, and above them, in the applications and agents closest to the customer. The more difficult position to defend is in the middle, where a business owns a general-purpose model and little else.

The region has positioned itself accordingly. Capital is flowing overwhelmingly into the infrastructure end, while leading models are licensed from abroad rather than competed with. Domestic model development continues, principally for Arabic-language and sovereign use cases, but is being pursued as a national capability rather than a commercial bet. That is an efficient allocation of national resource, and a materially different choice from the one most large economies have made.

Gulf capital does take exposure to the model layer, but as an investor rather than an operator. MGX, founded by Mubadala and G42, holds positions in OpenAI, Anthropic, and xAI.[13] The Qatar Investment Authority (QIA) has backed Anthropic and xAI, and HUMAIN is among xAI's more recent investors.[14] Taking equity in the middle of the chain, without the operating burden of competing there, is a more capital-efficient way to participate in a segment whose economics remain the least settled.

Stargate UAE shows the logic in physical form. G42 is funding and building a one-gigawatt compute cluster within a five-gigawatt campus in Abu Dhabi, with OpenAI and Oracle operating it and NVIDIA, Cisco, and SoftBank as partners.[2] The first 200-megawatt phase was expected online in the third quarter of 2026, in a build involving more than 5,000 workers and substantial concrete and steel requirements.[3] The division of roles maps closely onto the value chain: Emirati capital takes the infrastructure end, while the American partners bring the platform and operating layers above it.

> Margin is concentrating at the two ends of the chain: beneath the models, in chips, power, and data centers, and above them, in the applications and agents closest to the customer

OpenAI is the one participant whose core business sits in the middle of the chain, and its presence in an infrastructure consortium is itself evidence of the pressure on that position—the leading model company extending downward into compute and upward into applications rather than defending the model layer alone.

HUMAIN's partnership with Together AI makes the same choice in commercial terms. The two plan a 250-megawatt Saudi facility selling training and inference capacity to developers and enterprises, including customers outside the Kingdom, with first-year annualized revenue potential of more than US$5 billion.[4] It sits alongside a US$5 billion Amazon Web Services (AWS) "AI Zone," a US$10 billion collaboration with Advanced Micro Devices (AMD) and deployments of Qualcomm inference hardware.[5,6] What is being built here is an infrastructure business—capacity sold by the hour, underpinned by cost advantages in power and land that few jurisdictions can match—rather than a model business. The revenue figure matters less as a forecast than as evidence of where contracted money is available today: at the infrastructure end of the chain, rather than in the middle.

A third advantage sits alongside power and land, and it is less durable than either. Operational sovereignty has become a genuine purchase criterion in this region: an institution choosing where to run a regulated workload is asking where the data is permitted to sit and who is able to interrupt it; and benchmark performance is a threshold to clear rather than the differentiator. That position is commercially valuable today, though less uniformly than it appears—in-country residency is already being matched by global providers, increasingly in partnership with local operators.

What is not easily matched is who runs the infrastructure, which nationals hold access, and whose law governs the operating entity. That is the narrower advantage, and the one worth pricing deliberately.

The constraint runs across the whole infrastructure layer

Building AI capacity at gigawatt scale depends on a long chain of inputs, each carrying its own lead time and its own point of failure: advanced chips released under export license and delivered in tranches, land with grid access, high-voltage electrical equipment, cooling systems operating at power densities with limited industry precedent, and commissioning teams who have delivered facilities of this class before. Different inputs bind at different moments, and which one is binding changes by site and by quarter. A project can be fully funded, fully permitted, and still a year away from carrying load.

Globally, the evidence shows up in the distance between announcement and delivery. Transformer deliveries that took 24 to 30 months before 2020 can now stretch to as long as five years.[7]

Lead times for generator step-up transformers passed 160 weeks by the first quarter of 2026, against an average of 143 weeks in 2024, while high-voltage circuit breakers reached 125 weeks from 77 weeks in 2023.[7,8] In the United States, only 13 per cent of the capacity that entered interconnection queues between 2000 and 2019 had reached commercial operation status by the end of 2024, with 77 per cent withdrawn.[9]

The Gulf region is better placed on some of these inputs than others. Generation is abundant and competitively priced, land is available at scale, and projects of this scale have not been constrained by access to capital. Where the region shares the global position is in the queue for heavy electrical equipment, in the phased delivery schedules attached to chip allocations, and in access to people who have commissioned facilities of this class before—simply because very few exist anywhere. The commercial implication is that the milestone worth contracting against is the date a site is live and carrying customer load, rather than the date any single input arrives.

Access to compute has become an instrument of statecraft

Advanced compute is now allocated through government decisions as much as commercial ones. The United States approved access for G42 to the equivalent of up to 35,000 NVIDIA GB300 chips,[3] and NVIDIA's Saudi allocation began with a first tranche of 18,000 Blackwell units directed to HUMAIN.6 Control structures reflect the strategic weight of these assets: Stargate UAE is funded and built by G42 inside a campus established as a UAE national initiative, with the operating mandate contracted to international partners.[10]

Securing these arrangements has been a considerable diplomatic and commercial achievement, and one few other jurisdictions have managed.

One consequence follows directly. Gulf buyers have diversified deliberately across NVIDIA, AMD, and Qualcomm. That reduces exposure to any one supplier's allocation queues, manufacturing capacity, and pricing. It does not reduce exposure to the policy itself: the leading suppliers are American companies operating under a single licensing regime, so vendor choice is a supply-chain hedge rather than a sovereign one.

> The Gulf region is better placed on some of these inputs than others. Generation is abundant and competitively priced, land is available at scale, and projects of this scale have not been constrained by access to capital.

The first phase imported demand, the second phase builds it

Both flagship projects are structured so that the operating and customer-facing roles sit with the international partner in this first phase. At Stargate UAE, the operating mandate rests with OpenAI and Oracle,[2] with Oracle describing the arrangement as pairing its AI-optimized cloud with sovereign infrastructure so that government agencies and commercial institutions can connect their data to leading models. That places the access layer, and with it the customer relationship, on the partner's platform. In the Saudi arrangement, the partners themselves describe the division as combining HUMAIN's large-scale infrastructure and capital deployment with Together AI's cloud platform, inference technology, and global customer ecosystem.[4]

As an opening move, this is well judged. Utilization is what makes a multi-billion-dollar asset financeable, and the fastest route to utilization is to bring in an operator with an existing demand book rather than build one domestically from a standing start. It also imports operating expertise for a facility class with very few global precedents. Sequencing in this order—secure the asset, import the operating capability, then build the domestic platform—is a well-established route in capital-intensive industries, and a faster one than attempting all three at once.

What follows is therefore a question of phasing rather than of judgement. In infrastructure businesses generally, the operating and customer-facing layers are where a significant share of lifetime margin accrues, and the terms governing them are easier to shape early in a partnership than late. The strategic question is not whether to move up the chain, but when, and with which capabilities built domestically.

The region is already acting on this. HUMAIN's leadership has set out an investment approach under which the company backs businesses that commit to using Saudi data centers for part of their computing requirements or to establishing a workforce in the Kingdom, alongside a domestic investment vehicle and a substantially scaled global AI venture fund.[1] That is demand creation through capital deployment: a direct and unusually deliberate answer to the question of where future load originates, and a model other capacity owners in the region are likely to study.

Telecommunications operators entering the same business approach the question from a different starting point. stc has extended its memorandum with HUMAIN to establish a joint venture through center,[3] its data center subsidiary.[11] Operators bring land, grid connections, fiber backhaul and, importantly, established enterprise relationships—a domestic demand base that pure infrastructure owners have to construct from scratch. Their question is how far up the chain to extend, and how fast.

We do, however, see fragility in the market

Three areas warrant continued attention, none specific to any single operator or jurisdiction.

The first is that commercial capital is less patient than state capital, and the financing base is shifting towards it. HUMAIN has supplemented state backing with external arrangements, including a framework of up to US$1.2 billion with the National Infrastructure Fund, alongside preparations for a public listing.[1,12] That is the normal path for infrastructure at this size. The exposure it introduces is sequencing risk: if delivery slips, the cost of capital rises at the point in the build where momentum matters most, and part of the region's cost advantage is given back.

The second is that the chip constraint is political rather than technical. Unlike equipment lead times, which shorten as supply chains respond, an export authorization is subject to revision, and allocations arrive in tranches. It cannot be engineered away through procurement, since the leading suppliers operate under the same licensing regime. What can be managed is the consequence: how quickly workloads re-host, where they sit, and what the contracts say when policy moves.

The third is that global demand for AI compute is being repriced continuously as the largest buyers revise their capital plans. Capacity is long-lived and illiquid; demand is neither. Every capacity owner faces this asymmetry. The region's scale makes it worth managing explicitly through contract structure rather than assuming it away.

Five priorities for 2027

Contract against the date a site goes live. Commitments indexed to the point a facility carries customer load hold up better than those indexed to the arrival of chips, equipment, or power individually.

Advance the demand book in step with the build program. Anchor tenancy, offtake, and pricing are most valuable when they mature alongside construction rather than behind it.

Treat the move up the value chain as a sequencing decision, and make it explicit. Owning the operating capability, followed by the platform and application layers above it, and ultimately the customer relationship itself is a far better position than remaining the compute landlord. These terms are more tractable early in a partnership than at renewal.

Monetize the sovereignty advantage while it remains distinct. Residency is commoditizing as global providers build in-country regions with local partners. The premium worth sizing and pricing is operated sovereignty—who runs the infrastructure, which nationals hold access, and whose law governs the operating entity.

Structure for policy change rather than procure around it. Vendor choice does not reduce exposure to a single licensing regime, which is why contract terms that anticipate policy changes matter more than supplier diversification.

The Gulf has demonstrated something few regions could: the ability to commit capital at scale, negotiate access to tightly constrained technology and move from decision to construction faster than established markets. Those are durable advantages, and they have been earned rather than assumed.

The next phase asks a different question. With the capital committed, the land secured and the chip allocations negotiated, attention turns to how quickly commitments convert into capacity that carries customer load—and to how much of the value beyond that capacity the region builds for itself. The current partnership structures, in which a sovereign entity owns the campus while an international partner brings operating expertise and an existing customer base, are the right way to de-risk a first phase and reach utilization quickly.

What gets built in the second phase, in domestic demand, operating capability and the commercial layers above the asset, will determine how much of the value created by this build-out is retained in the region. Based on the evidence of the last twelve months, that work is already under way.

The strategic question is not whether to move up the chain, but when, and with which capabilities built domestically

By **Hassan Malik**, Partner and TMT Leader and **Mirza Ali**, Director, Monitor Deloitte, Deloitte Middle East

Endnotes

1. Fortune, "Saudi AI giant races to fund data center expansion", 9 September 2026. https://fortune.com/2026/09/09/saudis-humain-turns-to-outside-investment-as-the-kingdom-reins-in-fiscal-spending/.
2. G42, "Global Tech Alliance Launches Stargate UAE", 22 May 2025. https://www.g42.ai/resources/news/global-tech-alliance-launches-stargate-uae.
3. The National, "Stargate UAE's first phase to be completed in third quarter of 2026", 5 December 2025. https://www.thenationalnews.com/business/2025/12/05/stargate-uaes-first-phase-to-be-completed-in-third-quarter-of-2026/.
4. Together AI and HUMAIN, joint press release issued via Business Wire, 31 August 2026. https://finance.yahoo.com/technology/ai/articles/together-ai-humain-form-strategic-120000648.html.
5. Data Center Dynamics, "Humain secures 211 plots of land in Saudi Arabia for data centers", 27 July 2026. https://www.datacenterdynamics.com/en/news/humain-secures-211-plots-of-land-in-saudi-arabia-for-data-centers/.
6. Reuters, via Yahoo Finance, "Saudi's Humain to launch data centers with US chips in early 2026, Bloomberg News reports". https://finance.yahoo.com/news/saudis-humain-launch-data-centers-191120744.html.
7. The Standard, reporting Bloomberg, "US AI expansion hit by power shortages". https://www.thestandard.com.hk/innovation/article/328687.
8. Reuters, "US power companies scramble to secure equipment as surging data center demand strains supplies", 9 July 2026. https://www.kitco.com/news/off-the-wire/2026-07-09/us-power-companies-scramble-secure-equipment-surging-data-center.
9. Lawrence Berkeley National Laboratory, "Queued Up: 2025 Edition, Characteristics of Power Plants Seeking Transmission Interconnection". https://emp.lbl.gov/publications/queued-2025-edition-characteristics.
10. Gulf Business, "Stargate UAE: G42, OpenAI, Oracle, NVIDIA, SoftBank, Cisco to build AI cluster". https://gulfbusiness.com/g42-openai-oracle-nvidia-to-build-stargate-uae/.
11. stc group, Investor Relations H1 2026 results presentation. https://www.stc.com/content/dam/groupsites/en/pdf/InvestorRelationsQ2-2026En.pdf.
12. Business Wire, "HUMAIN and Infra Announce AI and Digital Financing Framework Agreement", 22 January 2026. https://www.businesswire.com/news/home/20260122600897/en/HUMAIN-and-Infra-Announce-AI-and-Digital-Financing-Framework-Agreement.
13. CNBC, "OpenAI, Anthropic backer MGX raises one of the biggest AI funds ever as it closes at $49 billion", 1 July 2026. https://www.cnbc.com/2026/07/01/mgx-ai-fund-uae-49-billion.html.
14. Semafor, "UAE ups pre-IPO stake in Anthropic", 1 June 2026. https://www.semafor.com/article/06/01/2026/uae-ups-pre-ipo-stake-in-anthropic.