AIFC

EMERALD CAPITAL PARTNERS

AI AS INFRASTRUCTURE

2026

Investment, compute, power and capital formation across Kazakhstan and the Turkic space

Astana International Financial Centre

AIFC

AI as Infrastructure
Investment, compute, power and capital formation across Kazakhstan and the Turkic space

Astana International Financial Centre

2026

Foreword

From the Governor of the Astana International Financial Centre

Artificial intelligence is moving beyond the boundaries of the technology sector. It is reshaping how capital is allocated across economies, how organisations operate and how countries approach productivity, infrastructure and long-term competitiveness.

Dear Readers,

Artificial intelligence is moving beyond the boundaries of the technology sector. It is reshaping how capital is allocated across economies, how organisations operate and how countries approach productivity, infrastructure and long-term competitiveness.

This transformation also brings the physical foundations of the digital economy into sharper focus. Computing capacity, data centres, reliable energy systems and international connectivity increasingly influence where technological activity, investment and innovation take place. Governance, access to technology, human capital and international cooperation will also determine how effectively countries participate in this emerging landscape.

Kazakhstan has recognised the strategic importance of this transition. The declaration of 2026 as the Year of Digitalization and Artificial Intelligence reflects the country's commitment to making technology a central pillar of economic modernisation. Through the Digital Qazaqstan strategy, Kazakhstan is advancing AI adoption, strengthening its institutional framework and developing the infrastructure and skills required for a more competitive digital future.

Flagship initiatives such as the Data Center Valley in Ekibastuz demonstrate how Kazakhstan can connect its energy resources, geographic position and investment potential with growing global demand for computing capacity. Together with an expanding technology ecosystem and international partnerships, these developments can position the country as a digital infrastructure and innovation hub connecting Europe and Asia.

The wider Turkic region is undergoing a similar transformation, creating opportunities for deeper cooperation in infrastructure, connectivity, investment, talent and digital services.

The AIFC has a particular interest in how technological change interacts with capital formation, institutional development and cross-border investment. This report examines the global shift towards infrastructure-intensive AI and considers its implications for Kazakhstan and the wider Turkic region.

I hope it supports informed dialogue and practical cooperation among policymakers, investors, operators and technology partners.

Sincerely,
Renat Bekturov
Governor, AIFC

Authors

Ayan Tuleshev
Director, Industry Analysis Department

Diana Assaubayeva
Senior Manager, Industry Analysis Department

Dana Abdykaeva
Senior Manager, Industry Analysis Department

Aisha Abdezova
Manager, Industry Analysis Department

Disclaimer

This report (the "Report") has been prepared for informational and analytical purposes only. It does not constitute legal, financial, investment, regulatory, tax, or professional advice, nor should it be construed as a recommendation or offer to buy, sell, or hold any financial instruments or to follow any investment strategy.

The information contained in this Report is provided on an "as is" basis. While the Report draws on sources considered reliable at the time of publication, we make no representations or warranties, express or implied, as to the accuracy, completeness, timeliness, or reliability of the information, analysis, or conclusions contained herein.

This Report is intended for general informational purposes only and does not constitute a comprehensive analysis. It should not be relied upon as the sole basis for any decision. Readers are encouraged to conduct their own independent analysis and seek professional advice in light of their specific circumstances.

The content of this Report may include opinions, estimates, forecasts, and forward-looking statements that are inherently uncertain and subject to change without notice. Actual results may differ materially from those expressed or implied by such statements. Any reliance placed on the information contained in this Report is at the reader's own risk.

This Report may reference data, publications, or materials from third-party sources for informational or contextual purposes. Such references do not imply endorsement, and we accept no responsibility for the accuracy, legality, or content of any third-party materials.

Nothing in this Report shall be construed as establishing any contractual, fiduciary, or advisory relationship between the reader and the authors, co-authors, or affiliated institutions. To the fullest extent permitted by applicable law, we accept no liability for any direct, indirect, incidental, special, consequential, or punitive damages arising from the use of, or inability to use, this Report.

Use of Artificial Intelligence Tools. In accordance with the requirements of the Law of the Republic of Kazakhstan "On Artificial Intelligence" No. 230-VIII dated 17 November 2025, we hereby disclose that artificial intelligence tools may have been used in the preparation of this Report solely for certain auxiliary technical, editorial, and research-related tasks. The use of such tools was limited in scope and shall not be construed as evidence that this Report, as well as the conclusions and analytical judgments contained herein, were generated automatically by an artificial intelligence system.

In the event of any discrepancy between the English version of this Report and any translation into other languages, the English version shall prevail. We reserve the right to amend, update, or revise this Report at any time without prior notice.

Abbreviations

The following abbreviations and technical terms are used throughout this report.

AI
Artificial intelligence

AIFC
Astana International Financial Centre

AIX
Astana International Exchange

API
Application programming interface

AWS
Amazon Web Services

BIS
Bank for International Settlements

Capex
Capital expenditure

DBRS
DBRS Morningstar (credit-rating agency)

EU
European Union

FDI
Foreign direct investment

GPU
Graphics processing unit

GRES
State district power station (Gosudarstvennaya Raionnaya Elektrostantsiya)

GW / MW / kW
Gigawatt / megawatt / kilowatt (units of electrical power)

HPC
High-performance computing

ICMA
International Capital Market Association

IDC
International Data Corporation

IEA
International Energy Agency

IMF
International Monetary Fund

ISSAI
Institute of Smart Systems and Artificial Intelligence, Nazarbayev University

KZT
Kazakhstani tenge

OECD
Organization for Economic Co-operation and Development

PPP
Public–private partnership

PUE
Power usage effectiveness (data-centre energy-efficiency metric)

REIT
Real estate investment trust

ROI
Return on investment

SCO
Shanghai Cooperation Organization

SLA
Service-level agreement

SPV
Special-purpose vehicle

Sukuk
Shariah-compliant fixed-income instrument

Tbps
Terabits per second

TWh
Terawatt-hour

UAE
United Arab Emirates

UNCTAD
United Nations Conference on Trade and Development

USD
United States dollar

VAT
Value-added tax

Table of contents

About this report

PURPOSE

The purpose of this report is to examine how artificial intelligence is shifting from a software-led technology cycle into an infrastructure-led investment cycle, and what this transition means for Kazakhstan and the Turkic States.

Rather than treating AI only as a model, application or innovation-policy issue, the report focuses on the physical, digital and financial foundations required to scale AI: compute, data centres, cloud platforms, power supply, grid access, fibre connectivity, data governance and long-term capital.

The report frames Kazakhstan's opportunity not as a race to produce frontier AI models, but as a potential role in building a regional Turkic compute and cloud platform for selected AI deployment, sovereign, regulated and enterprise workloads.

SCOPE

The report covers the global AI infrastructure buildout, sovereign AI policy models, Kazakhstan's AI implementation architecture and the regional infrastructure landscape across the Turkic States (the Organization of Turkic States, OTS).

The analysis examines global investment flows, data-centre capacity, compute constraints, electricity demand, cloud infrastructure, sovereign AI strategies and financing models. It then applies this framework to Kazakhstan by assessing AI regulation, national compute initiatives, commercial data-centre capacity, public-sector AI demand, connectivity, power-system readiness and the potential role of the AIFC in capital intermediation.

The report does not seek to forecast AI adoption in every sector or evaluate individual AI applications in detail. Instead, it assesses whether Kazakhstan can convert data-centre pipeline, power, connectivity and finance into a financeable regional AI infrastructure platform.

INTENDED AUDIENCE

The intended audience includes policymakers, institutional investors, development finance institutions, cloud and data-centre operators, technology companies, financial institutions, regulators and corporate leaders.

For policymakers, the report identifies the infrastructure, regulatory and institutional conditions required to support AI deployment. For investors and operators, it highlights the financeability conditions that matter for AI infrastructure projects, including anchor demand, power contracts, grid connection, cloud partnerships, data-governance clarity, permitting and the reliability of key counterparties.

APPROACH AND DATA

The analysis is based on publicly available data, including institutional reports, company announcements, market studies and specialist research from international organizations, consulting firms, technology-sector sources and regional authorities.

The report uses global sources such as Stanford HAI, OECD, IEA, IMF, BIS, McKinsey, UNCTAD and other institutional references to frame the global AI infrastructure cycle. Kazakhstan and regional analysis draw on official government sources, policy documents, company disclosures, data-centre market reviews, telecom and cloud announcements, and publicly reported project information.

Unless otherwise indicated, quantitative data reflect the latest information available at the time of writing. Announced investment values, megawatt capacity, GPU numbers, commissioning dates and project timelines should be interpreted as indicative unless confirmed through official commissioning, independent certification, financial-close documentation or sustained commercial operation. The report distinguishes operational assets from projects that have been announced, are planned, are under development or remain at the concept stage, based on the information available.

Executive Summary

Artificial intelligence is entering an infrastructure-led phase in which competitive advantage increasingly depends on access to compute, data centres, advanced chips, cloud platforms, power and connectivity.

This shift is driving a major new cycle of infrastructure investment. Global corporate AI investment reached approximately USD 581.7 billion in 2025, up around 130% year on year, while McKinsey estimates that meeting projected compute demand could require USD 6.7 trillion in cumulative data-centre capital expenditure by 2030, including approximately USD 5.2 trillion for AI workloads.

Deliverable power is emerging as a critical physical constraint. Global data-centre power-capacity demand is projected to rise from 82.3 GW in 2025 to 219.0 GW by 2030, with inference becoming the largest workload category. Around 20% of planned data-centre projects globally could face grid-related delays, making deliverable power and timely grid connection key determinants of where projects can be built and financed. Progress should therefore be measured not by announced capacity, but by capacity that is financed, constructed, connected, commissioned and utilised.

Sovereign AI is best understood as dependency management rather than full-stack ownership. Data localisation alone does not create operational control. For mid-sized economies, the more credible model is selective autonomy: retaining stronger control over sensitive workloads and critical dependencies while relying on trusted international providers and contractual safeguards elsewhere in the AI stack.

Kazakhstan enters this cycle with a comparatively differentiated starting position. The country has a dedicated AI framework, 896 operational NVIDIA H200 GPUs across two TOP500-ranked systems, 91% commercial data-centre utilisation, and an expanding infrastructure pipeline led by Data Center Valley in Ekibastuz, planned initially at 300 MW with a long-term ambition of up to 1 GW. Demand signals are strengthening through public-sector AI deployment, enterprise cloud adoption and IT-services exports, but long-term contracted workloads are still forming. Power delivery, grid connection and enforceable offtake therefore remain critical to converting infrastructure readiness into commercially viable capacity.

Regional scale can materially improve infrastructure economics by widening the addressable market and supporting higher utilisation. The five member states of the Organization of Turkic States represent approximately USD 1.97 trillion in combined GDP, creating scope for a federated model in which sensitive systems remain national while selected compute, cloud and digital services are delivered across compatible platforms. Such an approach could enable cross-border scale without requiring a single supranational cloud or weakening national control over sensitive systems.

Execution should therefore be phased and demand-led. Sovereign cloud, regulated-sector workloads, GPU-as-a-Service and anchor-tenant projects can establish contracted demand and utilisation first; larger regional structures can follow as the market matures. AIFC-based structures can support this transition through project structuring, private credit, sukuk, capital-market access and enforceable contractual frameworks. Ultimately, success will be measured not by announced capacity, but by commissioned infrastructure, contracted workloads and sustained utilisation.

CHAPTER 1.

The AI Industrial Buildout: From Technology to Infrastructure

Key numbers

USD 581.7 BLN
Global corporate AI investment in 2025

156 GW
Projected global data-centre demand from AI training and inference workloads in 2030

USD 5.2 TRILLION
Estimated cumulative AI-related data-centre capex through 2030

20%
Estimated share of planned data-centre projects globally that could face grid-related delays

950 TWH
Projected global data-centre electricity consumption in 2030

Section summary

Artificial intelligence is entering an infrastructure-led phase in which growth depends increasingly on data centres, advanced chips, cloud platforms, power, cooling and connectivity. This investment cycle is affecting capital formation, electricity demand, equipment imports and the geography of digital infrastructure, while longer-term productivity gains will depend on adoption and Organizational change.

AI infrastructure remains concentrated across firms, technologies and jurisdictions. However, mid-sized economies do not need to control the full stack to participate. Credible positions remain available as infrastructure and supply-chain nodes, deployment economies or capital and contracting platforms, provided that reliable infrastructure is combined with demand, institutional capacity and technology access.

Key insights

1. AI is affecting the economy through two distinct channels: near-term capital formation and longer-term productivity gains.

2. AI infrastructure is an interconnected investment stack rather than a single asset class. Power, connectivity and data-centre facilities have relatively long economic lives and infrastructure-like characteristics, while computing equipment and cloud services carry shorter technology cycles and greater operating risk.

3. Compute availability is shaped by interdependent constraints. AI capability increasingly depends on reliable access to advanced chips, high-bandwidth memory, advanced packaging, specialised networking, data-centre capacity, cloud platforms, software ecosystems and sufficient power capacity. A bottleneck at any layer can limit deployment.

4. Power availability is becoming a critical location factor. Grid capacity, connection timelines, equipment supply chains, electricity costs and operational reliability increasingly determine where AI infrastructure can be developed.

5. The AI economy is concentrated, but participation pathways remain open. Frontier-model development and advanced compute are concentrated among a limited number of firms and markets, but infrastructure hosting, domestic deployment and capital intermediation create viable opportunities for mid-sized economies.

AI as a macroeconomic variable

1.1

Artificial intelligence is becoming a macroeconomic variable because its development and deployment are increasingly compute-, power- and equipment-intensive. Rising use of AI adoption drives investment in semiconductors, data centres, network infrastructure and power systems, affecting fixed capital formation, equipment imports, investment location decisions and, over time, productivity and competitiveness.

The most visible near-term economic effects of AI are higher capital investment and stronger demand for data centres, computing equipment, power and connectivity, while broader economy-wide productivity gains remain more difficult to isolate. The macroeconomic transmission therefore operates in sequence: AI adoption increases demand for **computing capacity; this demand** drives investment in physical and digital **infrastructure**; and the longer-term **return depends on how effectively businesses and public institutions use that capacity to increase output, reduce costs and improve service delivery**.

Recent data show that hundreds of billions of dollars are being directed into AI through corporate investment, acquisitions and capital-market transactions. The Stanford 2026 AI Index reports that global corporate AI investment reached approximately **USD 581.7 billion in 2025**, an increase of approximately 130% from 2024. This measure includes mergers and acquisitions, minority-stake transactions, private investment and public offerings. Private investment accounted for approximately USD 344.7 billion of the total. These figures measure financing and transaction activity rather than direct expenditure on physical AI infrastructure.[1]

Venture-capital data provide a complementary, but narrower, measure of investor activity. According to the OECD, AI companies attracted approximately USD 258.7 billion in venture capital in 2025, equivalent to 61% of total global venture-capital investment of USD 427.1 billion. Companies focused on IT infrastructure and hosting attracted USD 109.3 billion, accounting for more than 42% of AI venture-capital investment and representing the largest amount among the industry segments assessed.[2]

[1] Stanford HAI — AI Index Report 2026: Economy

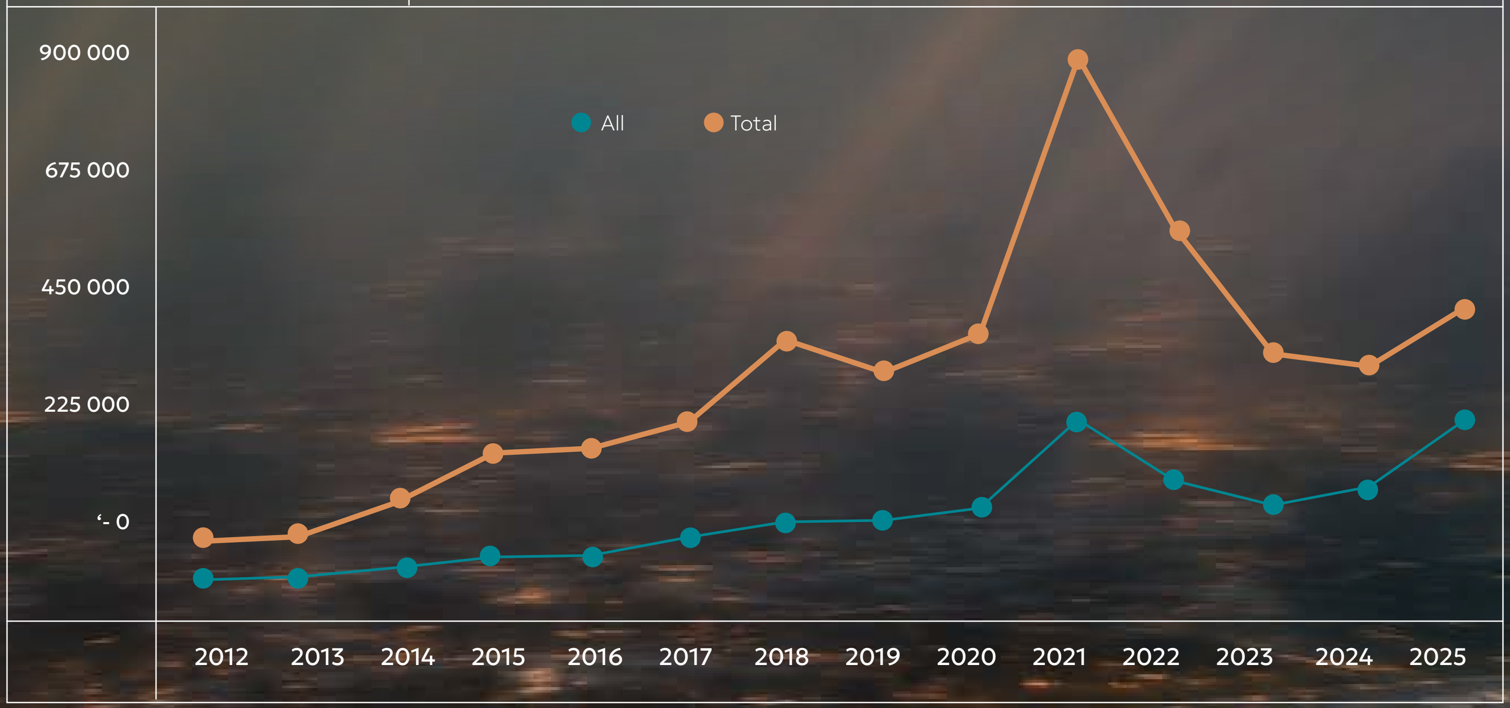
[2] OECD — AI firms capture 61% of global venture capital in 2025

Figure 1

Source: OECD, Venture Capital Investments in Artificial Intelligence through 2025, based on Preqin data, 2026

Global venture capital investment, total and AI-related, 2012–2025

Note: The Stanford and OECD measures should not be added together, as they apply different classifications and include overlapping private financing activity.



The United States provides a clear example of **how AI-related investment can reach a scale that influences** national economic indicators. S&P Global estimates that data centres and related high-technology spending generated around 80% of the growth in US final private domestic demand in the first half of 2025 and raised second-quarter GDP by approximately 0.5 percentage point relative to its previous investment trend.[3]

These figures capture the **near-term economic contribution of AI-related investment through spending on** construction, equipment, software development and energy systems. **They do not, however, measure realised productivity gains**. The longer-term question is whether AI **enables workers and organisations to generate more output from the same amount of labour, capital and time**. McKinsey estimates that generative AI could add 0.1–0.6 percentage points to annual global labour-productivity growth through 2040. **These gains could arise through the automation of routine tasks, faster analysis of information and improved decision-making.**[4]

The scale of investment alone does not determine whether productivity gains will be realised. The wider economic benefit will depend on how effectively countries combine computing capacity with reliable power, connectivity, skilled workers, **institutional capacity and sustained AI adoption** across businesses and public organisations.

[3] S&P Global — Data Center Frontiers: Data center investment moves the macro needle

[4] McKinsey & Company — The economic potential of generative AI: The next productivity frontier

The economics of AI infrastructure investment

1.2

AI infrastructure comprises interconnected investments in power, connectivity, data-centre facilities, computing equipment and cloud services. These layers depend on one another operationally but differ materially in capital intensity, economic life, revenue model and risk exposure.

Unlike conventional software, much of the physical capacity must be permitted, financed, connected and constructed before future demand is fully known. The central investment question is therefore not only how much capacity is proposed, but whether it can be delivered on time, placed into commercial operation and utilised throughout its economic life.

Table 1 — Source: AIFC analysis

Economic characteristics of the AI infrastructure investment stack

Investment layer	Principal assets and revenue model	Core economic exposure
Power infrastructure	Generation, transmission, substations and backup systems; revenue through tariffs, power sales or contracted offtake	Permitting, construction, grid connection, regulation and offtake
Connectivity	Fibre routes, subsea cables and interconnection; revenue through capacity leases, transit and access charges	Route utilisation, redundancy, transit pricing and network effects
Data-centre facilities	Land, buildings, electrical distribution and cooling; revenue through colocation, wholesale leases or powered-shell contracts	Location, power access, construction, occupancy and retrofit requirements
Computing equipment	GPUs, servers, storage and networking hardware; revenue through capacity rental or bundled services	Utilisation, obsolescence, replacement cycles and residual value
Cloud and managed services	Cloud platforms, GPU-as-a-service, inference services and managed environments; revenue through consumption, subscriptions or reserved capacity	Demand, pricing, competition, service quality and operating execution

The distinction between these asset types matters because their economic lives of the underlying assets differ substantially. Data-centre shells, substations and fibre connections may remain productive for decades, while electrical and cooling systems may require upgrades as power densities increase. GPUs and servers operate on substantially shorter technology and replacement cycles. Cloud and managed AI services add a further layer of demand, pricing and competitive risk. AI infrastructure should therefore be understood as an interconnected investment system rather than a homogeneous asset class.

Scale and concentration of capital deployment

The scale of the current investment cycle is significant by historical technology-sector standards. McKinsey estimates that meeting projected global compute demand could require approximately **USD 6.7 trillion** in cumulative data-centre capital expenditure by 2030. Of this amount, **around USD 5.2 trillion** would be associated with AI workloads and USD 1.5 trillion with traditional computing. Across both categories, more than USD 4 trillion could be allocated to computing hardware.[5]

Near-term corporate expenditure is also rising. Market projections cited by Reuters in June 2026 placed combined hyperscaler capital expenditure at approximately USD 725 billion for 2026. The estimate covers broad investment programmes across data centres, chips, cloud infrastructure and related systems and remains subject to changes in corporate guidance, delivery schedules and market conditions.[6]

Cross-border project announcements provide a separate measure of the investment cycle. Preliminary UNCTAD data show that announced greenfield data-centre investment exceeded USD 270 billion in 2025 and represented more than one-fifth of the value of global greenfield projects. These figures measure announced project values, not completed construction, operational capacity or realised foreign direct investment.[7]

[5] McKinsey & Company — The cost of compute: A $7 trillion race to scale data centers

[6] Reuters — Tech companies tap debt markets to fund AI and cloud expansion

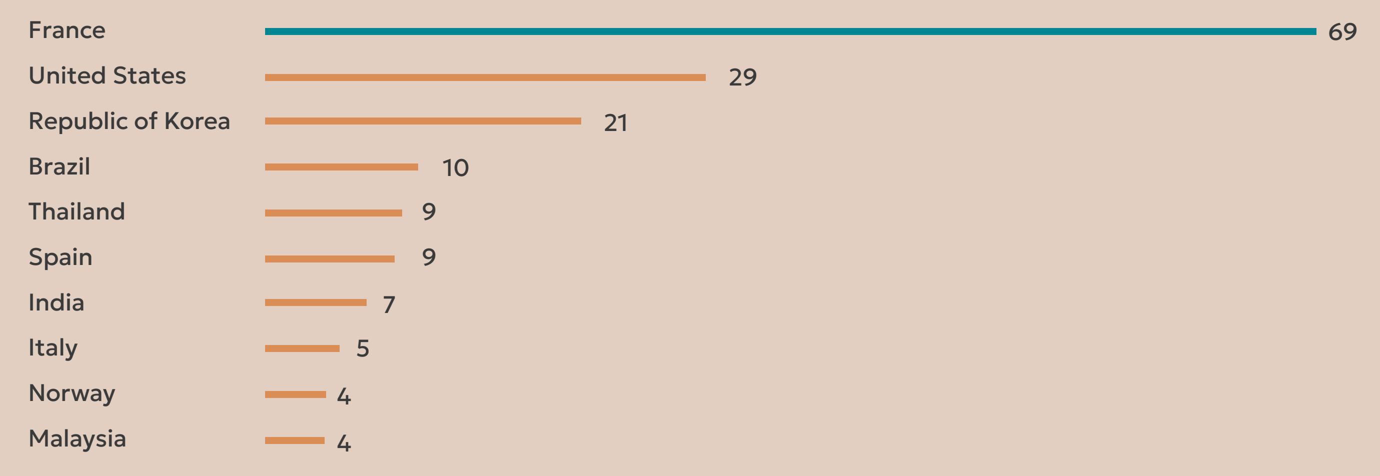
[7] UNCTAD — Data centres are reshaping the global investment landscape

Figure 2

Source: UNCTAD, Global Investment Trends Monitor No. 50, January 2026

Major recipients of announced data-centre greenfield investment, 2025 (USD billions)

Note: These indicators measure different dimensions of the investment cycle. McKinsey's estimate represents modelled cumulative capital requirements; the hyperscaler figure reflects annual corporate capital expenditure across chips, cloud and related infrastructure; and the UNCTAD figure measures announced cross-border project values. None should be interpreted as a measure of completed investment or operational capacity.



Risk allocation across the investment stack

AI infrastructure projects face risks that emerge both within individual layers and at the points where those layers depend on one another. A completed facility may remain idle if its grid connection or equipment delivery is delayed. Computing equipment may become operational before sufficient customer demand has been secured. Changes in processor performance and rack density may also require upgrades to electrical and cooling systems.

Table 2

Source: AIFC analysis

Risk categories in AI infrastructure investment

1. Development risk	Arises from permitting, construction, grid connection and equipment-delivery delays. Power and major data-centre campuses require multi-year development timelines, while technology and workload requirements can change more rapidly.
2. Commercial risk	Arises when capacity is delivered ahead of sufficient demand or depends excessively on a limited number of customers. Underutilisation can reduce returns, while insufficient capacity can increase access costs and dependence on external providers.
3. Technology risk	Is concentrated in computing equipment and facility design. GPUs and servers may become economically obsolete before the underlying buildings and power connections, while rising rack densities can require electrical and cooling retrofits.
4. Financial risk	Reflects leverage, refinancing requirements and the mismatch between long-lived infrastructure and shorter-lived computing equipment.
5. Policy and dependency risk	Arises from dependence on a limited number of chip suppliers, cloud platforms, equipment manufacturers and technology jurisdictions, as well as exposure to export controls, data rules and changes in energy regulation.

These risks are interconnected. Grid-connection and equipment delays can leave completed facilities idle; weak workload demand can reduce utilisation and returns; and changes in chip performance or rack density can alter the economics of existing assets. Investment performance therefore depends on coordination across infrastructure delivery, technology procurement, customer contracting and financing.

Cloud platforms as the channel for access and monetisation

Data centres and computing equipment provide the physical capacity for AI workloads, while cloud platforms make that capacity accessible through APIs, managed environments and enterprise services. Cloud platforms therefore connect physical infrastructure with commercial use by providing access to computing capacity and enabling operators to sell it as a service. Physical capacity alone is insufficient to attract workloads: operators must also provide cloud interconnection, specialised hardware, service capability and identifiable enterprise demand. For investors, technical capacity alone does not establish a viable business model: utilisation depends on customer access, cloud partnerships, service integration and contractual visibility over future workloads.

Compute, workloads and power constraints 1.3

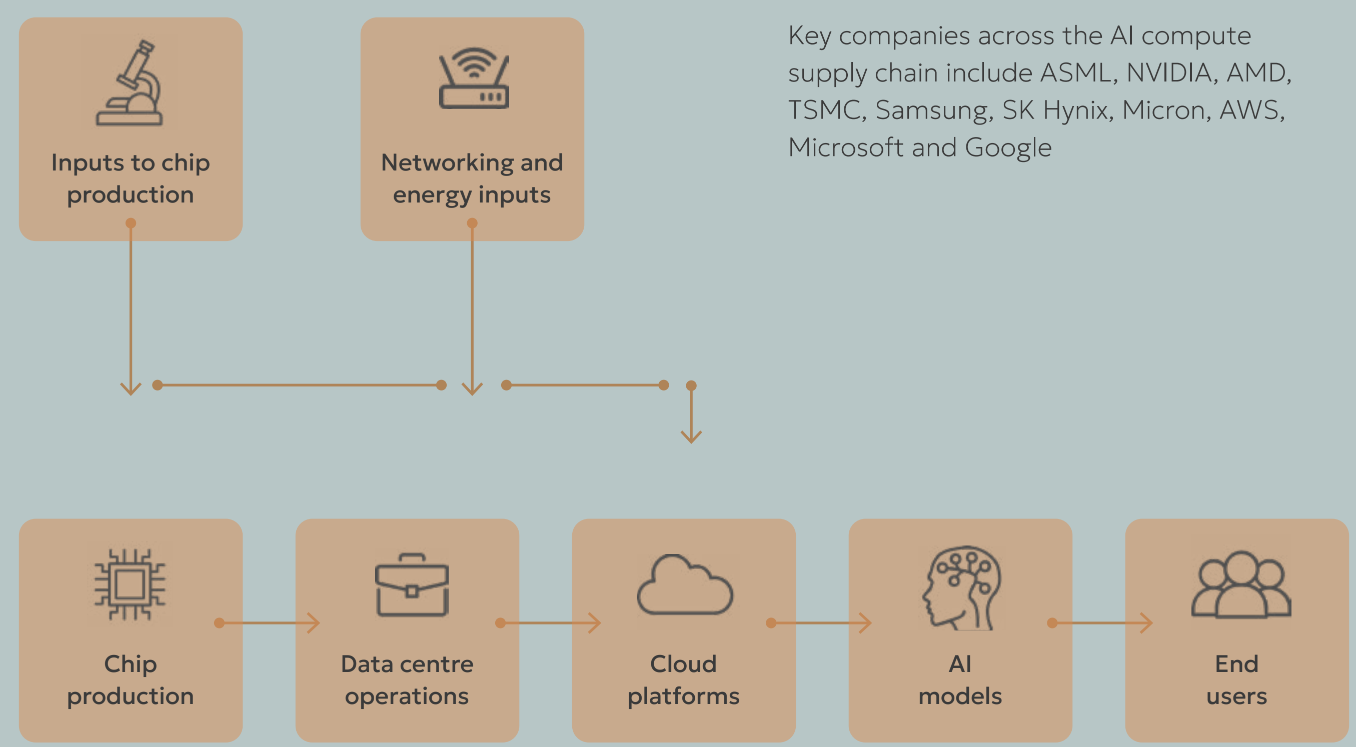
In the AI economy, compute capacity refers to the effective processing capability available to train and operate AI models. It depends on an integrated system of processors, memory, networking, software, data-centre infrastructure and reliable electricity. A constraint at any layer can delay deployment, reduce utilisation or leave completed assets commercially idle.

The compute supply chain is concentrated among a limited number of global firms and jurisdictions. Access to advanced processors, memory, manufacturing capacity and specialised equipment may therefore be affected by production constraints, commercial allocation decisions, delivery schedules and export controls. Compute supply resilience depends not only on the ability to finance equipment, but also on sustained access to the technologies and services required to operate it.

Figure 3

Source: Adapted from OECD, Competition in Artificial Intelligence Infrastructure (2025)

The AI compute supply chain: from production inputs to end users



Key companies across the AI compute supply chain include ASML, NVIDIA, AMD, TSMC, Samsung, SK Hynix, Micron, AWS, Microsoft and Google

Workload characteristics influence location

AI workloads are commonly divided into training and inference. Training develops or materially improves a model, while inference uses a trained model to generate outputs. These workloads have different infrastructure requirements, but the distinction does not determine location on its own.

Frontier-model training generally requires large accelerator clusters, high-performance networking, advanced cooling and substantial power capacity. Because large-scale training is generally not delivered as a real-time user-facing service, it can be located in power-rich areas where land, grid access and cooling conditions are favourable.

Inference supports deployed applications such as search, recommendation systems, enterprise assistants, public-service platforms, financial analytics and industrial systems. As AI adoption expands, inference is expected to account for a growing share of compute demand.

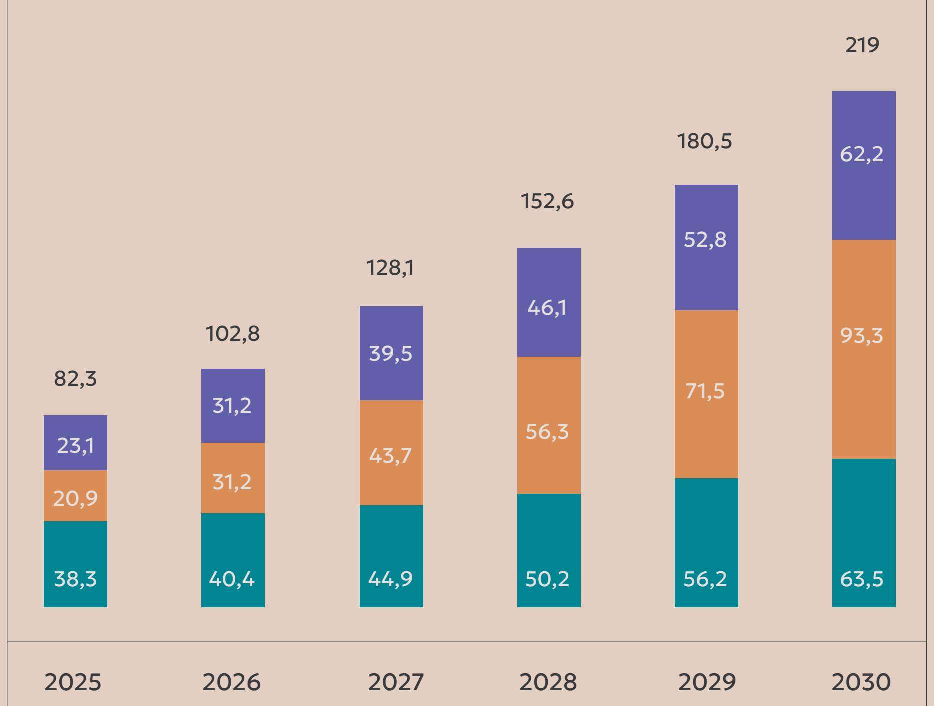
Figure 4

Source: McKinsey & Company, The next big shifts in AI workloads and hyperscaler strategies, December 2025

Global data-centre power-capacity demand by workload, 2025–2030

Non-AI | AI inference | AI training | Total

McKinsey projects that global data-centre power-capacity demand could increase from 82.3 GW in 2025 to 219.0 GW in 2030. Within this total, power-capacity demand associated with AI inference is projected to rise from 20.9 GW to 93.3 GW, while AI training demand is projected to increase from 23.1 GW to 62.2 GW. Under this scenario, inference would become the largest workload category by 2030.



The growth of inference does not mean that all capacity must be located close to users. Many applications can continue to be served economically from established external cloud regions. Local or regional infrastructure becomes more relevant where workloads involve:

- **regulatory, contractual or security constraints;**
- **recurring demand sufficient to support local utilisation; and**
- **material data-transfer costs or operational continuity requirements;**
- **latency or real-time performance requirements.**

Regulatory requirements may create inherently local demand, while latency and data-transfer advantages must still be assessed against the cost and performance of external cloud alternatives. Infrastructure location should therefore be determined by workload requirements. Markets should identify applications that require or economically justify local processing, assess recurring demand and only then determine the appropriate infrastructure configuration.

Deliverable power determines project timing

Power availability is a site-specific delivery issue rather than a matter of national power capacity in the aggregate. A data-centre project requires electricity at the relevant location, scale, reliability level and commissioning date. Installed generation capacity elsewhere in the power system does not necessarily mean that a particular site can be connected on commercially acceptable terms.

The International Energy Agency projects that global data-centre electricity consumption could increase from **approximately 460 TWh in 2024 to around 945 TWh** by 2030, representing just under 3% of global electricity consumption. Electricity use by AI-optimised data centres is expected to grow considerably faster than overall data-centre demand.[8]

Facility requirements also vary by workload and equipment configuration. Accelerator-dense clusters require more electricity per rack than conventional enterprise computing and may require liquid cooling, upgraded electrical distribution and additional backup systems. Existing facilities may therefore require upgrades to electrical distribution, cooling and backup systems before they can accommodate newer accelerator clusters.

Grid-connection timing is becoming a material development constraint. The IEA estimates that around 20% of planned data-centre projects could face delays unless grid constraints are addressed. Transmission infrastructure can take several years to plan and construct, while delivery times for transformers, cables and other equipment have also increased.[8]

Investors therefore assess connection timing, reliability and expansion potential alongside electricity cost. A low tariff provides limited advantage where firm capacity cannot be delivered within the project timetable.

[8] International Energy Agency — Key Questions on Energy and AI: Executive Summary

Selective participation by mid-sized economies

1.4

For the purposes of this report, a mid-sized economy is a market capable of hosting and deploying AI services but without the scale or technology position to compete across the **full global AI value chain.**

Such economies do not need to reproduce advanced semiconductor manufacturing, frontier-model development and global cloud platforms domestically. Their economic opportunity lies in selecting functions where local capabilities, demand conditions and regional access provide a credible advantage.

Participation can take several forms. An economy may rely primarily on external cloud and AI services, develop local capacity for selected workloads, support domestic adoption across businesses and public institutions, or provide financial and commercial infrastructure for regional projects. The appropriate combination depends on where local provision creates a measurable advantage over established external alternatives.

Table 3 Source: AIFC analysis.

Participation pathways for mid-sized economies

Economic function	Role in the AI economy	Conditions for viability	Principal risk
Host and operate capacity	Provide data centre, compute, cloud or interconnection capacity to domestic and regional users	Deliverable power, competitive connectivity, equipment access, operating capability and contracted demand	Capacity is developed without sufficient utilisation or a clear advantage over external providers
Develop and aggregate demand	Build recurring workloads through enterprise adoption, public services and sector-specific applications	Economically valuable use cases, accessible data, organisational readiness and customers able to make sustained commitments	Pilot projects fail to develop into scaled and recurring demand
Support project development and capital formation	Provide the institutional, financial and contractual environment required to develop infrastructure and technology projects	Credible project pipelines, identifiable customers, viable revenue models and enforceable contracts	Financial and institutional structures are developed before projects have sufficient demand or operational readiness

These functions reinforce one another. Infrastructure requires recurring demand, workloads require reliable capacity, and financing requires credible projects supported by both. Economies should therefore avoid developing physical capacity independently of customer demand or treating financing as a substitute for viable project fundamentals.

Implications for Kazakhstan

Kazakhstan's opportunity lies in selective participation rather than replication of the full AI value chain. Its most credible position is likely to combine demand-led domestic and regional capacity with financial and contractual support for viable projects.

Progress should be assessed through commissioned capacity, utilisation and contracted workloads rather than announced megawatts or project values alone.

CHAPTER 2.

Sovereign AI: From Concept to Policy

Key numbers

5,427
Data centres located in the United States, more than ten times the number in any other country

1,590 EFLOPS
China's officially reported intelligent computing capacity at the end of 2025

19
AI Factories established by the European Union, supported by 13 regional AI Factory Antennas

1 GW
Planned capacity of Stargate UAE, with an initial 200 MW expected to become operational in 2026

Section summary

Sovereign AI is best understood as a framework for managing dependencies across compute, cloud, data, models, operations and legal jurisdiction. It does not require domestic ownership of the entire technology stack. As AI enters public services, regulated sectors and critical infrastructure, countries must decide which capabilities require direct control, which can be accessed through trusted providers, and which external dependencies need contractual, regulatory or operational safeguards. The United States, China and the European Union illustrate different combinations of commercial scale, domestic capability-building and trusted governance. The United Arab Emirates, Saudi Arabia, Bahrain and Singapore demonstrate more selective approaches based on infrastructure partnerships, institutional coordination, public cloud demand and specialised deployment. For mid-sized economies, the practical objective is an operating model that classifies workloads, identifies material dependencies and applies proportionate levels of control.

Key insights

1. **Data localisation does not by itself create AI sovereignty.** Operational control also depends on access to compute, cloud platforms, models, technical support, legal jurisdiction and service continuity.

2. **Sovereignty requirements should be calibrated to workload sensitivity.** General productivity tools, regulated applications and strategically sensitive systems require different levels of domestic control, trusted external access and operational safeguards.

3. **Selective autonomy is more credible than full-stack self-sufficiency for most economies.** National resources should be concentrated on capabilities where domestic influence creates clear economic, operational or security value, while other layers are accessed through trusted partnerships.

4. **Execution is the ultimate test of sovereign AI policy.** Strategies, laws and infrastructure announcements establish direction, but credibility depends on systems that are deployed, governed, resilient and used in practice.

Sovereign AI as a dependency-management framework

2.1

Sovereign AI refers to a country's ability to access, deploy and govern AI systems in accordance with its legal framework, security requirements and strategic priorities. It does not necessarily require domestic ownership of every layer of the AI value chain. In practice, it is a framework for deciding which capabilities require direct control, which can be accessed through trusted external providers and which dependencies need regulatory, contractual or operational safeguards.

This distinction has become more important as AI moves from general-purpose applications into public administration, financial services, healthcare, critical infrastructure and industrial operations. In these settings, AI systems do more than process information: they generate outputs, influence decisions and operate within workflows that can have legal, commercial, operational and security consequences.

Sovereign AI is consequently broader than data sovereignty. A country may require domestic data storage while remaining dependent on imported processors, foreign cloud platforms, externally developed models, proprietary software or offshore technical support. Data localisation alone does not provide operational control if essential computing, software or support services remain concentrated among a small number of external providers.

For most economies, full domestic control across advanced semiconductor production, frontier-model development, hyperscale cloud and large-scale computing infrastructure is neither economically efficient nor technically realistic. These capabilities require levels of capital, market demand, specialised knowledge and supply-chain access that are concentrated among a limited number of firms and jurisdictions.

A more practical objective is selective autonomy. This means identifying the workloads and dependencies that matter most and applying an appropriate level of control to each. The required approach may differ substantially by use case. General productivity tools may be accessed through international cloud platforms under standard commercial and data-protection arrangements. AI systems used in financial supervision, public-sector decision-making, healthcare, energy networks or other sensitive environments may require stronger controls over data access, model behaviour, cybersecurity, auditability, continuity and legal jurisdiction.

Three policy choices follow from this approach:

1. Domestic control, where disruption, unauthorised access or external jurisdiction would create unacceptable operational, security or legal risk.

2. Trusted external access, where international providers can offer greater scale, capability or efficiency under suitable legal and commercial arrangements.

3. Dependency safeguards, where external reliance remains necessary but can be managed through supplier diversification, access controls, audit rights, continuity planning and regulatory oversight.

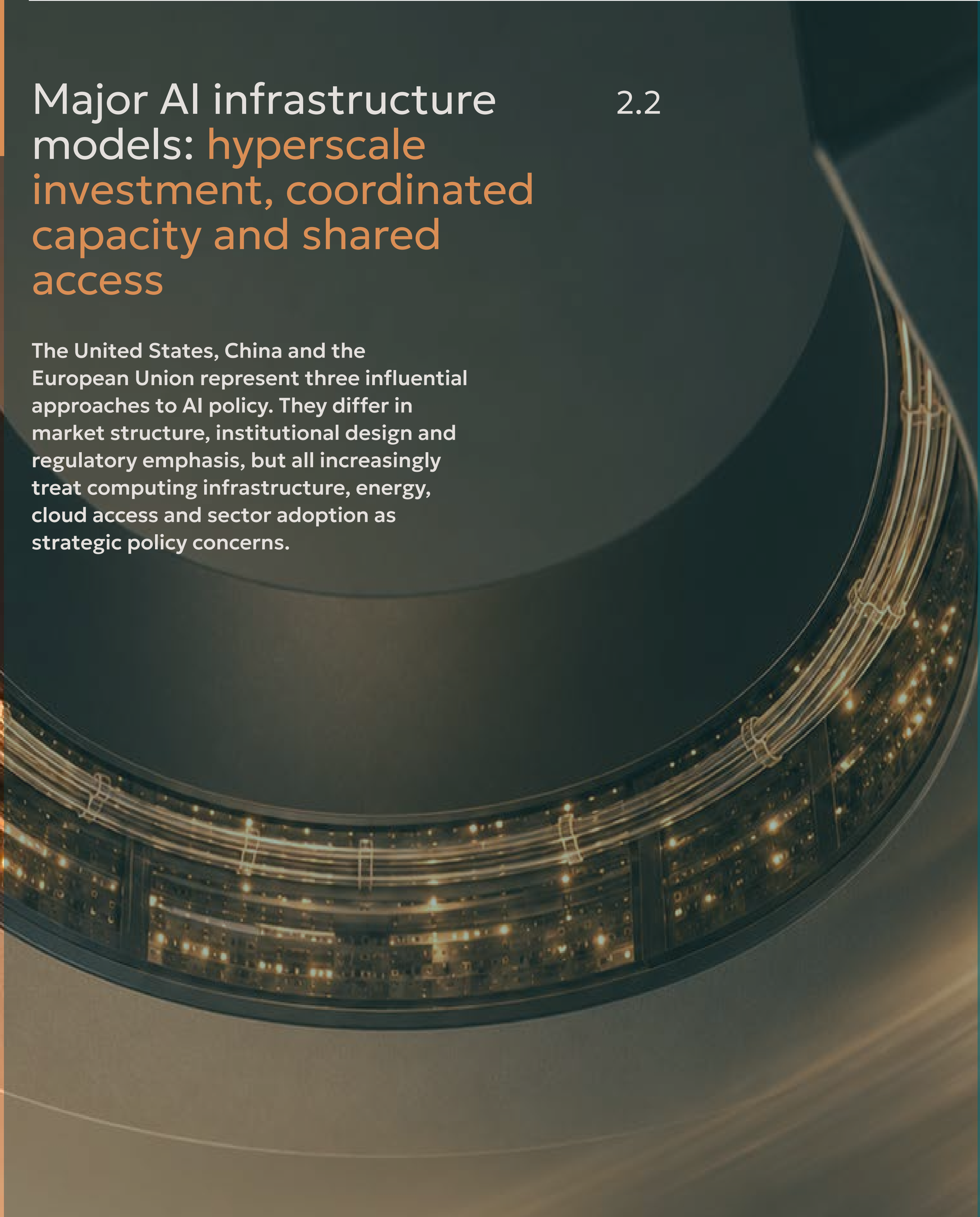
The optimal choice depends on four considerations: the sensitivity of the workload, the concentration and substitutability of suppliers, the consequences of disruption and the cost of domestic provision. Sovereignty should therefore be assessed at the level of specific systems and use cases rather than through a uniform objective of localisation.

Table 4 — Source: AIFC Industry Analysis Department.

Sovereign AI dependency map

Dimension	Why it matters	Policy relevance
Compute and cloud access	AI systems require processors, data centres, storage, networking, cloud platforms and reliable power	Hosting requirements, provider concentration, access conditions and continuity.
Data	Models depend on lawful, high-quality and contextually relevant national, sectoral and language datasets	Data access, protection, sharing, cross-border transfer and permitted use.
Models	Externally developed models may not reflect local languages, laws, institutions or sector requirements	Evaluation, adaptation, fine-tuning and the ability to replace unsuitable models.
Operations	Sensitive systems require secure access, monitoring, reliable performance and incident response	Operational responsibility, cybersecurity, auditability and service continuity.
Regulation	AI deployment creates questions of liability, intellectual property, user rights and legal accountability	Applicable law, provider obligations, sector oversight and dispute resolution.
Capital and procurement	Financing and procurement arrangements influence supplier dependence, technology access and control over critical assets	Funding continuity, procurement terms, access rights and contractual safeguards.

These dimensions are interdependent: domestic compute has limited value without usable data, suitable models and operating capability, while imported technologies may still require local testing, adaptation and accountability mechanisms. Few countries can manage every dependency through domestic resources. The objective is therefore to combine direct control, trusted access and safeguards in proportion to the economic and operational importance of each workload, enabling participation in the global AI ecosystem while keeping material dependencies visible and manageable.



Major AI infrastructure models: hyperscale investment, coordinated capacity and shared access

2.2

The United States, China and the European Union represent three influential approaches to AI policy. They differ in market structure, institutional design and regulatory emphasis, but all increasingly treat computing infrastructure, energy, cloud access and sector adoption as strategic policy concerns.

The differences are relative rather than absolute. The United States places greater weight on private capital, frontier technology companies and commercial scaling. China combines domestic capability-building with coordinated industrial deployment. The European Union emphasises trusted adoption through common regulation and publicly supported compute. Each model combines markets, public policy, infrastructure and governance in different proportions.

Comparative infrastructure assessment

The three models address different parts of the AI infrastructure challenge.

The United States is strongest at mobilising private capital and building hyperscale capacity rapidly. Its main constraint is the physical availability of power, grid connections and semiconductor supply.

China is strongest at coordinating computing capacity with energy systems, telecommunications networks and industrial policy. Its main constraints are utilisation efficiency and access to the most advanced parts of the semiconductor supply chain.

The European Union focuses on widening access to advanced computing resources through shared public infrastructure and common rules. Its main challenge is achieving sufficient scale and converting public compute into commercially competitive firms and applications.

For emerging AI infrastructure markets, the relevant lesson is not to copy one model in full. A viable approach may combine the US emphasis on commercial investability, China's coordination of power and industrial demand, and the EU's emphasis on shared access, interoperability and trusted deployment.

United States

Hyperscaler-led infrastructure scale

The United States has the world's largest commercially financed AI infrastructure ecosystem. Its model combines hyperscale cloud providers, frontier model developers, advanced semiconductor designers, specialised data-centre operators, deep capital markets and large enterprise customers.

2.2.1

The Stanford AI Index Report 2026 estimates that the United States hosts approximately 5,427 data centres, more than ten times the number located in any other country. This installed base supports large-scale cloud computing, frontier-model training and rapidly expanding inference demand. This large installed base provides a favourable foundation for cloud and AI expansion, although facility counts do not measure their power capacity, accelerator density or ownership.[9]

Private investment reinforces this infrastructure advantage. US private AI investment reached approximately USD 285.9 billion in 2025. US-based institutions also produced 59 notable AI models in 2025, reflecting their ability to secure the large-scale computing resources required for frontier development.[9]

The model is not entirely market-driven. America's AI Action Plan and related executive actions identify data centres, electricity generation, grid infrastructure and semiconductor production as strategic national priorities. Federal policy seeks to accelerate permitting for qualifying AI infrastructure projects, including data centres requiring more than 100 MW of new power demand and their associated generation and transmission facilities.[10]

The principal advantage of the US model is the speed at which private firms can aggregate capital, chips, energy contracts and technical expertise. Hyperscalers can invest across the full stack, including custom accelerators, data centres, cloud platforms, foundation models and enterprise applications.

Its main limitation is physical concentration. Continued expansion depends on the availability of electricity, transmission connections, transformers, cooling systems, suitable land and local permits. The scale of individual projects can also create pressure on regional power systems and increase dependence on a small number of cloud and semiconductor suppliers.

[9] Stanford HAI — AI Index Report 2026: Economy

[10] The White House — America's AI Action Plan

China

Coordinated computing capacity and industrial demand 2.2.2

China's AI infrastructure model is characterised by policy coordination across computing capacity, electricity supply, telecommunications networks, domestic technology development and industrial deployment. Rather than treating data centres as standalone assets, national policy increasingly positions them within a wider system linking regional development, energy resources, digital connectivity and technological self-reliance.

A central component is the East Data, West Computing strategy, which seeks to shift suitable workloads from major eastern demand centres towards western regions with greater availability of land and energy.[11] National computing hubs and data-centre clusters are connected through high-capacity networks, allowing latency-tolerant workloads such as model training, batch processing and data storage to be located away from congested metropolitan markets. Low-latency applications, however, may still need to remain closer to users.

Infrastructure expansion has supported increasingly competitive model development. According to the Stanford AI Index Report 2026, Chinese institutions produced 35 notable AI models in 2025, compared with 59 in the United States. More importantly, the performance gap between leading Chinese and US models narrowed sharply, reaching approximately 2.7 percentage points by March 2026.

Demand is reinforced by the **AI Plus initiative,** which promotes AI deployment across manufacturing, science, transport, healthcare, consumer services and public administration. China's industrial base provides a major channel for deployment: approximately 295,000 industrial robots were installed in 2024, representing 54% of global installations, while domestic manufacturers supplied 57% of new units.[12]

The principal advantage of China's model is the coordination of supply and demand. Compute capacity can be planned alongside energy availability, fibre connectivity, public procurement and industrial policy. This can create a clearer demand pipeline for data-centre investors and cloud operators than expansion based primarily on expectations of future commercial use. The main risk is overexpansion, especially where demand remains geographically or commercially weak. Installed capacity does not automatically ensure high utilisation, competitive performance or adequate returns. China also remains exposed to constraints in leading-edge semiconductor fabrication, high-bandwidth memory, manufacturing equipment and parts of the accelerator software ecosystem.

[11] AI Proem — China's Eastern Data and Western Computing

[12] SESEC — Bimonthly Newsletter, July–August 2025

European Union

2.2.3 Shared compute and trusted infrastructure access

The European Union's approach places greater emphasis on shared access to publicly supported computing infrastructure. Rather than attempting to reproduce the US hyperscaler model directly, the EU seeks to connect supercomputing resources with universities, start-ups, SMEs and industrial users that may not be able to procure advanced AI compute independently.

The AI Factories programme combines EuroHPC supercomputers with AI-optimised hardware, data resources, technical expertise and user-support services. By April 2026, the European Commission reported 19 AI Factories and 13 associated AI Factory Antennas. These infrastructure and service ecosystems are intended to broaden access to model training, fine-tuning and inference capacity across member states.

The infrastructure model is therefore based not only on constructing computing facilities, but also on organising access. This is particularly relevant for smaller firms and research institutions that face high entry costs in commercial cloud markets.[13]

The next proposed layer is the AI Gigafactory programme. The InvestAI initiative aims to mobilise EUR 200 billion for AI-related investment, including EUR 20 billion to support up to five AI Gigafactories.[14] These facilities are intended to provide substantially larger computing capacity for the development of advanced models.

Regulation is integrated into this infrastructure model. **The AI Act** provides a common legal framework for AI deployment, while public computing programmes seek to ensure that trusted and compliant applications have access to the necessary infrastructure. The EU model therefore combines compute access with regulatory assurance and cross-border coordination.

Its principal advantage is inclusiveness. Publicly supported infrastructure can reduce dependence on a small number of non-European cloud providers and give research organisations, start-ups and industrial users access to advanced computing resources.

Its limitations are fragmentation and scale. European capacity is distributed across multiple countries, institutions and funding mechanisms. Public facilities may widen access, but they do not automatically create the concentration of capital, chips, engineering talent and electricity supply found in US hyperscale ecosystems. The success of the model will depend on utilisation, user demand and the ability to convert shared compute into commercially competitive products.

[13] European Commission — AI Factories

[14] European Commission — EU launches InvestAI initiative to mobilise €200 billion for AI

Selective autonomy approaches 2.3

Most economies cannot reproduce the scale of the United States, China or the European Union. A more practical approach is selective autonomy: concentrating policy, capital and institutional capacity on specific parts of the AI system while obtaining other capabilities through international partnerships.

The United Arab Emirates, Saudi Arabia, Bahrain and Singapore illustrate four different mechanisms. The cases were selected not as complete national models or direct templates for replication, but because they demonstrate distinct approaches to managing AI-related dependencies.

Table 5

Source: AIFC Industry Analysis Department

Selected approaches to AI autonomy

Economy	Entry point	Mechanism	Transferable lesson	Limitation
UAE	Large-scale compute	Programme-level partnerships	National influence can be built through structured partnerships	High capital and execution requirements
Saudi Arabia	Coordinated capability-building	Sovereign-backed platform	A central platform can coordinate procurement, infrastructure and deployment	Concentrated execution risk
Bahrain	Cloud-first government with an in-country hyperscaler region	Public-sector cloud-first policy	Anchor government demand can support commercial infrastructure	Reliance on regional demand and providers
Singapore	Trusted deployment	Governance and specialised capability	Trusted deployment can create value under infrastructure constraints	Continued dependence on external capacity

United Arab Emirates

partnership-structured infrastructure

Key numbers

USD 355 million
UAE AI hardware spending in 2025

USD 892 million
Projected AI hardware spending, 2029

5 GW
Total UAE–US AI Campus capacity

NO. 1
Global ranking by population-level AI adoption

The United Arab Emirates is developing AI capacity through large-scale infrastructure programmes and international technology partnerships. It combines strong domestic adoption with access to foreign chips, cloud platforms and frontier models. The UAE ranked first globally in population-level AI adoption at the end of 2025, when 64% of its working-age population was estimated to use AI.[15] AI hardware spending reached approximately USD 355 million in 2025 and is projected to rise to USD 892 million by 2029.[16]

The centrepiece of this approach is the planned 5 GW UAE–US AI Campus in Abu Dhabi. Within the campus, Stargate UAE is being developed as a 1 GW compute cluster, with the first 200 MW scheduled to enter operation in 2026. G42 is leading construction, OpenAI and Oracle will operate the cluster, NVIDIA will supply advanced computing systems, and Cisco and SoftBank are participating in the wider partnership.[17]

The model does not depend on domestic ownership of every technology layer. Instead, the UAE uses sovereign coordination, capital and local infrastructure to structure access to global technology and operating expertise. High adoption and rising hardware spending provide a domestic demand base, while the scale of the campus creates the potential to serve wider regional workloads. The principal lesson is that national influence over AI infrastructure can be built through partnership architecture, even where advanced chips, models and platforms continue to come from international providers.

15 Microsoft AI Economy Institute — Global AI Adoption 2025

16 WIRED Middle East — The Middle East’s AI spending is becoming a two-horse race

17 G42 — Global technology alliance launches Stargate UAE

Saudi Arabia

institutional coordination through sovereign capital

Saudi Arabia's approach is centred on HUMAIN, a PIF-owned company established in May 2025 to operate across data centres, cloud infrastructure, AI models and applications.[18] The model brings strategic direction, capital allocation, hardware procurement and international technology partnerships within a single sovereign-backed platform. This allows Saudi Arabia to coordinate the development of the full AI value chain rather than pursue separate infrastructure, cloud and model initiatives.

HUMAIN is targeting 6.6 GW of data-centre capacity by 2034 and has formed partnerships with NVIDIA, AMD, AWS, Qualcomm, Groq and xAI covering AI training, inference and cloud infrastructure.[19] Construction of its first data centres in Riyadh and Dammam has begun, while the first shipment of advanced NVIDIA GPUs arrived in December 2025.[20] This expansion addresses a growing domestic market: Saudi AI hardware spending is projected to increase from USD 378 million in 2025 to USD 1.09 billion by 2029. In parallel, 33.1% of Saudi establishments reported using AI technologies in 2025, including more than half of establishments in information and communications, financial services and education.[21]

Saudi Arabia therefore illustrates how sovereign capital and institutional coordination can bring together infrastructure, technology partners and commercial adoption across the AI value chain. The model's distinguishing feature is not only the scale of investment, but the concentration of capital, infrastructure development and operating capabilities within a single national platform.

Key numbers

USD 378 million
Saudi AI hardware spending in 2025

USD 1.09 billion
Projected AI hardware spending, 2029

6,6 GW
HUMAIN target data-centre capacity by 2034

33%
Saudi business AI adoption

[18] Public Investment Fund — HUMAIN

[19] Data Center Dynamics — HUMAIN seeks US data-centre equity partner and targets 6.6 GW by 2034

[20] Saudi Gazette — HUMAIN receives first shipment of NVIDIA AI chips

[21] WIRED Middle East — The Middle East's AI spending is becoming a two-horse race

Bahrain

cloud-first sequencing

Key numbers

85%+
Government workloads hosted in the cloud

1,385+
Government services migrated to cloud infrastructure

50,000
Bahrainis targeted for AI training by 2030

USD 1.2 billion
Projected cloud contribution to Bahrain's GDP, 2026

Bahrain illustrates a demand-led approach in which public-sector procurement was used to establish a domestic cloud market before large-scale infrastructure was deployed. Its 2017 Cloud First Policy directed government entities to prioritise cloud services over new on-premises systems, shifting public IT expenditure towards scalable infrastructure and creating an early source of anchor demand.[22]

AWS announced a Bahrain Region in 2017 and launched the AWS Middle East (Bahrain) Region in July 2019. It was the first AWS Region in the Middle East and opened with three Availability Zones, supporting locally hosted, resilient and fault-tolerant services. Government adoption subsequently reached significant scale: more than 85% of government workloads have moved to the cloud, covering 72 entities, more than 1,385 government services and 570 e-services.[23]

The same infrastructure is now supporting the transition from cloud adoption to AI deployment. In December 2025, Bahrain's Information & eGovernment Authority announced that Amazon Bedrock would be integrated into the government cloud, giving public entities access to managed generative-AI models and tools through the existing AWS environment.[24] Bahrain's experience therefore shows how cloud procurement, regulatory acceptance and public-sector demand can create a foundation for subsequent AI adoption without first building a nationally owned computing platform.

The principal lesson is sequencing. Bahrain established demand and common procurement rules, attracted an in-country hyperscale region and then expanded the platform towards AI services. This approach lowers the initial infrastructure burden and accelerates access to advanced capabilities. Its limitation is that the availability, pricing and development of those capabilities remain closely linked to the strategy of the global cloud provider.

22 Amazon Web Services — Bahrain's Cloud First success story
23 Amazon Web Services — Bahrain's Cloud First success story
24 Bahrain Information & eGovernment Authority — AWS Bahrain launches Amazon Bedrock

Singapore

trust-led deployment and regional adaptation

Key numbers

1.4+ GW
Existing data-centre capacity

S$1+ billion
Public AI research and talent investment, 2025–2030

UP TO S$500 million
Government allocation for GPUs and other advanced AI chips

UP TO S$150 million
Enterprise Compute Initiative for access to AI tools and computing capacity

Singapore has developed AI relevance through governance, sector adoption and local adaptation, while land and energy constraints limit further infrastructure expansion. Its approach combines access to international technology with domestic capability in regulation, assurance and specialised applications.

In May 2026, Singapore updated its National AI Strategy around ten refreshed priorities and strengthened national coordination through the National AI Council. The country already has more than 1.4 GW of data-centre capacity.[25] Its Green Data Centre Roadmap provides for at least 300 MW of additional capacity, with a further 200 MW linked to the use of green-energy options.

Singapore has also committed more than S$1 billion to public AI research and talent development from 2025 to 2030 and aims to expand its pool of AI practitioners to 15,000.[26] Governance tools such as AI Verify provide standardised testing across transparency, explainability and safety, while SEA-LION offers open multilingual and multimodal models adapted to Southeast Asian languages and cultural contexts.

Singapore demonstrates that infrastructure, trust and specialised capability can reinforce one another. Compute capacity supports research and deployment, while governance tools, talent and regional-language models allow the country to capture value beyond physical hosting and strengthen its role as an applied-AI hub for Southeast Asia.

[25] IMDA Singapore — Red Dot Analytics helps data centres keep cool

[26] Singapore Ministry of Digital Development and Information — National AI R&D Plan

Comparative implication

The four cases show that selective autonomy can begin with different entry points: infrastructure partnerships, institutional coordination, public cloud demand or trusted deployment. Each approach retains external dependencies. The policy objective is therefore to reserve stronger domestic influence for critical capabilities while governing the remaining dependencies through partnerships, contracts and operational safeguards.

Data embassies: sovereign continuity through trusted foreign hosting

2.4

A data embassy allows a government to store protected copies of selected critical data and information systems in another country so that essential public services can continue if domestic infrastructure is disrupted by a cyberattack, natural disaster or other crisis.

The main systems generally remain at home, while the foreign facility serves as a secure backup and, where agreed, an emergency operating environment. Unlike ordinary overseas cloud storage, the arrangement is governed by a treaty or dedicated legislation specifying who controls the data, who may access the facility, what legal protections apply and how security and operations are managed. A data embassy is therefore not a diplomatic mission with automatic immunity, but a specialised continuity mechanism supported by clearly defined legal and operational safeguards.

For analytical purposes, two broad approaches can be distinguished: treaty-based continuity arrangements initiated by the state that owns the systems, and statutory hosting frameworks established by a host country for qualifying foreign customers.

Treaty - based continuity models

Estonia established the first treaty-based data embassy. The 2007 cyberattacks demonstrated the vulnerability of a highly digitalised state to disruption and made geographically separated continuity capacity a national-security priority. Estonia and Luxembourg signed the agreement on 20 June 2017, and the data embassy became operational in 2018.[27]

Under the agreement, Estonian server resources are hosted in a Luxembourg state-owned Tier IV data centre. The facility remains on Luxembourg territory, but Estonia retains operational control, while the treaty protects the hosted data and systems against unauthorised access and interference. The infrastructure can support not only data backup but also the operation of critical services if systems in Estonia become unavailable.[28]

The initial deployment covered ten priority systems and datasets, including the population, business and land registers, the identity-documents register, the state gazette, treasury information and judicial systems. Estonia allocated approximately EUR 2.2 million over the first five years, including around EUR 1 million for establishment and an average of EUR 236,000 annually for rental and data communications.[29] The project demonstrates that a data embassy can begin with a limited group of continuity-critical systems rather than replicating the entire government technology environment.

Monaco subsequently adopted a similar approach. Monaco and Luxembourg signed a hosting agreement on 15 July 2021, providing legal protection for designated Monégasque data and systems located in Luxembourg. For a state covering approximately two square kilometres, substantial geographic separation between primary and recovery infrastructure cannot be achieved domestically. External hosting therefore provides the functional equivalent of a distant disaster-recovery site supported by state-to-state guarantees.[30]

[27] Government of Estonia — Estonia to establish the world's first data embassy in Luxembourg

[28] e-Estonia — Data Embassy Factsheet

[29] ERR News — Data embassy in Luxembourg to cost €2.2 million over five years

[30] Legilux — Luxembourg-Monaco agreement on hosting sovereign data and information systems

Bahrain's statutory sovereign-hosting framework

Bahrain provides a different host-side model based primarily on domestic legislation. Legislative Decree No. 56 of 2018 was introduced to facilitate the provision of cloud-computing services from data centres in Bahrain to qualifying foreign public and private entities. Its scope extends beyond government continuity systems and covers infrastructure-, platform- and software-as-a-service.[31]

The central feature of the framework is the allocation of jurisdiction over subscriber content. Data, software and other content belonging to a qualifying foreign subscriber remain subject to the laws and jurisdiction of the foreign state in which the subscriber is domiciled or established. The legislation also provides a process through which orders issued by the relevant foreign authorities may be recognised and enforced in Bahrain.

The framework applies only to foreign jurisdictions and hosting environments formally designated by Bahrain's Council of Ministers. A designation may be withdrawn, subject to a transition period allowing affected subscribers and service providers to make alternative arrangements. By March 2026, the designated list covered 19 foreign jurisdictions. Separately, Bahrain had announced data-sovereignty initiatives with international partners including Italy and Cyprus.[32]

These examples differ in both legal form and operational maturity. Estonia operates an established data embassy hosting designated government systems. Monaco has created a comparable legal basis through a bilateral agreement, although public information on the scale and active use of its hosted environment remains more limited. Bahrain, by contrast, has established an enabling statutory regime for qualifying foreign customers rather than a single data-embassy deployment. Together, the cases show that trusted foreign hosting can take different institutional forms, but its credibility depends on clear jurisdiction, operational authority, technical separation and enforceable safeguards.

[31] Bahrain Business Laws — Law on providing cloud-computing services to foreign parties

[32] Bahrain Information & eGovernment Authority — Italy granted sovereignty over data hosted in Bahrain

Implications for mid-sized economies: from sovereignty ambition to operating model

2.5

For mid-sized economies, the practical challenge is to translate selective autonomy into an operating model that classifies workloads, maps dependencies and assigns proportionate control mechanisms. National strategies should distinguish between ordinary AI use, regulated applications and strategically sensitive systems, as each category may require a different level of control over data, computing infrastructure, models, operations and legal jurisdiction.

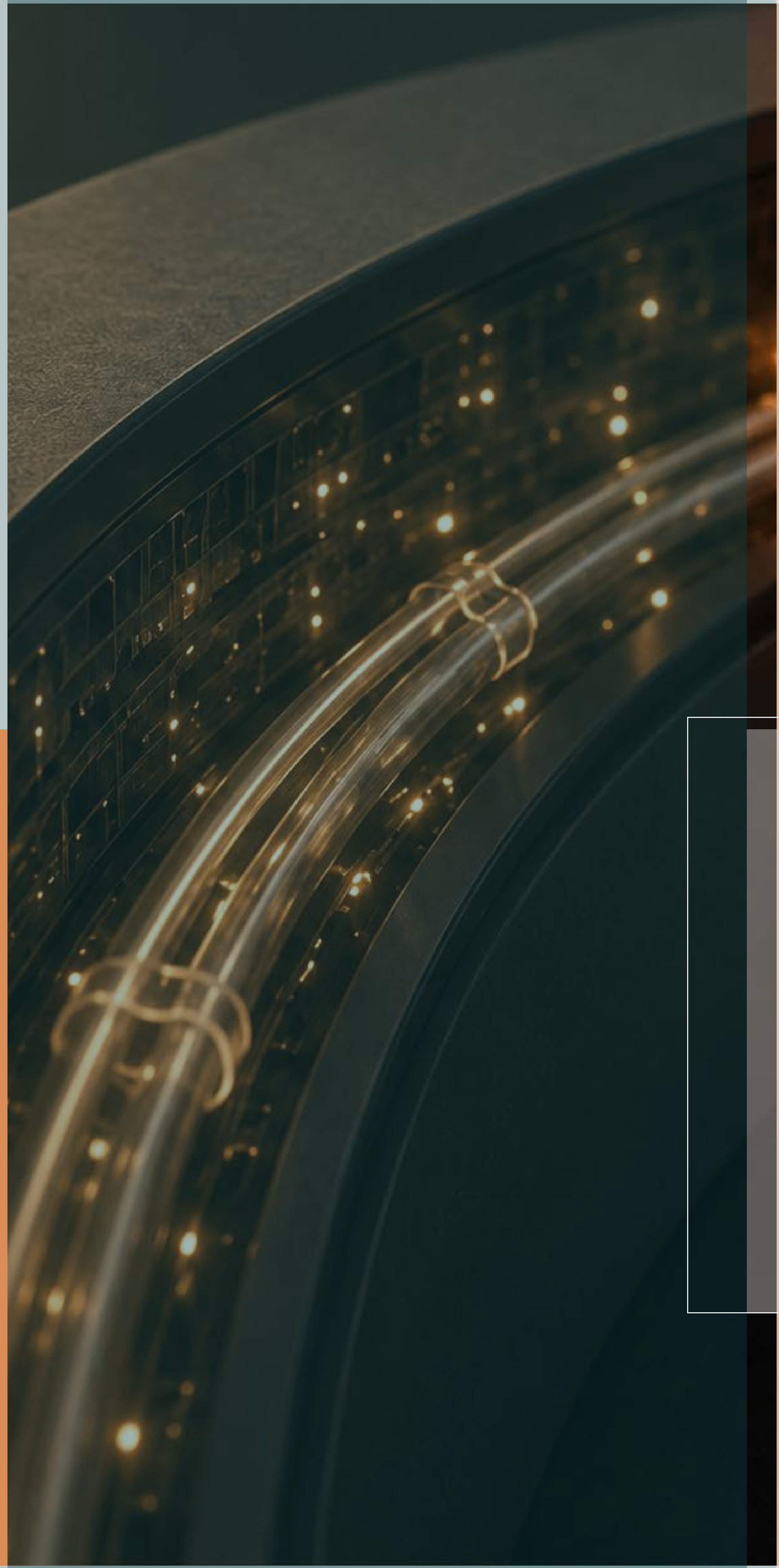
A practical operating model involves four decisions.

Table 6

Source: AIFC Industry Analysis Department.

Sovereign AI decision framework for mid-sized economies

Operating decision	Core question	Policy implication
Classify workloads	Which applications are low-risk, regulated or strategically sensitive?	Apply different requirements according to the legal, operational and security consequences of each workload
Map dependencies	Which external providers, technologies and jurisdictions are essential to deployment?	Identify concentration, continuity, access and substitution risks across compute, cloud, data, models and operations
Select the level of control	Is domestic capability necessary, or can trusted external access provide an acceptable outcome?	Use domestic control where risks are material and partnerships where external provision offers greater scale or efficiency
Establish delivery conditions	What infrastructure, regulation, procurement, skills and institutional coordination are required?	Convert policy objectives into deployable systems, credible projects and measurable adoption outcomes



Workload classification and dependency mapping should be undertaken together. A dependency becomes strategically material where a system performs an important function, relies on a concentrated or difficult-to-replace provider and would be costly to relocate or continue during disruption. The same provider relationship may be acceptable for a general productivity application but insufficient for financial supervision, healthcare, public administration or critical infrastructure.

Domestic ownership is only one control mechanism. Alternatives include trusted cloud arrangements, supplier diversification, contractual access and audit rights, local operating requirements, data-transfer safeguards and continuity plans. The selected mechanism should be proportionate to the consequences of failure and the cost of domestic provision.

Three cross-cutting principles follow:

1. Requirements should be proportional to workload sensitivity. Uniform localisation rules can impose unnecessary costs on low risk uses while failing to address the specific dependencies of sensitive systems.

2. External access should include accountability and continuity safeguards. Legal jurisdiction, operational responsibility, provider concentration and substitution options should be considered before deployment.

3. Capability should be measured through operational use. Strategies, laws and infrastructure announcements create direction, but credibility depends on systems that are deployed, governed and used in practice.

For Kazakhstan, this framework provides a basis for assessing which capabilities require stronger domestic control, which can be accessed through partnerships and where additional safeguards may be justified. Chapter 3 applies this approach to the country's infrastructure, institutions, research base and emerging AI workload demand.

CHAPTER 3.

Kazakhstan: AI Infrastructure Readiness

Key numbers

~896 NVIDIA H200 GPUs
Combined operational capacity of Alem.Cloud and Al-Farabium

91%
Commercial data-centre rack utilisation in 2025

1 GW
Planned long-term power capacity of Data Center Valley, starting from 300 MW

USD 1.143 billion
Kazakhstan's domestic IT-services exports in 2025

APPROXIMATELY USD 30 billion
Projected investment in the full-scale development of Data Center Valley in Ekibastuz

USD 0.025 per kWh
Indicative long-term electricity price proposed for Data Center Valley investors

Section summary

Kazakhstan is moving beyond broad digitalisation and entering the early development of an AI infrastructure market. Its profile already stands out within the region, supported by dedicated AI legislation, a specialised ministry, operational GPU clusters, an expanding commercial data-centre market, a pipeline of larger and more advanced projects, and emerging public-sector demand.

The decisive test is execution: converting announced capacity into commissioned facilities, secured power supply, cloud partnerships, anchor tenants and bankable project structures. Power delivery, grid access, resilient international connectivity, emissions performance and enforceable offtake agreements will determine which projects reach operation. As regional competition intensifies, Kazakhstan's position will depend not on the scale of individual announcements, but on coordinated delivery across infrastructure, regulation, power, connectivity, demand and finance.

Key insights

1. **Kazakhstan is moving from policy formation into implementation.** Dedicated AI legislation, a specialised ministry, operational national computing infrastructure and growing public-sector adoption provide a strong institutional foundation for wider AI deployment.

2. **Kazakhstan is establishing a tangible AI infrastructure base.** High commercial data-centre utilisation, operational GPU clusters and a growing pipeline of cloud and data-centre projects demonstrate clear market momentum. The next stage is to convert announced capacity into operational infrastructure supported by power, connectivity, tenants and financing.

3. **Power is emerging as a key competitive advantage.** Kazakhstan's generation base and the Ekibastuz Data Center Valley proposition create a strong platform for large-scale compute infrastructure at competitive energy costs. Grid delivery, tariff certainty and lower-carbon power pathways will shape its long-term competitiveness.

4. **Demand is broadening across government, enterprise and the technology sector.** Public-sector AI agents, national language models, enterprise cloud adoption, IT exports and startup activity are expanding the domestic workload base. This creates a pathway toward recurring contracts, committed capacity and more predictable utilisation.

5. **Regional positioning is becoming more active.** Its combination of scalable power, operational computing capacity, improving international connectivity and a supportive institutional framework provides a platform for serving demand across Central Asia and the wider Turkic region. The pace of project delivery will determine how much of this opportunity Kazakhstan captures.

Kazakhstan's macroeconomic and investment context

3.1

Kazakhstan enters the current AI-infrastructure investment cycle with a larger economy and a higher level of fixed-capital investment than at the beginning of the decade. Nominal GDP increased from approximately **USD 197 billion in 2021 to USD 306 billion in 2025,**[33] **while real GDP expanded by 6.5% in 2025**[34]**.** These indicators reflect the growing scale of Kazakhstan's economy and the broader macroeconomic environment in which the country's digital and AI infrastructure market is developing.

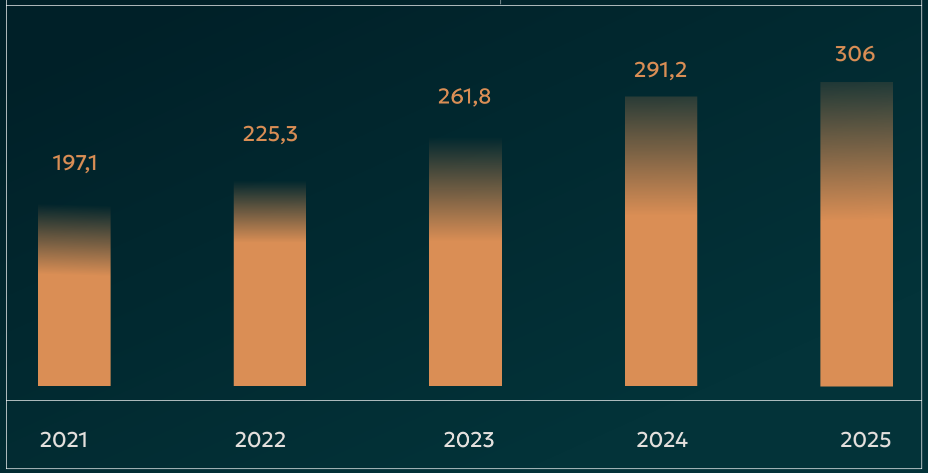
[33] World Bank — GDP (current US$): Kazakhstan

[34] Bureau of National Statistics of Kazakhstan — National accounts publication

Figure 5

Source: Bureau of National Statistics of the Republic of Kazakhstan.

Kazakhstan's nominal GDP, 2021–2025, USD billion



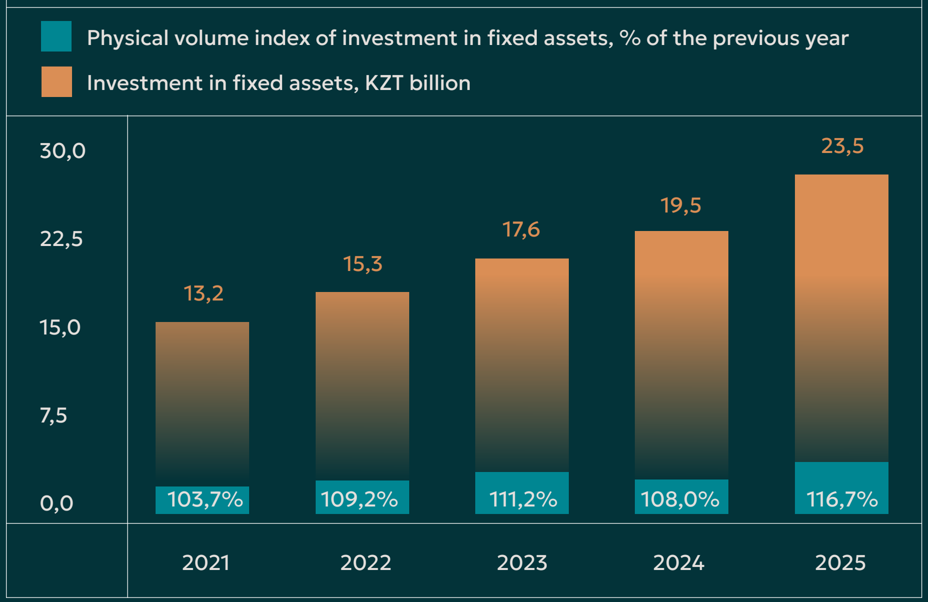
The information and communication sector generated KZT 3.65 trillion of gross value added in 2025, equivalent to 2.3% of GDP, and recorded real growth of 6.6%. Fixed-capital investment directed to the sector reached KZT 481.8 billion, an increase of 32.0% in comparable prices and 2.1% of total fixed-capital investment.[35] These results demonstrate the growing scale of digital-sector activity and capital formation.

[35] Bureau of National Statistics of Kazakhstan — Investment statistics publication

Figure 6

Source: Bureau of National Statistics of the Republic of Kazakhstan.

Investment in fixed capital, 2021–2025, KZT trillion



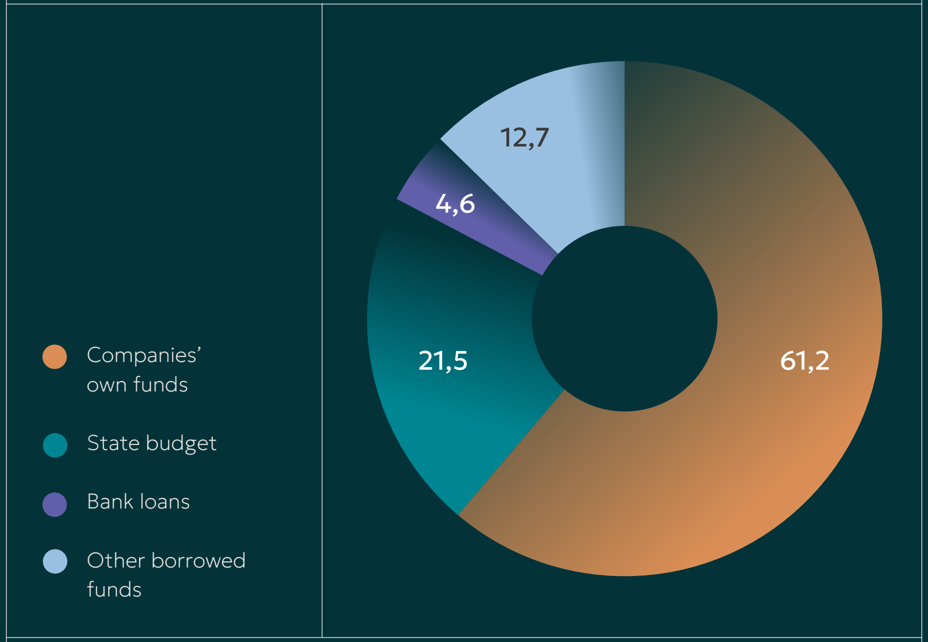
Kazakhstan also recorded substantial growth in aggregate investment activity. Investment in fixed capital increased **from KZT 13.2 trillion in 2021 to KZT 23.5 trillion in 2025,** representing cumulative nominal growth of approximately 77%. In 2025 alone, its physical volume rose by 16.7%.

The structure of investment financing provides additional context. Enterprises' own funds accounted **for 61.2% of fixed-capital investment in 2025, while the state budget contributed 21.5%, other borrowed funds 12.7% and bank loans 4.6%.** Thus, it indicates that corporate balance sheets and public funding remained the principal sources of capital formation, while bank lending played a comparatively limited aggregate role.[19]

Figure 7

Source: Bureau of National Statistics of the Republic of Kazakhstan.

Sources of financing for investment in fixed assets, 2025, %



Domestic financing conditions remain relatively restrictive despite the recent easing of monetary policy. The National Bank raised the base rate to 18.0% in October 2025 before reducing it to 17.0% in June 2026 and 16.75% in July. For capital-intensive projects, borrowing costs remain relevant, particularly where revenues are denominated in tenge while servers, accelerators, networking and cooling equipment are priced in foreign currency.

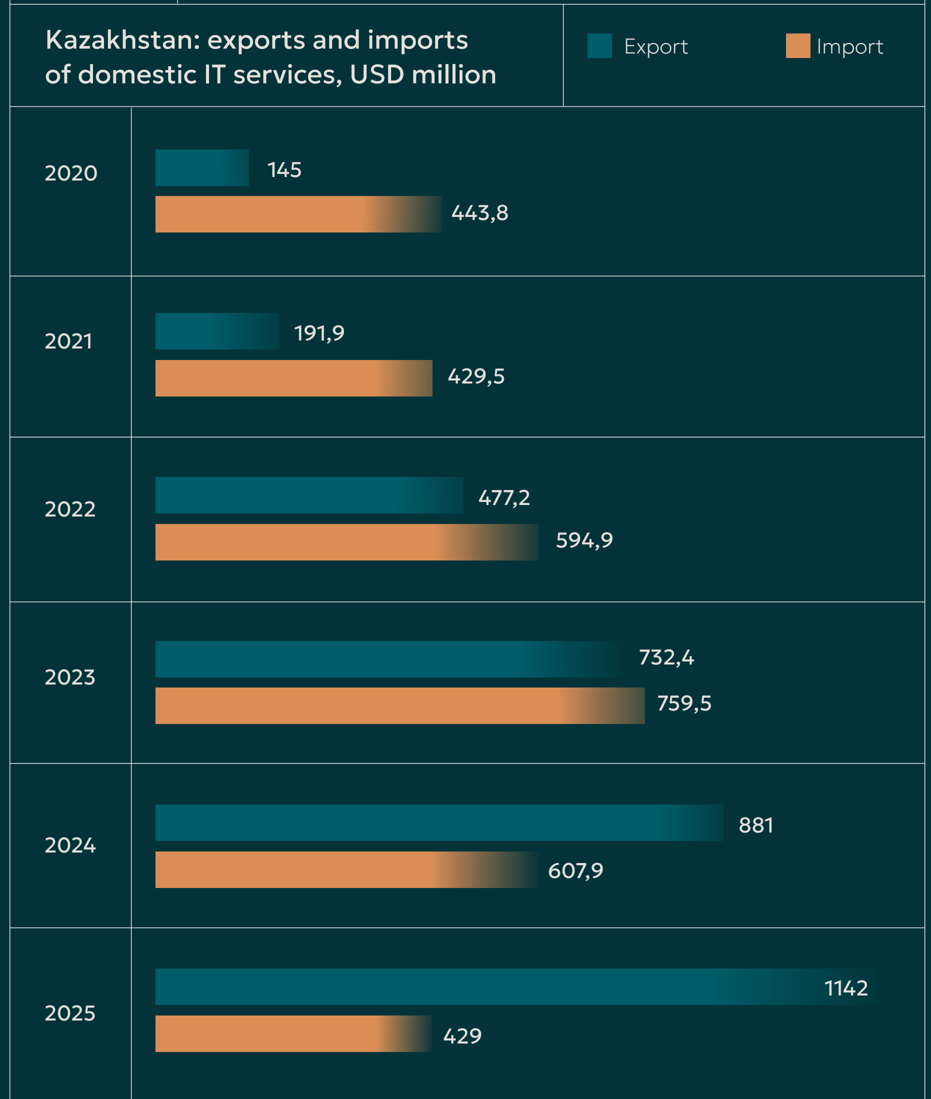
Kazakhstan's international trade in digital services provides a complementary indicator of technology-sector activity. Exports of telecommunications, computer and information services reached USD **1.355 billion in 2025, compared with imports of USD 735 million, resulting in a services surplus of approximately USD 620 million.** Within this category, domestic IT services accounted for USD 1.143 billion of exports and USD 429 million of imports. The broader category includes activities beyond AI and cloud infrastructure, but its export performance indicates an expanding base of cross-border technology services.[36]

[36] National Bank of Kazakhstan — International trade in services statistics

Figure 8

Source: National Bank of Kazakhstan.

Kazakhstan: exports and imports of domestic IT services, USD million



Kazakhstan's economy and aggregate fixed-capital investment grew between 2021 and 2025, alongside increased investment in the information and communication sector. These indicators describe the country's broader macroeconomic and investment environment. The commercial case for AI infrastructure depends on project-level factors, including customer demand, equipment costs, currency exposure, financing tenor, access to technology and the allocation of risks among developers, operators, investors and customers. The following section examines Kazakhstan's policy and institutional framework for AI implementation.

3.2

Kazakhstan's AI implementation architecture

Between 2024 and 2026, Kazakhstan developed a more structured framework for AI policy and implementation. The Concept for the Development of Artificial Intelligence for 2024–2029 set the initial priorities, followed by dedicated legislation, a specialised ministry, national computing initiatives, a government-wide strategy and international partnerships.

The legal framework was expanded through Law No. 230-VIII "On Artificial Intelligence". Signed on 17 November 2025 and effective from 18 January 2026, the law introduced a risk-based approach to AI regulation. It classifies AI systems according to their impact on safety and individual rights and establishes requirements for system owners, operators and users. These include transparency requirements, protection of user rights, risk management and restrictions on certain applications. High-risk systems are subject to additional governance and monitoring requirements.[37]

In January 2026, the President declared 2026 the Year of Digitalisation and Artificial Intelligence. Government Resolution No. 60 of 31 January 2026 approved the corresponding action plan and provided for a commission chaired by a Deputy Prime Minister. The plan sets targets across AI, data, cybersecurity, telecommunications, digital assets and government platforms, including the adoption of 90% of planned regulatory measures and the migration of at least 30% of state information systems to the QazTech platform. Responsible government bodies report quarterly on implementation to the Ministry of Artificial Intelligence and Digital Development. The Digital Code, which entered into force in July 2026, adds a wider layer of regulation covering digital data, platforms, automated decision-making, smart contracts and other elements of the digital environment. Together, the AI Law and the Digital Code form the statutory framework for AI governance in Kazakhstan. Their implementation involves secondary regulation, supervisory guidance, technical standards and enforcement procedures.

Institutional responsibility has also been centralised. The Ministry of Artificial Intelligence and Digital Development was established in September 2025 as the main government body responsible for AI policy, digitalisation, e-government, data governance, cybersecurity and communications.[38]

The Digital Qazaqstan Strategy, adopted in June 2026, brought these policy areas into a wider national digital-transformation agenda. It sets priorities for AI adoption, digital public services, infrastructure and human-capital development. In parallel, the government advanced specific initiatives, including national GPU capacity, AI platforms and the proposed Data Center Valley project in Ekibastuz.[40]

Kazakhstan already operates national computing assets and is deploying AI applications across public institutions. Larger infrastructure initiatives add a longer-term capacity layer by connecting national AI policy with data-centre development, cloud services, power infrastructure and regional connectivity.

Kazakhstan has also expanded its international technology partnerships. In June 2026, it joined Pax Silica, a United States-led initiative focused on trusted supply chains for the AI economy. The initiative covers areas including critical minerals, energy, semiconductors, data centres, computing capacity, advanced manufacturing, research and workforce development. Kazakhstan also signed the related AI Opportunity Partnership joint statement.[41]

Kazakhstan has therefore established components of a national AI framework: dedicated legislation, a specialised ministry, public computing initiatives and international partnerships. The next stage is more operational. It requires secondary regulation, effective supervision, procurement standards, project-level coordination and clearer rules for data access, cross-border services and high-risk applications.

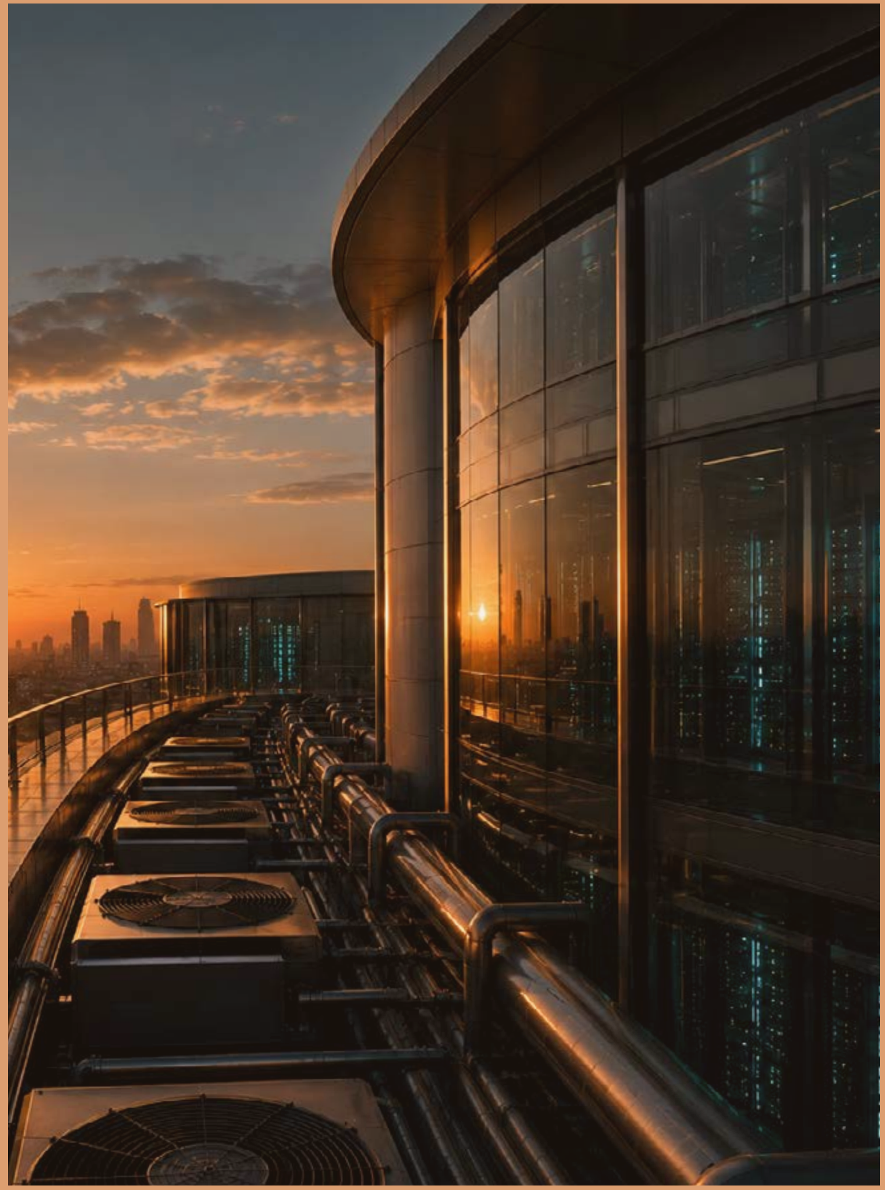
[37] Unicase Law — Kazakhstan's First Artificial Intelligence Law to Take Effect on 18 January 2026

[38] Adilet — Digital Code of the Republic of Kazakhstan

[39] The Astana Times — Kazakhstan to establish Ministry of Artificial Intelligence and Digital Development

[40] Ministry of Artificial Intelligence and Digital Development — Digital Qazaqstan Strategy

[41] The Astana Times — Kazakhstan joins US-led Pax Silica initiative to advance AI economy



The next stage is implementation: commissioning infrastructure, securing reliable power connections, making computing capacity commercially available and building recurring public- and private-sector workloads. Outcomes will depend on delivery across regulation, infrastructure, procurement and demand.

3.3

Kazakhstan's AI infrastructure base and expansion pipeline

Kazakhstan is moving from a concentrated domestic hosting market towards a multi-layer AI-infrastructure platform. Its emerging ecosystem combines commercial data-centre capacity, sovereign and enterprise cloud, domestically operated accelerated computing, new colocation and hyperscale facilities, and the flagship Data Center Valley development in Ekibastuz.

The country already has a tangible operating base. Commercial rack utilisation reached 91% in 2025, indicating strong use of existing capacity. Alem. Cloud and Al-Farabium together provide 896 NVIDIA H200 GPUs, serving public institutions, researchers, start-ups and commercial customers. Both systems rank among the world's 125 most powerful supercomputers, giving Kazakhstan internationally benchmarked high-performance computing capability.

Data Center Valley represents the largest proposed expansion. The platform is designed for phased development from an initial 300 MW to 1 GW. In June 2026, the Government announced a USD 10 billion package of agreements with Firebird and NVIDIA-related participation, covering a computing cluster based on 100,000 advanced GPUs, including NVIDIA GB300 and Vera Rubin technologies. According to government estimates, the project could generate at least USD 3 billion in annual export revenue once implemented.

Together, these projects extend Kazakhstan's infrastructure beyond standalone facilities towards a regional computing platform. Implementation will require investment in power generation, grid connections, cooling, fibre infrastructure, equipment, operational expertise and anchor demand. The following subsections examine the existing operating base, the near-term project pipeline, Data Center Valley and the power-system conditions for further expansion.

Commercial data-centre capacity and domestic accelerated computing

3.3.1

Kazakhstan's commercial data-centre market expanded steadily between 2022 and 2025. IDC estimates that installed commercial rack capacity increased from approximately 3,310 racks to 4,009 racks, while utilised rack capacity rose from 2,571 to 3,648 racks. Utilisation reached 91% in 2025, compared with 77.6% in 2022.[42]

IDC projects installed capacity of approximately 5,100 racks in 2026 and 7,100 racks by 2029. Used capacity is projected to reach 6,323 racks by 2029, with utilisation remaining above 80%. The market therefore combines continued capacity expansion with a relatively high level of existing use.

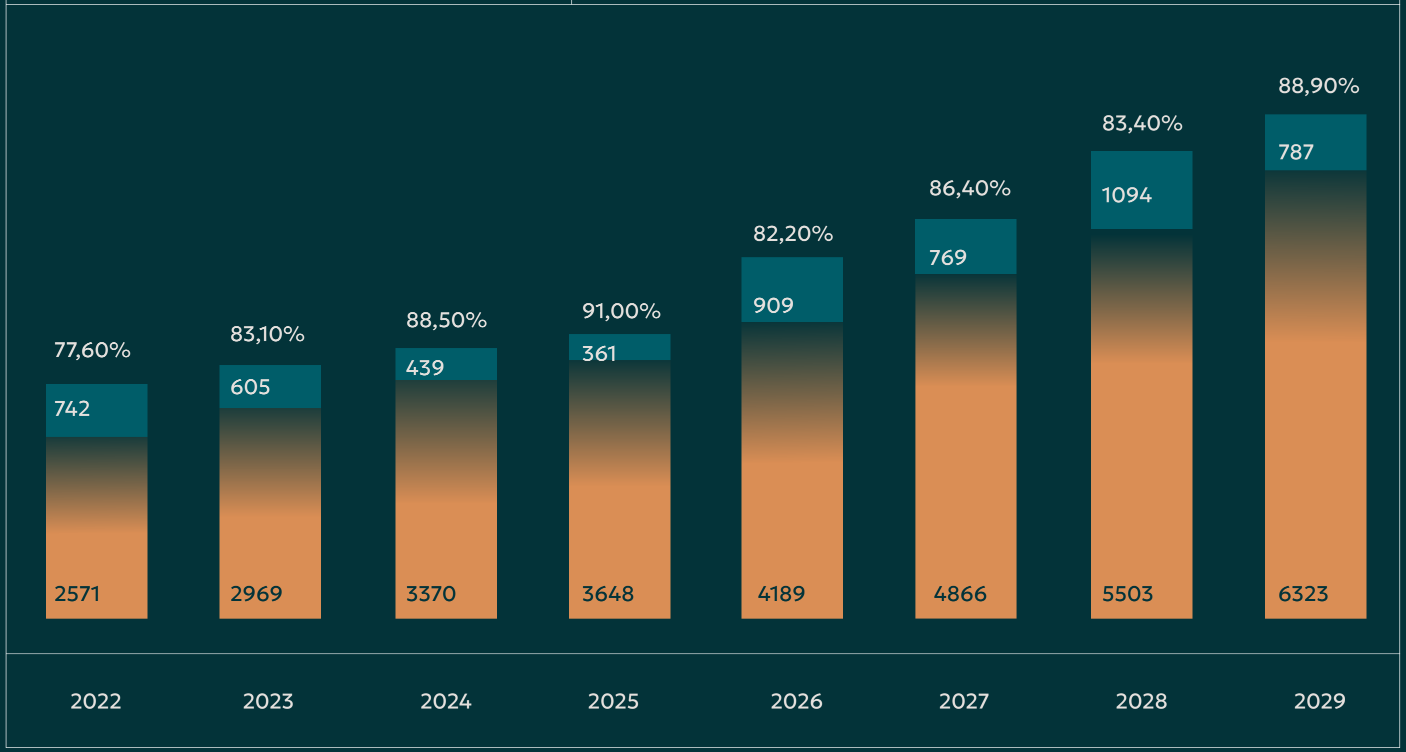
[42] IDC — Kazakhstan Public Cloud Services and Commercial Data Centre Market Review and Forecast

Figure 9

Source: IDC, Kazakhstan Public Cloud Services and Commercial Data Centre Market Review and Forecast, February 2026;

Commercial data-centre rack capacity and utilisation, 2022–2029

Used capacity
Free capacity
Commercial data centre rack utilisation rate, 2022–2029



The commercial market provides colocation, hosting, storage, disaster-recovery and cloud services. Its customer base includes public institutions, financial organisations, telecommunications operators and other enterprises. The next group of projects broadens this base through managed cloud services, GPU access and facilities designed for higher-density computing.

Domestic accelerated computing

Kazakhstan operates two large computing systems based on NVIDIA H200 accelerators: Alem. Cloud and Al-Farabium. Together, they provide a documented domestic base of 896 H200 GPUs.

Alem.Cloud is located in Astana and operated by JSC National Information Technologies (the operator of the "e-government" information and communication infrastructure and the National Artificial Intelligence Platform (NAIP) in Kazakhstan). The system uses HPE Cray XD670 servers and contains 496 NVIDIA H200 GPUs. Its computing resources are intended for government bodies, universities, research organisations, start-ups and companies working with AI, big data and scientific computing.[43] In the June 2026 TOP500 ranking, Alem.Cloud was placed 104th globally, with measured High Performance Linpack performance of 20.48 petaflops and theoretical peak performance of 25.34 petaflops.[44]

Al-Farabium is located in Almaty and operated by Kazakhtelecom as a commercial GPU cloud. The system includes 400 NVIDIA H200 GPUs, 100 TB of RAM, 56 TB of GPU memory and 6 PB of high-performance storage. Customers can access virtualised GPU configurations and bare-metal servers through hourly billing.[45] Al-Farabium ranked 122nd in the June 2026 TOP500 list, with measured performance of 17.93 petaflops and theoretical peak performance of 26.84 petaflops. [45]

The two systems have different service models. Alem.Cloud combines public-sector, academic, startup and enterprise access, while Al-Farabium provides accelerated computing through a commercial cloud platform. This creates a domestic computing layer alongside Kazakhstan's conventional data centre and cloud infrastructure.

[43] National Information Technologies — National Supercomputer Alem.Cloud

[44] TOP500 — June 2026 ranking

[45] Kazakhtelecom — Al-Farabium GPU cloud

Near-term cloud and colocation expansion 3.3.2

Projects in Almaty and Astana are expanding Kazakhstan's enterprise-cloud, sovereign-cloud and commercial colocation capacity. Their contribution to national AI-infrastructure readiness should be assessed according to disclosed capacity, financing status, construction progress, power availability and expected commissioning dates. The projects remain at different stages of development and should not be aggregated as operational capacity.

Beeline Hyper Cloud

Beeline Kazakhstan, part of VEON Group, began construction of its Hyper Cloud data centre in Almaty in December 2025. The Tier III facility is intended to provide sovereign cloud, enterprise cloud and AI-compute services, with customer data stored and processed within Kazakhstan.[46]

The announced service portfolio includes IaaS, PaaS, BaaS, GPU-as-a-Service and Security-as-a-Service. VEON targets commercial launch by the end of 2026. The company has not publicly disclosed the facility's IT capacity, rack count or investment value, limiting direct comparison with the larger data-centre projects under development.

[46] VEON — Beeline Kazakhstan breaks ground for Hyper Cloud data centre

AKASHI Data Center

AKASHI is developing a Tier IV data-centre campus in Astana. In June 2026, AKASHI and the Eurasian Development Bank agreed key terms for USD 90 million of financing for a first phase comprising 10 MW of IT capacity and 520 server racks. The EDB announcement describes a wider project of 41 MW and 3,520 racks, with total planned investment of approximately USD 230 million.[47]

Current company materials describe a wider four-building campus with planned capacity of ~4,200 racks and up to 100 MW. The facility is designed for enterprise, government, cloud, AI and high-performance computing workloads, including higher-density configurations and liquid cooling.

[47] Eurasian Development Bank — Financing for Kazakhstan's first Tier IV data centre

GK Hyperscale data centres

Kazakhstan has also signed an investment agreement with Singapore-based GK Hyperscale for two Tier III hyperscale data centres in the Akmola and Karaganda regions. The facilities are planned to provide a combined IT capacity of 200 MW, supported by announced investment of USD 1.5 billion. Construction was scheduled to begin in 2026, with the first module targeted for operation in 2027.

The project is intended to support cloud, AI and high-performance computing workloads, with part of the commercial proposition based on exports. The announced programme also includes additional investment in power-generation modernisation, wind capacity and energy storage, reflecting the need to develop large compute capacity together with firm electricity supply.

The agreement represents a potentially significant expansion of Kazakhstan's hyperscale pipeline. However, the 200 MW figure remains planned rather than commissioned capacity. Delivery depends on construction progress, dedicated power infrastructure, equipment procurement and the conversion of expected export demand into binding customer commitments.

Freedom Sovereign AI Hub

Freedom Holding announced an agreement with the Ministry of Artificial Intelligence and Digital Development in November 2025 to develop a sovereign AI hub in Kazakhstan. The proposed USD 2 billion programme is intended to use NVIDIA-based infrastructure at a site with 100 MW of available power. The project could add a domestically operated AI-compute platform focused on Kazakhstan and the wider Central Asian market.

Data Center Valley: large-scale AI infrastructure development

3.3.3

Data Center Valley is Kazakhstan's main large-scale AI infrastructure project. The site comprises a land reserve of approximately 1,400 hectares in Ekibastuz, near the GRES-1 power plant. It is designed to combine computing capacity, access to power, telecommunications, engineering infrastructure and investment incentives within a single industrial campus.[48]

The project is being developed through cooperation between the Government of Kazakhstan, Kazakhtelecom and U.S.-based AI-infrastructure company Firebird, with NVIDIA providing technology expertise. Agreements announced in June 2026 provide for up to USD 10 billion of investment and a computing platform incorporating as many as 100,000 advanced GPUs, including NVIDIA GB300 and Vera Rubin systems. The campus is intended to support AI training and inference, cloud services, high-performance computing and export-oriented digital workloads.[49]

The development is structured around a starting platform of approximately 300 MW, with phased expansion towards 1 GW. The first 125 MW data centre is scheduled to enter operation in the first half of 2027, followed by a second 125 MW facility in 2028. Subsequent facilities are planned between 2029 and 2033, potentially bringing the campus to as many as ten data centres.[50]

The commercial model combines energy access with prepared engineering infrastructure. The Ministry of Artificial Intelligence and Digital Development is developing a dedicated electricity-pricing mechanism intended to provide investors with long-term tariff visibility at approximately USD 0.025 per kWh. Investors are also expected to receive allocated land, water and telecommunications connections and access to a prepared engineering base.[51] Published technical parameters include an intended Tier III standard and a target power-usage-effectiveness ratio of 1.25.

Water requirements will depend on the cooling and heat-rejection systems selected for individual facilities. A water-intake point from the Satpayev Canal has been identified, and published regional planning information indicates a daily requirement of approximately 2,300 cubic metres for utility purposes, together with additional provision for fire safety.[50] Closed-loop liquid-cooling systems combined with dry coolers can limit direct water consumption, while evaporative and hybrid systems require a more stable supply. Cooling design, recycled-water use, wastewater management and water-usage-effectiveness targets should therefore form part of each facility's technical and environmental requirements.

The project offers several routes for private participation. Investors may install computing equipment in an operator-developed facility, construct their own data centre on an allocated site with prepared infrastructure, or commission a turnkey facility designed to specific technical requirements. The proposed investment regime includes exemptions from corporate income, property and land taxes, together with legal-protection and stability provisions.

Preliminary investor demand exceeds 100 MW, close to the planned capacity of the first facility. Converting this interest into long-term capacity contracts will support financing for subsequent phases and utilisation of the infrastructure.

At full scale, Data Center Valley would create a power-linked computing platform serving domestic users and export markets for cloud, AI and high-performance-computing services. The Government estimates that the project could generate at least USD 3 billion in annual export revenue. Implementation will require data-centre construction to remain aligned with power generation, grid capacity, cooling infrastructure, international connectivity and contracted customer demand.[52]

48 Prime Minister of Kazakhstan — Implementation of the Data Center Valley project

49 The Astana Times — Kazakhstan advances 1 GW Data Center Valley project in Ekibastuz

50 Kapital.kz — First Data Center Valley facility planned for 2027

51 Forbes Kazakhstan — Proposed electricity tariff for Data Center Valley investors

52 Prime Minister of Kazakhstan — Government, Firebird and NVIDIA sign $10 billion AI agreements

Power-system balance and capacity expansion

3.3.4

AI infrastructure requires more than sufficient electricity generation at the national level. Large GPU clusters need firm power at a specific location, high-capacity grid connections, stable power quality, redundancy and the ability to increase consumption as additional computing equipment is installed.

Current power balance

In 2025, Kazakhstan generated 123.1 TWh of electricity and consumed 124.6 TWh. Domestic generation was therefore approximately 1.5 TWh below consumption. At the end of the year, installed generation capacity stood at approximately 26.7 GW. Available capacity during periods of maximum demand was lower because of maintenance, equipment condition, seasonal constraints and reserve requirements.[53]

For the 2025–2026 heating season, KEGOC projected maximum demand of approximately 17.6 GW against available domestic generation of around 16.7 GW, corresponding to a peak-capacity gap of about 0.9 GW. This distinction between installed and available capacity is particularly relevant for data centres, which require continuous power, stable frequency and voltage, reserve supply and limited exposure to curtailment.[54]

Data Center Valley would create a new source of concentrated electricity demand. The project currently has access to 300 MW of power capacity and is designed for phased expansion to 1 GW. At continuous full load, 300 MW corresponds to approximately 2.6 TWh of annual electricity consumption, equivalent to around 2.1% of Kazakhstan's electricity generation in 2025. At 1 GW, annual consumption would reach approximately 8.8 TWh, or around 7.1% of 2025 generation. Actual demand will increase as individual facilities and computing equipment are commissioned and utilised.

[53] Prime Minister of Kazakhstan — Kazakhstan's energy system expected to reach a surplus by 2029

[54] Qazaq Green / EXIA — Kazakhstan Energy Outlook 2025

Generation expansion

Kazakhstan is implementing a programme of generation modernisation, expansion and new construction. Under the Energy Sector Development Plan to 2035, the Ministry of Energy expects more than 26.3 GW of additional capacity to be commissioned. The programme includes approximately 5 GW from the modernisation and expansion of existing plants, 10.5 GW of new conventional generation, 8.4 GW of renewable-energy projects and 2.4 GW of nuclear generation.[55]

By the end of 2029, the Ministry expects approximately 13.3 GW of capacity additions, including 7.4 GW of base generation and 5.9 GW of renewable-energy capacity. Of this amount, approximately 12.56 GW would come from new facilities and 0.74 GW would replace existing capacity. The Ministry expects the planned additions to cover overall electricity demand from early 2027 and subsequently establish a stable system surplus.

The contribution of new capacity to continuous data-centre supply will vary by technology. Thermal and nuclear generation can provide firm supply. Hydropower and gas-fired generation can also support system flexibility, depending on plant design, water availability and operating conditions. Wind and solar generation require balancing resources, storage or complementary firm generation to support continuous computing loads.

In parallel, the Ministry is implementing the National Project for the Development of Coal Generation Based on "Clean Coal" Technologies. The programme covers 19 projects with combined new and modernised capacity of approximately 7.8 GW. Eight projects involve new generating facilities, while 11 cover the expansion or modernisation of existing plants. Key projects include Ekibastuz GRES-3, a thermal power plant in Kurchatov and new combined heat and power plants in Kokshetau, Semey and Ust-Kamenogorsk.

These projects are intended to expand baseload supply and replace ageing capacity. At the same time, the emissions intensity of electricity supplied to data centres will affect the commercial positioning of export-oriented computing services, particularly among customers with renewable-energy procurement and emissions-reduction requirements.

The planned generation programme provides a broader basis for the expansion of AI infrastructure. Its relevance to individual data-centre projects will depend on the location and timing of new generation, transmission capacity, grid connections and the availability of firm power at each site.

[55] Kapital.kz — Kazakhstan plans more than 26 GW of additional capacity by 2035

Power supply for Data Center Valley

Data Center Valley is being developed with dedicated electricity infrastructure connected to the Ekibastuz energy hub. The project includes approximately 15 kilometres of transmission lines and step-down substations with an initial capacity of 300 MW and potential expansion to 1 GW. The acquisition of electricity-network assets required for a direct connection to the energy hub has also been completed.[56]

The proposed commercial framework is intended to provide investors with long-term electricity-price visibility at approximately USD 0.025 per kWh. Contracts with terms of 10 to 20 years are expected to apply take-or-pay principles, under which investors pay for reserved power capacity irrespective of actual consumption.

Take-or-pay contracts can improve revenue certainty for power and infrastructure providers and encourage investors to commission computing equipment on schedule. They do not by themselves determine how new generation, transmission connections, reserve capacity and balancing services will be financed. These elements will need to be incorporated into the commercial structure of individual facilities and development phases.

The Ministry of Artificial Intelligence and Digital Development and the Ministry of Energy are implementing a 70/30 mechanism for the modernisation and expansion of generating capacity. Under this structure, up to 70% of newly added capacity may be sold directly to major consumers, including data centres, while the remaining 30% would be supplied to the national power system.

Direct connection and the 70/30 mechanism represent two distinct approaches. Direct connection uses existing generation and dedicated network infrastructure at the Ekibastuz energy hub, while the 70/30 mechanism links access to newly added capacity with private investment in generation modernisation or expansion.[57]

[56] Prime Minister of Kazakhstan — Inspection of the Data Center Valley project in Ekibastuz

[57] Ministry of Artificial Intelligence and Digital Development — Data-centre power-supply mechanism

Grid capacity and system flexibility

Large AI facilities require redundant power paths, on-site backup generation, uninterruptible power systems and tested switching arrangements. Under the intended Tier III standard, each capacity component and distribution path can be removed on a planned basis for maintenance or replacement without interrupting the critical IT load. Facility-level resilience complements, rather than replaces, firm grid supply and adequate reserve capacity.

Broader grid reinforcement is progressing in parallel. KEGOC plans to complete the South Kazakhstan reinforcement and West Kazakhstan interconnection projects in 2027, increasing transmission capacity in the south and integrating western generation into the national grid. Its longer-term pipeline includes approximately 7,000 kilometres of additional transmission lines through 2035.[58]

[58] KEGOC — Power-grid development and reinforcement projects

Assessment

Kazakhstan's generation and transmission pipeline offers a route for supplying the first large AI-computing facilities while expanding the wider power system. The initial phases of Data Center Valley are structured around existing generation at the Ekibastuz energy hub, a direct grid connection and dedicated transmission and substation infrastructure.

Expansion towards 1 GW will require data-centre commissioning to remain aligned with additional generation, transmission capacity, reserve supply and balancing resources. The 70/30 mechanism provides a framework for linking part of this demand to private investment in generating capacity.

Connectivity and the role of local AI inference

3.4

Reliable, high-capacity connectivity is an important enabling condition for AI infrastructure. Networks connect users and applications with computing resources, support the transfer of datasets and allow digital services to operate across domestic and international markets. Bandwidth determines the volume of traffic that can be carried, while latency, route diversity and network reliability affect service performance and continuity.

Kazakhstan does not currently host a full public cloud region operated directly by AWS, Microsoft Azure or Google Cloud. Workloads using these platforms may therefore be processed in Europe, the Middle East or Asia.

The telecommunications infrastructure planned for Data Center Valley is intended to reduce network latency from approximately 80 milliseconds to 57 milliseconds, a level described by the Government as comparable to major European traffic-exchange hubs.[59] Achieving this target would strengthen the commercial proposition of Kazakhstan's emerging computing infrastructure by improving access to regional and international users.

Against this background, domestic network use is expanding. Fixed wired broadband traffic reached 7.08 billion GB in 2025, approximately 2.5 times its 2019 level.[60] This is not a measure of AI demand but indicates the increasing scale of data-intensive activity supported by Kazakhstan's digital infrastructure.

[59] Prime Minister of Kazakhstan — Implementation of the Data Center Valley project

[60] Bureau of National Statistics of Kazakhstan — Information and communication technology statistics

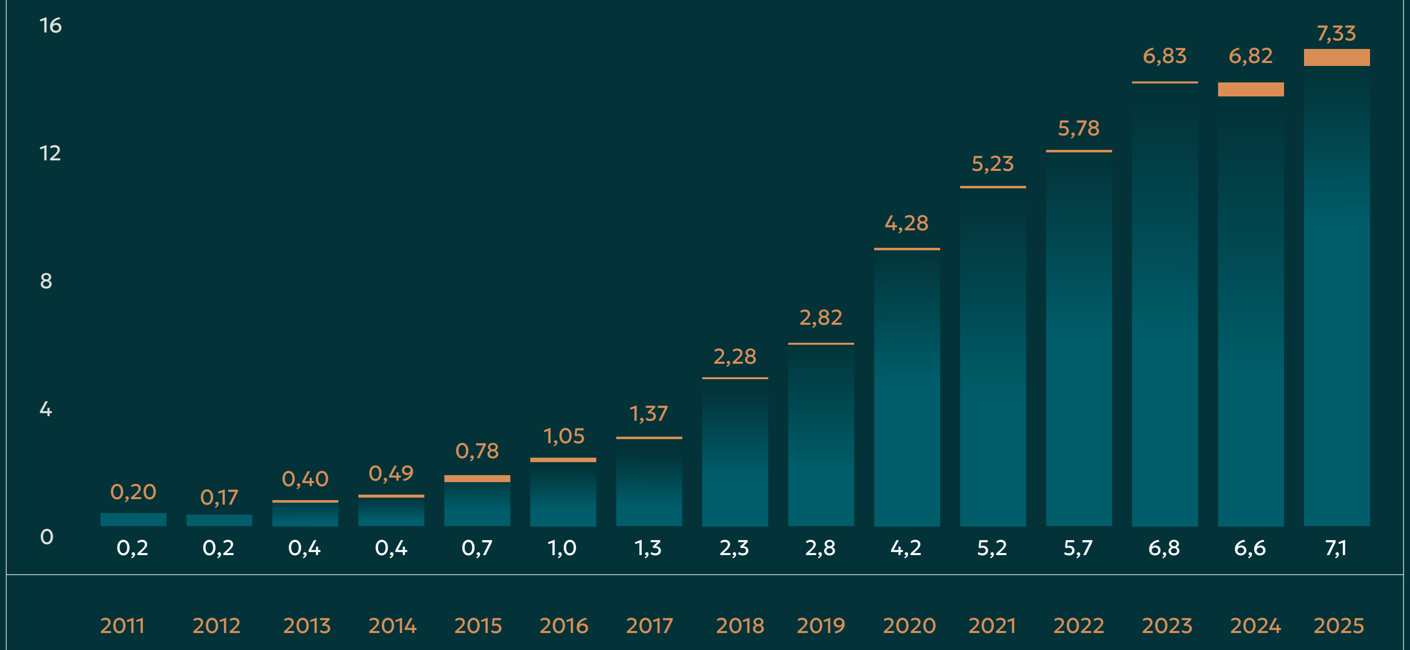
Figure 10

Source: Bureau of National Statistics of the Republic of Kazakhstan; AIFC analysis

Fixed broadband data traffic in Kazakhstan, 2011–2025

Wired broadband
Fixed wireless broadband
Total fixed broadband

Unit: billion GB



Domestic and international backbone capacity is expected to strengthen through the West–East fibre-optic hyper highway. The planned 4,511-kilometre route will connect Aktau with the Khorgos border crossing. By August 2026, more than 2,200 kilometres had been laid. The network is expected to provide capacity of at least 12 Tbps and enter operation by the end of 2026.[61]

At Aktau, the route will connect with the approximately 380-kilometre Trans-Caspian fibre-optic cable to Sumgait. Installation of the subsea cable was completed in August 2026, with commercial operation expected by the end of the year. The system has a design capacity of approximately 400 Tbps.[62]

Together, the West–East hyperhighway and Trans-Caspian cable will expand available bandwidth, diversify Kazakhstan's international routes and improve network resilience. For AI infrastructure, this will facilitate access to domestic computing capacity, accelerate the movement of large datasets and strengthen the reliability of cloud and digital-services exports.

Improved connectivity therefore reinforces Kazakhstan's wider AI-infrastructure proposition. Combined with expanding data centre and GPU capacity, it provides the network foundation required to serve domestic demand and connect Kazakhstan's computing infrastructure with regional and international markets.

[61] Ministry of Artificial Intelligence and Digital Development — West-East fibre-optic hyper highway

[62] Submarine Networks — Installation of the Trans-Caspian fibre-optic cable completed

3.5 The research base and human capital: from talent formation to applied AI

AI-infrastructure readiness depends not only on data centres, power and computing equipment, but also on the institutions and people able to develop, adapt, deploy and operate AI systems. This capability extends from foundational research and language-model development to software engineering, data science, cloud operations, cybersecurity and sector-specific implementation. Kazakhstan has begun to establish a visible research and education base, but advanced capability remains concentrated and the depth of the specialist workforce is not yet fully measurable.

Research and model-development capability

The Institute of Smart Systems and Artificial Intelligence at Nazarbayev University is the country's most established institutional anchor for AI research. Founded in 2019 and elevated to an autonomous research institution in 2024, ISSAI has developed datasets, models and deployment tools adapted to Kazakhstan's linguistic and operating environment. Its relevance extends beyond academic output: the institute has built local experience across data collection, model training, evaluation, speech technologies and hardware-supported inference.[63]

KazLLM is the most prominent example of this capability. ISSAI developed 8-billion and 70-billion-parameter versions using a multilingual training dataset comprising more than 150 billion tokens across Kazakh, Russian, English and Turkish. The models are based on Meta's Llama architecture and were trained using NVIDIA H100 cloud infrastructure. They were released in December 2024 under source-available under a non-commercial licence[64]. KazLLM therefore demonstrates domestic capability in multilingual data preparation, model adaptation, training and evaluation.

[63] Institute of Smart Systems and Artificial Intelligence — Official website

[64] Institute of Smart Systems and Artificial Intelligence — KazLLM

Expansion of the education pipeline

Human-capital development is becoming a broader national policy priority. Kazakhstan has announced an ambition to train one million people in AI-related skills over five years[65], including schoolchildren, students, civil servants and entrepreneurs. This programme can expand AI literacy and accelerate adoption across public administration, education and business. Broad training is relevant to infrastructure demand because the value of local computing capacity ultimately depends on the number of institutions and professionals able to use AI tools in everyday workflows.

University provision is also expanding from traditional computer-science education towards more specialised AI training. As of January 2026, 95 universities had incorporated AI-related disciplines into their curricula, while 30 higher-education institutions were implementing 42 dedicated programmes in artificial intelligence. Since 2025, practical AI skills have also been introduced across the wider higher-education curriculum, and 54% of faculty members had completed AI-related professional-development courses.[66]

[65] Astana Hub — Kazakhstan plans to train one million citizens in AI over five years

[66] Prime Minister of Kazakhstan — AI can increase the productivity of 70% of Kazakhstan's workforce

Integrated institutions

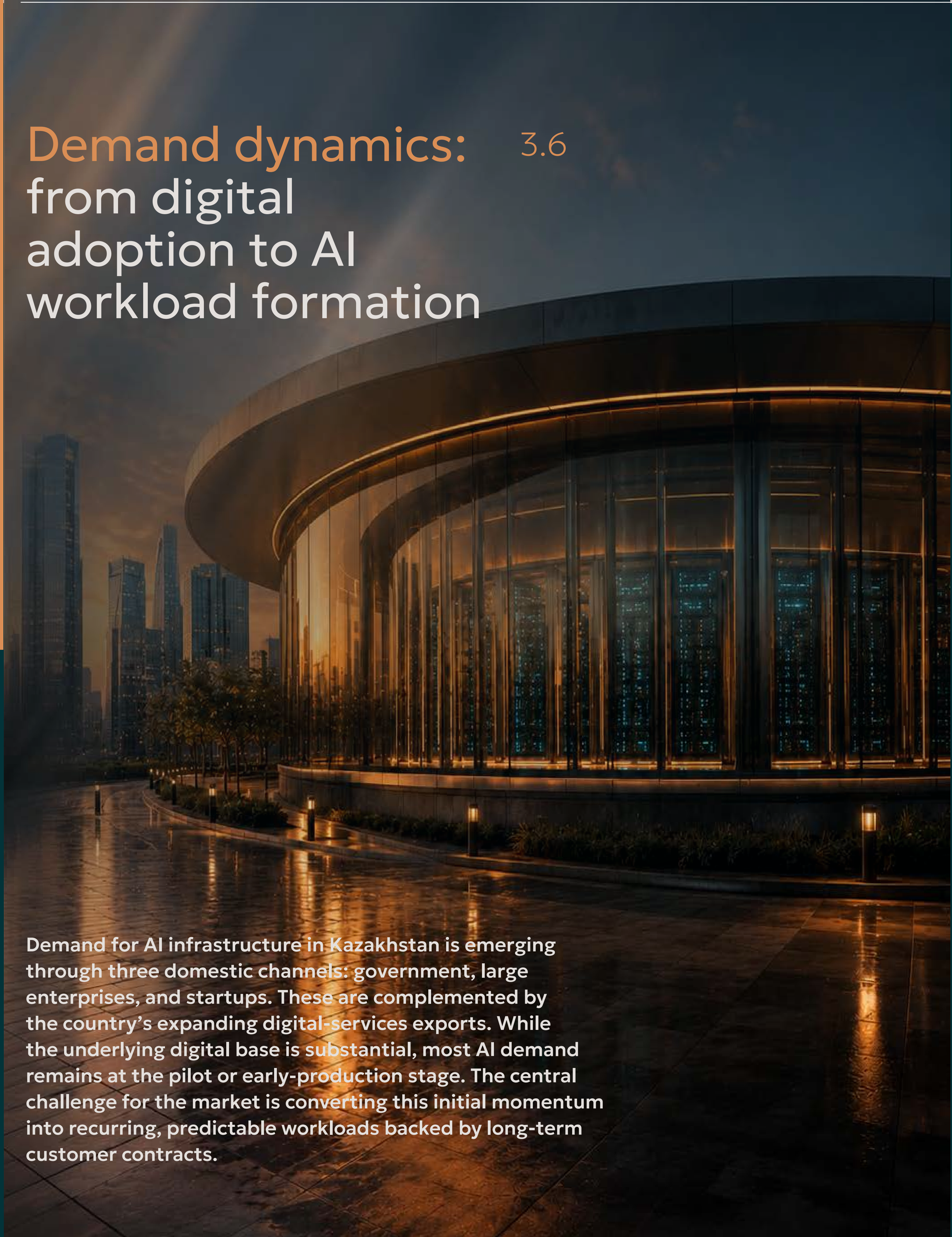
Alem.ai, the International Artificial Intelligence Centre inaugurated in Astana in 2025. Rather than operating only as a training facility, alem.ai is intended to bring together education, research, startups, government initiatives and technology partners within a common ecosystem. This model is relevant because it can shorten the path between training, experimentation and deployment by connecting learners and researchers with infrastructure, mentors, potential customers and investment institutions.[67]

Qazaq AI Research University represents the most specialised education component of this emerging ecosystem. Located within Astana's EXPO and alem.ai cluster, QAIRU opened admissions for its inaugural bachelor's cohort in June 2026. Its announced programmes include AI and Machine Learning at bachelor's level and AI+X, Applied AI Systems, and Business Analytics and AI at master's level. The university also intends to provide students and researchers with access to national computing infrastructure, including supercomputing resources, the alem.cloud platform and national language models, while linking coursework to industry and startup projects.[68]

[67] The Astana Times — Kazakhstan seeks to attract 10,000 AI specialists annually

[68] Qazaq AI Research University — Official website

Taken together, these initiatives point to a growing AI capability base in Kazakhstan. Their relevance to the infrastructure agenda is direct: local compute and cloud capacity become more valuable when they are used for fine-tuning, inference, applied experimentation and public- or enterprise-level deployment. Human capital is therefore not only a supporting condition, but also a source of future demand for sovereign, regulated and enterprise AI workloads.



Demand dynamics: from digital adoption to AI workload formation 3.6

Demand for AI infrastructure in Kazakhstan is emerging through three domestic channels: government, large enterprises, and startups. These are complemented by the country's expanding digital-services exports. While the underlying digital base is substantial, most AI demand remains at the pilot or early-production stage. The central challenge for the market is converting this initial momentum into recurring, predictable workloads backed by long-term customer contracts.

Government as an initial demand anchor

The public sector is likely to provide the earliest base of domestically anchored AI demand. Kazakhstan already has an extensive digital-government platform: approximately 90% of public services are available online, citizens received more than 51.5 million online public services in 2025, and eGov.kz has more than 15.1 million registered users.[69]

International benchmarks confirm the maturity of this digital base. Kazakhstan ranked 24th among 193 countries in the 2024 United Nations E-Government Development Index[70] and was classified in Group A—"extensive GovTech maturity"—in the World Bank's 2025 assessment.[71] These indicators demonstrate established government systems and digital-service channels, although they do not directly measure AI-compute demand.

AI adoption is beginning to add new workloads to this platform. More than 50 AI agents and services are reportedly under development across government bodies, the eGov AI assistant has been used more than 2.3 million times, and AlemGPT is being tested as a multi-agent public-service platform. Such applications can create recurring demand for inference, document processing, multilingual interaction and workflow automation.[72]

Public-sector applications involving sensitive personal or administrative data may favour domestic infrastructure because of security, data-governance and operational-control requirements. Government will become a meaningful anchor customer, however, only when pilot systems move into production and are supported by defined service levels, multi-year procurement and measurable capacity commitments.

69 eGov.kz — Results of digital public services in 2025

70 eGov.kz — Kazakhstan's position in the UN E-Government Development Index

71 World Bank — GovTech Maturity Index 2025 Update

72 Qazinform — Kazakhstan develops more than 50 AI-powered services for citizens

Enterprises as the test of commercial viability

Large enterprises will determine whether Kazakhstan can sustain a commercially viable AI-infrastructure market. The strongest potential demand lies in sectors with large proprietary datasets, complex operations and regulatory requirements, particularly banking, telecommunications, energy, mining and logistics. Relevant applications include fraud detection and compliance analytics in finance, network optimisation in telecommunications, and predictive maintenance and operational monitoring in industry.

Kazakhstan's public-cloud market increased from KZT 59.3 billion in 2022 to KZT 117.6 billion in 2025 and is forecast to reach approximately KZT 257.2 billion by 2029.[73] Official statistics indicate a broader base of organisational technology spending. Total ICT expenditure reached KZT 800.3 billion in 2025. External IT services and specialists accounted for KZT 352.3 billion, or 44% of the total, while in-house software development and licensed software represented a further KZT 122.1 billion and KZT 103.1 billion respectively.[74]

Material AI-infrastructure demand arises only when enterprises integrate models into core operations, connect them to proprietary data and require predictable capacity, security and service availability. The commercial test is therefore not whether companies experiment with AI, but whether these applications generate sustained workloads and sufficiently firm commitments to support infrastructure investment.

[73] Apple Hills — Cloud Study Kazakhstan 2026

[74] Bureau of National Statistics of Kazakhstan — ICT dynamic tables

Startups as a source of future workloads

Kazakhstan's startup ecosystem is expanding the range of AI applications but remains a relatively small and fragmented infrastructure customer. The country had more than 100 AI startups in 2025, while reported venture investment in AI increased from approximately USD 14 million in 2023 to USD 73 million in 2025, accounting for more than half of total venture investment.[75]

Most startup activity is concentrated in the applied software layer (enterprise SaaS, EdTech, MedTech, and localized marketing tools). These players generate demand primarily for model fine-tuning, adaptation, and inference. Individually, they lack the capital and scale to anchor major data centre investments or commit to long-term GPU contracts.[53]

Startups will become more relevant infrastructure customers where domestic compute is competitively priced, reliably accessible and connected to government, corporate and research opportunities. Stronger links between these groups would increase the likelihood that locally developed applications are also tested, deployed and scaled on infrastructure located in Kazakhstan.

[75] Astana Hub — Kazakhstan presents its first national AI development report

From demand signals to contracted utilisation

The structural outlines of Kazakhstan's AI demand are becoming clear: the public sector acts as a catalyst for sovereign workloads, large corporate enterprises represent the primary source of sustainable commercial revenue, and startups supply the innovation pipeline.

The immediate constraint is not a lack of digital activity, but rather the gap between pilot projects and long-term, contracted compute commitments. For the domestic AI infrastructure to become fully bankable for international investors, both state institutions and major enterprises must transition from "AI exploration" to "production-grade deployment." Until long-term capacity contracts are signed, the primary demand-side risk remains a structural mismatch where physical infrastructure expansion outpaces actual, contracted workload utilization.

Cross-border sovereign demand: data embassies

Kazakhstan has begun exploring the possibility of hosting digital assets belonging to foreign states. At the Digital Qazaqstan forum in March 2026, the Government presented the data-embassy concept as a means of hosting foreign digital assets while preserving their legal sovereignty and protection.[76] As discussed in Section 2.4, this would place Kazakhstan on the host side of the model and could extend potential demand beyond domestic government and commercial workloads.

Data embassies could create a specialised, export-oriented segment for Kazakhstan's digital infrastructure. Participating governments may require secure hosting capacity, resilient international connectivity, managed operations and regular continuity testing over extended periods. A successful arrangement could therefore generate recurring demand while strengthening bilateral digital cooperation and demonstrating confidence in Kazakhstan's legal, technical and cybersecurity environment.

The development of this market would be supported by agreements defining jurisdiction, authorised access, legal protection, encryption-key control, cybersecurity responsibilities, audit rights and termination procedures. Clear interaction with Kazakhstan's data-protection, localisation, criminal-procedure and national-security rules, together with technical separation of hosted environments, would provide the basis for accommodating sensitive foreign workloads.

[76] Kursiv — Kazakhstan proposes hosting other countries' sovereign data

From national readiness to regional platform logic

3.7

Kazakhstan has assembled most of the principal components required for AI-infrastructure development. These include operational GPU clusters, highly utilised commercial data centres, an expanding cloud and colocation pipeline, large power-linked projects, improving international connectivity, a growing research base and identifiable sources of domestic demand.

The maturity of these components nevertheless varies. Existing accelerated-computing clusters and commercial facilities are operational, while much of the planned increase in data-centre, generation and international-network capacity remains under construction or at the development stage. Kazakhstan's current position is therefore best understood as a combination of operating infrastructure and a substantially larger expansion pipeline.

The central task is to convert this pipeline into integrated and commercially utilised capacity. Data centres must be commissioned together with firm power, grid connections, cooling systems and telecommunications infrastructure. Computing equipment must be matched with operators, software environments and technical services that allow customers to deploy workloads without building the complete operating stack themselves.

Demand must develop in parallel. Government applications, regulated industries and large enterprises can provide the most immediate base of recurring workloads. Research institutions, startups and language-technology projects can broaden utilisation, but are unlikely individually to anchor large facilities. Converting experimentation and expressions of interest into production workloads, service-level requirements and longer-term capacity commitments will therefore be central to the commercial sustainability of new infrastructure.

The resulting market will depend on more than physical scale. Competitive electricity and compute pricing, reliable service, cybersecurity assurance, clear data-governance rules and access to skilled operators will influence whether users select infrastructure located in Kazakhstan. Phased development can help align capital expenditure with contracted demand and reduce the risk of commissioning capacity materially ahead of utilisation.

Kazakhstan's principal strength is the combination of assets that are often assessed separately: power availability, data-centre development, domestic compute, connectivity, digital-government capability and an emerging framework for infrastructure investment. Their economic value will depend on how effectively they are coordinated into reliable and commercially accessible services.

The next stage should therefore be assessed through operational indicators rather than announced capacity alone. Relevant measures include commissioned megawatts, installed and utilised GPU capacity, contracted customers, service availability, domestic workload migration and recurring digital-services revenue. Progress across these indicators would demonstrate that Kazakhstan is moving from infrastructure readiness towards a functioning AI-infrastructure market.

The following chapter places this national position within the wider Turkic region and examines where complementary capabilities and cross-border cooperation could create additional value.

CHAPTER 4.

The Organization of Turkic States:

from national capabilities to regional cooperation

Key numbers

USD 1.97 TRILLION
Combined GDP of OTS member states in 2025

USD 600 MILLION
Authorised capital of the Turkic Investment Fund

APPROXIMATELY 162,5 MILLION
Combined population of OTS member states in 2025

380 KM / UP TO 400 TBPS
Planned Trans-Caspian fibre-optic connection between Kazakhstan and Azerbaijan

END-2026
Target for elaborating and adopting the OTS AI and Digital Development Action Plan

Section summary

The Organization of Turkic States is developing a more structured agenda for digital cooperation, but it does not yet constitute an integrated AI-infrastructure market. Member states have developed different and potentially complementary capabilities across demand, compute, energy, cloud infrastructure and connectivity.

A model connecting compatible national platforms could enable selected research, language-technology and non-sensitive cloud services to be delivered across borders while preserving national control over sensitive systems. Kazakhstan could participate as one of several regional nodes, supported by its compute base, power-backed expansion potential, cross-Caspian connectivity and common-law framework for investment and contracting.

Key insights

1. Regional scale does not yet constitute an integrated market.
Economic scale and geographic reach coexist with nationally organised infrastructure, regulation, procurement and demand.

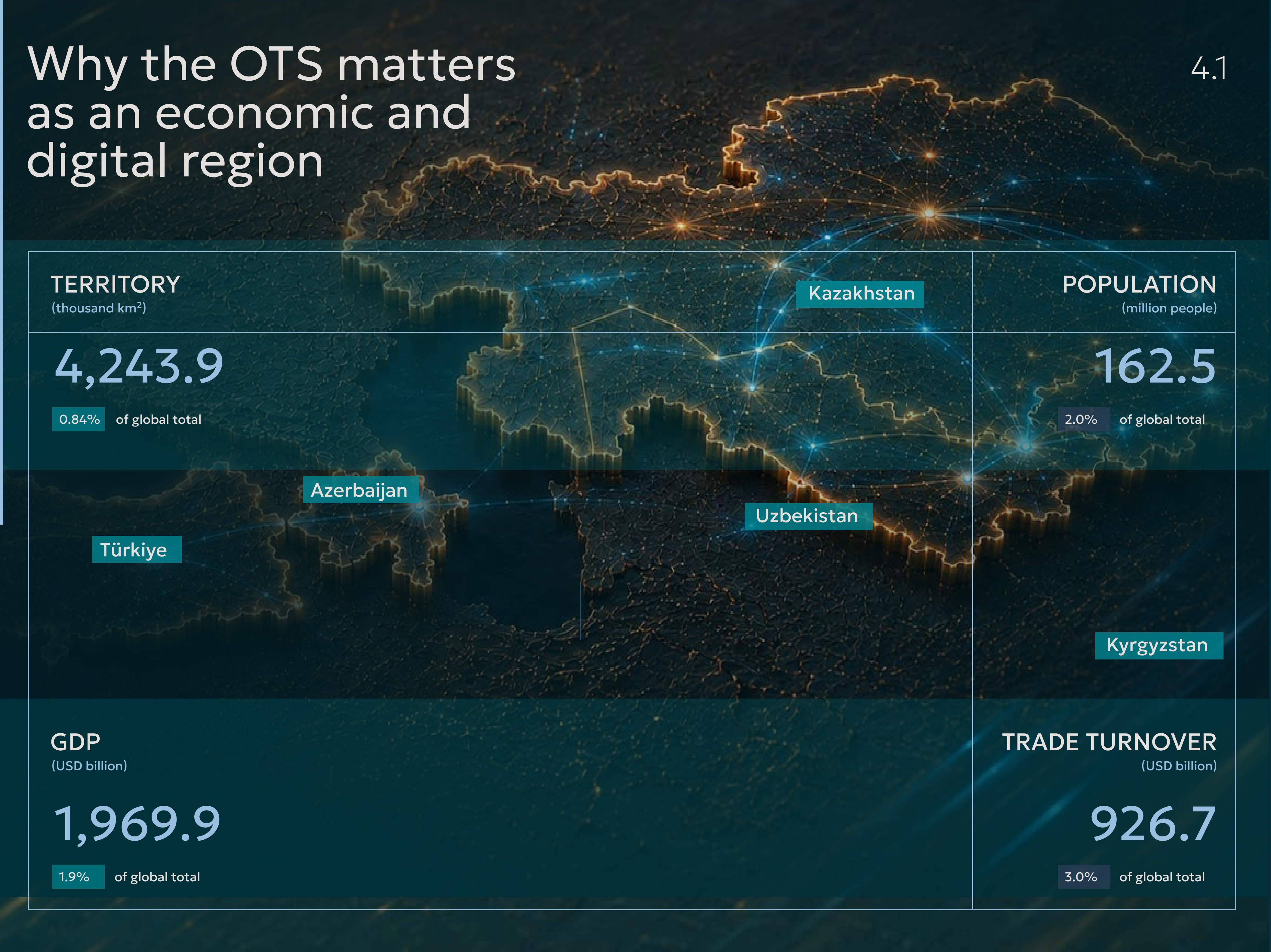
2. Regional value depends on identifiable cross-border workloads.
National capabilities become regionally relevant only where they can be linked to customers, legally permitted services and commercially viable delivery arrangements.

3. A federated model offers one practical path to cooperation.
National control over sensitive data and critical systems could be retained while selected computing, research, language-technology and non-sensitive cloud services are delivered across compatible national platforms. This is an analytical model rather than an adopted OTS framework.

4. Commercial execution is the decisive constraint.
Regional initiatives require operational connectivity, defined workloads, committed customers, enforceable agreements and financing structures capable of supporting long-term investment.

Why the OTS matters as an economic and digital region

4.1



TERRITORY (thousand km²)

4,243.9

0.84% of global total

POPULATION (million people)

162.5

2.0% of global total

GDP (USD billion)

1,969.9

1.9% of global total

TRADE TURNOVER (USD billion)

926.7

3.0% of global total

The Organization of Turkic States brings together Azerbaijan, Kazakhstan, Kyrgyzstan, Türkiye and Uzbekistan across a corridor extending from Türkiye through the South Caucasus and the Caspian region to Central Asia. Established in 2009, the OTS provides an institutional platform for cooperation across trade, transport, energy, digital development and emerging technologies. Hungary, Turkmenistan, the Turkish Republic of Northern Cyprus and the Economic Cooperation Organization participate as observers.[54]

The five member states recorded approximately USD 1.97 trillion in combined nominal GDP in 2025. Türkiye accounted for the largest share, with GDP of approximately USD 1.41 trillion, followed by Kazakhstan at USD 305.9 billion, Uzbekistan at USD 152.3 billion, Azerbaijan at USD 75.9 billion and Kyrgyzstan at USD 22.6 billion. These economies differ materially in market size, income levels and growth dynamics, creating distinct sources of digital demand across the region.[77]

[77] Center for Analysis of Economic Reforms and Communication — Turkic Economic Review 2025

Table 7 — Sources: Turkic Economic Outlook 2025; Oxford Insights, Government AI Readiness Index 2025; ITU;

Economic and digital profile of OTS member states

Note: Internet-use figures refer to the latest available year and may not be fully aligned across countries

	Türkiye	Kazakhstan	Uzbekistan	Azerbaijan	Kyrgyzstan
GDP, 2025 (USD bn)	1,413.1	305.9	152.3	75.9	22.6
GDP per capita, (USD thsd)	19	17,5	4,7	7,5	3,2
Real GDP growth, 2025	3.6%	6.5%	7.7%	1.4%	11.1%
Government AI Readiness 2025 (rank)	53	58	62	69	119
Internet users (% of population)	89.8%	93.3%	89.5%	90.8%	92.0%
Lead AI authority	Ministry of Industry and Technology	Ministry of Artificial Intelligence and Digital Development	Ministry of Digital Technologies	Ministry of Digital Development and Transport	Ministry of Digital Development and Innovative Technologies
AI legal and regulatory framework	Draft AI Bill No. 2/2234 (2024)	AI Law No. 230-VIII (2025)	AI-specific amendments, Law No. LRU-1115 (2026)	No standalone law; standards- and strategy-based framework	Digital Code and secondary high-risk AI requirements

AI-related investment depends on the scale and location of digital demand, access to electricity and connectivity, cloud adoption, institutional capability and the ability to serve customers across national markets. The OTS brings together economies with different but potentially complementary starting points across these areas.

The region's geographic position adds a further strategic dimension. Collectively, the member states connect Central Asia, the Caspian region, the South Caucasus, Türkiye and European networks. The same corridor that supports trade and energy flows can increasingly carry fibre connectivity, data traffic, cloud services and other digital workloads. Stronger digital infrastructure could therefore complement the region's established transport and energy corridors.

The combined economic scale and geographic reach of the OTS create a basis for deeper digital cooperation. This does not imply a single addressable market: regulation, procurement and infrastructure remain primarily national. The practical opportunity lies in connecting complementary capabilities through defined cross-border projects, common initiatives and commercially viable digital services.

From strategic coordination to an implementation agenda

4.2

Against this regional backdrop, the OTS has progressively translated its long-term digital priorities into agreements, institutional proposals and time-bound deliverables. Since 2021, the cooperation framework has expanded across connectivity, digital trade, cybersecurity, artificial intelligence, research, innovation and skills.

Turkic World Vision 2040, adopted in November 2021, established the long-term direction. It called for stronger intra-regional digital connectivity, greater use of big data, and wider application of digital technologies and artificial intelligence to support the development of digital, green and smart economies. Cooperation on e-government, cybersecurity, telecommunications infrastructure and innovation has developed in parallel through the OTS ICT ministerial and working-group framework.[78]

The cooperation framework became more specific in 2024. The Turkic States Artificial Intelligence Forum was held in Bishkek on 9–10 October, bringing together government representatives, policymakers and technology experts. The Bishkek Summit Declaration adopted the following month commended the forum and the separate Bishkek Declaration on Artificial Intelligence adopted during the event. It also supported digital training platforms, work towards a common platform for cooperation on cyber incidents and cyber-attacks, and the development of common ICT standards to improve interoperability.[79]

At the Eleventh OTS Summit in Bishkek on 6 November 2024, the member states signed the Digital Economy Partnership Agreement and a memorandum of understanding on the development of the Turkic Large Language Model. The Digital Economy Partnership Agreement is intended to facilitate digital trade and electronic interactions, while the memorandum provides a framework for cooperation on Turkic-language AI technologies.[80]

The regional financing architecture has expanded through the Turkic Investment Fund, an international financial institution established within the OTS framework. The Fund has authorised capital of USD 600 million, including USD 420 million in paid-in capital and USD 180 million in callable capital. Its priorities include physical and digital infrastructure, innovation, regional connectivity and co-financing with national and international institutions. Its planned instruments include loans, equity, trade finance and co-financing. While no dedicated AI-infrastructure window has been announced, the Fund could provide a platform for future regional digital projects.[81]

In 2025, the Gabala Declaration linked the digital agenda more directly with industrial development. It promoted the integration of artificial intelligence, green and digital technologies and smart-manufacturing systems into the industrial strategies of member states, with the stated objectives of improving productivity, sustainability and regional competitiveness. This extended regional cooperation from digital-policy coordination towards industrial application.[82]

The Informal Summit held in Turkistan in May 2026 marked a further shift towards a more structured implementation agenda. Convened under the theme "Artificial Intelligence and Digital Development", the summit addressed digital transformation, innovation, artificial intelligence, connectivity and sustainable economic development.[83]

The priorities identified in the Turkistan Declaration can be grouped into four areas:[84]

Digital infrastructure and connectivity, including the Digital Silk Way, high-speed networks, data centres, cloud infrastructure and high-performance computing;

AI capability, research and skills, including further development of the Turkic Large Language Model, research cooperation, education and talent development;

Digital trust and governance, including cybersecurity, data protection, closer legal coordination and further work on a draft agreement for the mutual recognition of electronic digital signatures;

Applied use cases, including manufacturing, agriculture, healthcare, public services, education, smart cities and environmental monitoring.

The declaration also emphasised the importance of advancing and adopting the Digital Turkic World Concept. It welcomed proposals to establish an OTS Digital Innovations Center and Cyber Security Council and instructed the relevant authorities to develop these initiatives and possible formats for cooperation. These proposals represent potential additions to the OTS institutional architecture and remain subject to further development by the member states.

The Turkistan Declaration also introduced a clearer coordination process by instructing the OTS Secretary General to organise annual ministerial meetings and expert forums and to work with national authorities on the development and adoption of an Action Plan on Artificial Intelligence and Digital Development by the end of 2026.

Together, these initiatives mark a transition from broad strategic coordination towards a more implementation-oriented regional framework. The planned Action Plan provides an opportunity to consolidate the priorities identified in Turkistan into a focused portfolio of connectivity, computing, research and applied-AI initiatives. Projects with defined participants, users, financing arrangements and delivery mechanisms could establish a pipeline of regionally coordinated AI and digital initiatives.

[78] Organization of Turkic States — Turkic World Vision 2040

[79] Organization of Turkic States — Turkic States Artificial Intelligence Forum launched in Bishkek

[80] Organization of Turkic States — Bishkek Declaration

[81] Turkic Investment Fund — Institutional presentation

[82] Organization of Turkic States — Gabala Declaration

[83] Organization of Turkic States — Informal Summit convened in Turkistan

[84] Organization of Turkic States — Turkistan Declaration

The regional AI infrastructure landscape

4.3

The AI infrastructure landscape across the OTS is developing through distinct national models rather than a uniform regional market. As Kazakhstan's AI-infrastructure readiness is examined in detail in Chapter 3, this section focuses on the other OTS member states and the capabilities most relevant to regional comparison and cooperation.

Figure 11

Source: Data Center Map, accessed mid-2026

Directory-listed commercial data-centre presence across OTS member states, mid-2026

Note: Facility counts indicate visible commercial presence but do not measure installed capacity, utilisation, technical quality or suitability for AI workloads

The countries exhibit different strengths across five dimensions: customer demand, operational computing capacity, data-centre infrastructure, access to power and international connectivity.

Türkiye provides the region's largest customer base and is moving towards locally available hyperscale-cloud services. Kazakhstan combines operational accelerated computing with power-backed expansion potential. Uzbekistan is building a sizeable investment pipeline supported by national targets and investment incentives. Azerbaijan combines sovereign-cloud infrastructure with an emerging cross-Caspian data route, while Kyrgyzstan has focused on foundational connectivity and shared public-sector digital infrastructure.

Türkiye

4.3.1

the region's largest digital market and an expanding cloud-infrastructure base

Türkiye has the largest population and economy among OTS member states, providing the region's broadest domestic market for digital services. Its information technology market reached USD 20.8 billion in 2024, while the communications technology market was valued at USD 15.9 billion. Together, these segments support demand for cloud computing, cybersecurity, data storage and AI-enabled applications across both the private and public sectors.[85]

[85] International Trade Administration — Türkiye Digital Economy

[86] Turkish Statistical Institute — Survey on ICT usage and AI adoption in enterprises, 2025

[87] Investment Office of Türkiye — Türkiye unveils AI Action Plan 2026–2030

[88] Investment Office of Türkiye — Turkcell and Google Cloud to establish Türkiye's first hyperscale regional data centre

AI adoption is expanding, although it remains uneven across the business sector. In 2025, **7.5% of enterprises with at least ten employees** reported using at least one AI technology, up from 2.7% in 2021. Adoption reached **24.1% among enterprises with 250 or more employees,** compared with **6.6% among businesses employing 10–49 people.** The gap shows that AI use remains concentrated among larger enterprises, leaving substantial scope for wider adoption among smaller businesses.[86]

The Ministry of Industry and Technology coordinates national AI policy and ecosystem development, including data, infrastructure, research and skills. The Directorate General of National Technology and Artificial Intelligence focuses on AI applications, data governance and related regulatory work within government. Türkiye has not enacted a dedicated AI law: Bill No. 2/2234, submitted in June 2024, remains before a parliamentary committee.

The Türkiye Artificial Intelligence Action Plan for 2026–2030 places data centres and cloud infrastructure within a broader programme for AI development. It targets at least **1 GW of installed data-centre capacity by 2030** and seeks to mobilise at least **USD 10 billion**, predominantly from private-sector sources, for AI infrastructure. The plan also provides for at least 2% of public investment programmes to be allocated to AI projects and positions public procurement as an initial source of demand for domestic AI solutions. A National Data Library is intended to make at least **2,000 public datasets** available for research, entrepreneurship and public-sector use.[87]

Türkiye's cloud market is also attracting global technology providers. The planned Google Cloud region forms part of Google's broader **USD 2 billion investment programme in Türkiye over ten years** and will be developed in cooperation with Turkcell. Separately, Turkcell plans to invest **USD 1 billion** in data centre and cloud technologies.[88] These remain forward-looking commitments but represent a significant step towards locally hosted hyperscale-cloud services.

Overall, Türkiye provides the strongest demand-side benchmark among the OTS economies considered in this report. Its large domestic market, higher enterprise AI adoption and announced hyperscale-cloud investments distinguish it from Kazakhstan's more infrastructure-led proposition. However, the 1 GW target and associated investment commitments remain forward-looking and their impact will depend on project delivery, power availability and the pace at which local cloud demand materialises.

Key numbers

86.1 MILLION
Türkiye's population at the end of 2025

AT LEAST 1 GW
Targeted installed data-centre capacity by 2030

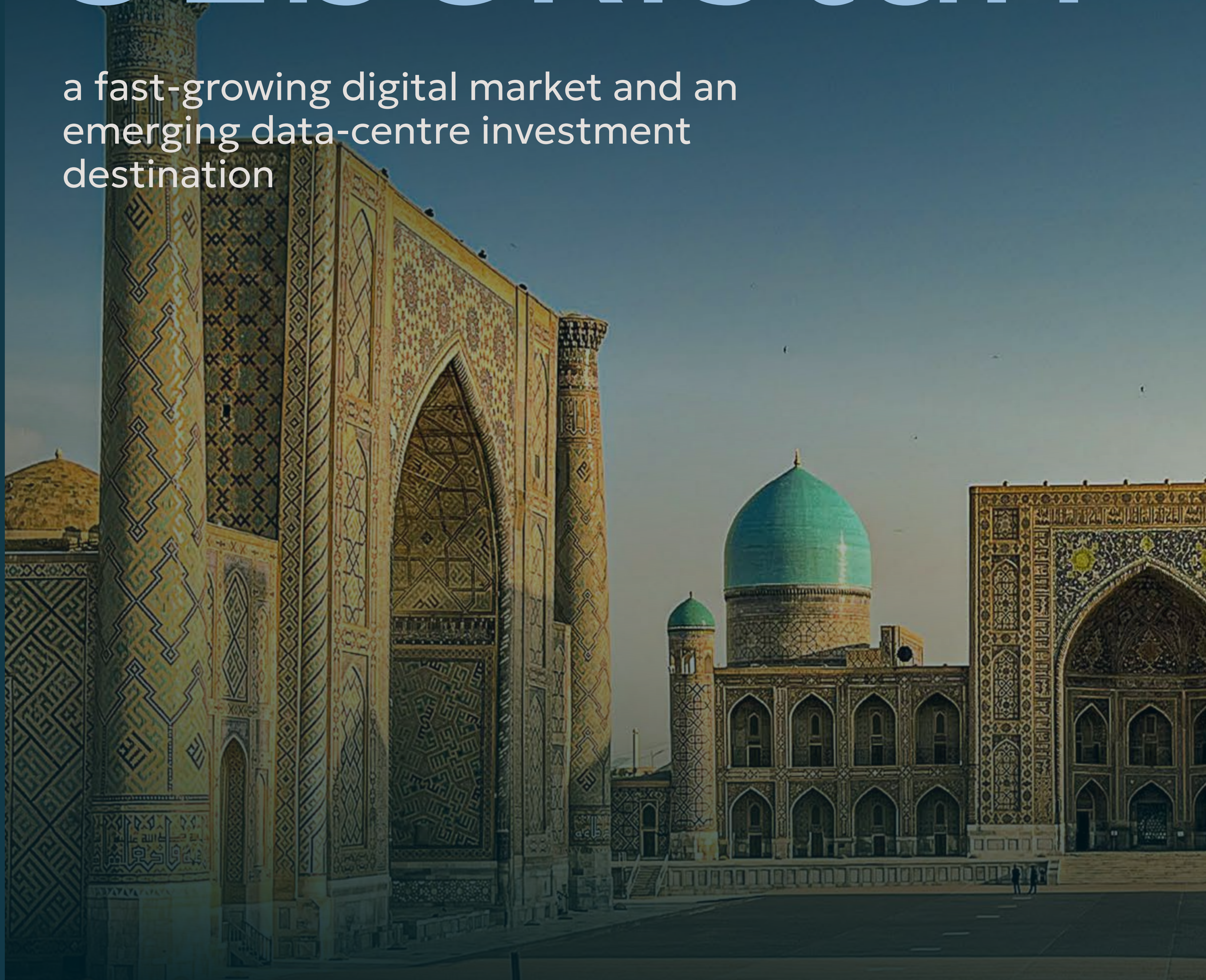
USD 2 BILLION
Google's planned ten-year investment in Türkiye, including a new Google Cloud region

7.5%
Share of enterprises using AI technologies in 2025

4.3.2

Uzbekistan

a fast-growing digital market and an emerging data-centre investment destination



Uzbekistan combines Central Asia's largest population with a rapidly expanding digital-services sector. Its permanent population reached **38.2 million at the end of 2025,** providing a sizeable potential user base for cloud computing, digital public services and AI-enabled applications. IT-services exports reached **USD 1 billion in 2025,** with the government targeting USD 5 billion by 2030. These trends are increasing domestic demand for computing, storage, connectivity and related digital infrastructure.

The government has established a five-year objective to develop **20 data centres with aggregate capacity exceeding 500 MW,** in cooperation with private investors. The target represents a substantial expansion from the country's current infrastructure base, although individual project schedules, capacities and financing structures remain at different stages of development.[89]

DataVolt represents the largest identified programme within this pipeline. The Saudi company plans a phased data-centre development of up to **500 MW,** associated with total investment of approximately **USD 3 billion.** Construction has begun on an initial **12 MW facility in Tashkent,** supported by around USD 150 million of investment. The national capacity objective and the DataVolt programme should not be treated as additive, as public disclosures do not clarify how much of the project is included within the wider 500 MW target.[90]

Uzbekistan is also using geographically targeted incentives to attract AI-infrastructure investment. In Karakalpakstan, projects investing more than **USD 100 million** in data centres and AI infrastructure may qualify for IT Park benefits, customs-duty exemptions on imported equipment and a preferential electricity tariff of **USD 0.05 per kWh.** At the June 2026 Tashkent International Investment Forum, the government stated that investors in the special zone would be exempt from taxes and duties until 2040, while road, electricity and water infrastructure would be financed by the state.

The institutional framework is led by the **Ministry of Digital Technologies,** supported by the Centre for the Development of Artificial Intelligence and the Digital Economy. The national AI Strategy to 2030, approved in October 2024, prioritises technical infrastructure, large datasets, skills and applied projects across sectors including finance, healthcare, transport, energy and agriculture. In January 2026, Uzbekistan enacted AI-specific legislative amendments under Law No. LRU-1115, introducing an initial binding framework for the use of AI rather than a standalone comprehensive AI act.

[89] Cushman & Wakefield — Uzbekistan data-centre market overview

[90] Gazeta.uz — DataVolt begins development of a data centre in Uzbekistan

Key numbers

38.2 MILLION
Uzbekistan's population at the end of 2025

MORE THAN 500 MW
Aggregate national data-centre capacity objective over five years

USD 3 BILLION
Planned phased DataVolt investment programme of up to 500 MW

USD 0.05/kWh
Preferential electricity tariff for qualifying AI-infrastructure projects in Karakalpakstan

For Kazakhstan, Uzbekistan represents an important regional benchmark, combining strong domestic digital demand with one of Central Asia's largest announced data-centre investment programmes. The country's policy emphasis differs from Kazakhstan's greater focus on large-scale energy-linked AI infrastructure and international compute exports.

4.3.3

Azerbaijan

a sovereign-cloud base and emerging cross-Caspian digital gateway

Key numbers

10.26 MILLION
Population as of 1 January 2026

EUR 43 MILLION
EIB Global financing for two new AzInTelecom data centres

380 KM
Planned Trans-Caspian submarine fibre-optic cable

UP TO 400 Tbps
Design capacity of the Trans-Caspian cable

Azerbaijan combines a relatively small domestic market with a strategic position on the digital corridor linking Central Asia, the South Caucasus and Europe. Its population exceeded 10.2 million in 2025,[91] while nationwide broadband coverage, expanding digital public services and the Government Cloud programme provide a foundation for demand in cloud computing, data storage and AI-enabled applications.

The operational base is centred on AzInTelecom's Government Cloud and AzInCloud platforms, which provide infrastructure and cloud services to public institutions, businesses and startups. In December 2024, EIB Global signed a **EUR 43 million** loan with AzInTelecom to finance two additional data centres. The project is intended to strengthen the resilience, security and efficiency of digital public services while expanding domestic hosting capacity.[92]

Azerbaijan's **Artificial Intelligence Strategy for 2025–2028,** approved in March 2025,[93] establishes a comprehensive roadmap for developing and applying AI technologies in the country. It prioritises data availability, computing infrastructure, skills, research, standards and the use of AI across government and the economy. The Ministry of Digital Development and Transport leads implementation, while the Digital Development Council, established in 2026, provides higher-level coordination across digitalisation, e-government, AI and innovation. Azerbaijan has not adopted a standalone comprehensive AI law, but it has begun introducing national AI standards and sector-specific governance measures.

Connectivity is a second pillar of the country's regional position. The Trans-Caspian cable under construction between Azerbaijan and Kazakhstan is designed to create a new east–west data route across the Caspian Sea. Public project descriptions indicate a route of roughly **380–400 kilometres** and design capacity of up to **400 Tbps.** Once commissioned, the cable could improve route diversity between Central Asia and Europe while potentially reducing latency for international cloud and data services.

> Azerbaijan's digital infrastructure is characterised by the combination of sovereign cloud services, expanding government digital platforms and its strategic location on the Trans-Caspian corridor. These assets strengthen domestic digital resilience while improving regional connectivity between Central Asia, the South Caucasus and Europe. However, Azerbaijan's current proposition is stronger in sovereign hosting and transit connectivity than in disclosed hyperscale or AI-compute capacity.

[91] State Statistical Committee of Azerbaijan — Population statistics

[92] European Investment Bank — €43 million loan to digitise Azerbaijan's public administration

[93] President of Azerbaijan — Artificial Intelligence Strategy of the Republic of Azerbaijan for 2025–2028

4.3.4

Kyrgyzstan

foundational public digital infrastructure

Key numbers

7.4 MILLION
Population as of 1 January 2026

58,899 KM
National fibre-optic network length at Digital CASA completion

212
Points of Presence connected through Digital CASA

Kyrgyzstan has one of the smaller domestic markets among the OTS economies assessed in this chapter, but it has made material progress in the foundational infrastructure required for digital public services. Its digital-infrastructure development is currently concentrated on national connectivity, government cloud services and shared public-sector platforms rather than large-scale commercial data-centre capacity.

The Digital CASA programme substantially expanded the country's connectivity base. By project completion, Kyrgyzstan's fibre-optic network had reached **58,899 km**, exceeding the project target of 39,030 km. The programme also deployed 30 backbone nodes, established two additional cross-border fibre connections with Kazakhstan and Uzbekistan, and connected all **212 Points of Presence**. A total of 1,432 state social facilities were directly connected under the project, while broader connectivity arrangements extended internet access to 3,882 social facilities nationwide. These investments expanded domestic coverage and strengthened the basis for delivering digital services beyond the largest urban centres.[94]

Shared public-sector infrastructure has developed alongside connectivity. By project completion, **21 government information systems from seven public bodies** had been migrated to the national G-Cloud platform, supporting more than **50 public services.**[95] This gives Kyrgyzstan an operational base for consolidating government systems, improving interoperability and reducing the need for each institution to maintain separate computing infrastructure.

The Digital Transformation Concept for 2024–2028 envisages a National AI Platform, model and dataset repositories, research support and Kyrgyz-language technologies.[96] The Digital Code subsequently established an initial legal framework for AI systems, including risk-management, transparency and data-quality requirements.[97]

[94] World Bank — Digital CASA: Kyrgyz Republic Implementation Completion and Results Report

[95] World Bank — Digital CASA: Kyrgyz Republic project report

[96] Digital Watch Observatory — Digital Transformation Concept of the Kyrgyz Republic for 2024–2028

[97] Ministry of Justice of the Kyrgyz Republic — Digital Code

> **Overall, Kyrgyzstan provides a useful regional comparison in digital public infrastructure rather than commercial AI compute. Its progress in national connectivity and shared government platforms differs from Kazakhstan's emerging emphasis on energy-linked, large-scale data-centre development. Publicly announced hyperscale and dedicated AI-compute project pipelines remain limited.**

Synthesis: what regional integration would require

4.4

The OTS has established a more structured agenda for digital cooperation, but its members do not yet constitute an integrated AI-infrastructure market. The region instead combines complementary national capabilities.

Türkiye provides the largest base of enterprise demand; Kazakhstan combines operational compute capacity with a growing pipeline of power-linked infrastructure projects; Uzbekistan is pursuing large-scale data-centre investment; Azerbaijan supports the development of cross-Caspian connectivity; and Kyrgyzstan has prioritised foundational digital infrastructure.

These capabilities are complementary rather than interchangeable, and no single country currently provides every component required for a competitive regional market. Complementarity alone, however, does not create integration. The principal gap is the absence of an operational and commercial framework through which infrastructure and workloads can be delivered reliably across national borders.

A more realistic long-term model is therefore a federated regional architecture rather than a unified regional cloud. Under this model, each country would retain control over national infrastructure, regulation and sensitive data, while selected research, language-technology, enterprise and cloud workloads could be delivered across compatible national platforms. Workload location would depend principally on data sensitivity, regulatory constraints, latency requirements and the potential benefits of regional scale.

Figure 12

Indicative workload-location matrix for cooperation across the Turkic region

NEED FOR NATIONAL CONTROL OF DATA AND PROCESSING	BENEFIT FROM CROSS-BORDER POOLING: LIMITED Little value from sharing infrastructure or demand	BENEFIT FROM CROSS-BORDER POOLING: STRONG Pooling can improve access, economics or utilisation
HIGHER Sensitive or regulated workloads	NATIONAL SOVEREIGN INFRASTRUCTURE Primary systems and sensitive data remain domestic ILLUSTRATIVE WORKLOADS • Critical government systems • Regulated finance and healthcare Protected foreign replicas may support continuity under specific legal arrangements.	FEDERATED NATIONAL PLATFORMS Shared tools; national data and operational control ILLUSTRATIVE WORKLOADS • National-language AI • Public-sector applications • Fraud analytics • Enterprise assistants Model components, evaluation tools and standards may be shared; sensitive data and core processing remain domestic.
LOWER Fewer sovereignty constraints	EDGE AND ON-SITE PROCESSING Primary systems and sensitive data remain domestic ILLUSTRATIVE WORKLOADS • Industrial control • Safety-critical autonomy Cloud can support model development, while operational processing remains on site or at the edge.	SHARED REGIONAL CAPACITY Strongest near-term case for cross-border delivery INITIAL APPLICATIONS • University and startup GPU access • Research computing • Regional language-model inference • Non-sensitive cloud services LATER, AS DEMAND MATURES • Model fine-tuning workloads Aggregating demand can improve utilisation and broaden access.

HOW TO READ THE FRAMEWORK
This is a categorical matrix, not a scatter plot. Workloads are grouped by delivery model; their position within a cell does not indicate a score or ranking.

The matrix distinguishes four broad delivery models. Critical government and regulated workloads are generally suited to national sovereign infrastructure, although protected foreign replicas may support continuity under specific interstate arrangements. Industrial-control and safety-critical workloads are more appropriately processed on site or at the edge because latency and resilience requirements limit the value of remote delivery.

Federated national platforms provide an intermediate model for workloads that can benefit from regional cooperation while retaining national operational control. National-language AI, public-sector applications, enterprise assistants and analytical tools may share selected models, datasets and technical components across borders, while sensitive data and core processing remain within the relevant national environment.

The most plausible initial applications for shared regional capacity are research computing, GPU access for universities and startups, Turkic-language model development and inference, selected non-sensitive cloud services and large-scale model fine-tuning. Aggregating these workloads across compatible regional nodes could improve utilisation, broaden access to advanced compute and reduce unnecessary duplication of specialised infrastructure.

Turning these opportunities into an operational market would require four enabling conditions: resilient cross-border connectivity; sufficiently compatible cybersecurity, cloud-assurance and data-governance requirements; clear allocation of operational and contractual responsibilities; and customers prepared to make credible capacity commitments. Larger regional clusters would become commercially viable only when these conditions support predictable utilisation and contractable revenue.

Within a federated regional architecture, Kazakhstan could participate as one of several infrastructure nodes, supported by its existing compute base, pipeline of power-linked infrastructure projects and planned expansion of cross-Caspian connectivity. Potential applications include shared research capacity, regional language-model inference and selected enterprise workloads. Kazakhstan's eventual role would depend on converting announced projects into operational capacity supported by firm power, reliable connectivity, competitive service terms and sustained customer demand.

Regional integration does not require all workloads to move across borders. Its value lies in identifying the workloads for which shared infrastructure, aggregated demand and compatible national platforms create greater value than isolated provision.

The next chapter examines the financing structures through which such projects could be delivered.

CHAPTER 5.

Capital Formation for AI Infrastructure

Section summary

AI infrastructure is shifting from a technology purchase to a long-duration infrastructure investment. More than USD 5 trillion of global AI and data-centre investment could be required by 2030, but capital will favour projects that convert technical capacity into predictable cash flows. Bankability depends on deliverable power, phased construction, contracted demand, credible operators and enforceable risk allocation—not headline megawatts or GPU counts.

Financing should follow the asset stack: long-term capital for land, power and facilities; shorter-tenor structures for GPUs; and recurring-revenue models for cloud and managed services. PPPs can anchor sovereign and public-service demand but cannot substitute for commercial discipline.

For Kazakhstan, the strongest near-term pathway is demand-led: sovereign cloud, regulated-sector workloads, GPU-as-a-Service and anchor-tenant facilities.

Key insights

1. AI infrastructure is becoming an infrastructure-finance market. Its physical foundations require long-duration capital, contracted revenue and disciplined risk allocation.

2. The asset stack is layered. Data-centre facilities, power, cooling, fibre, GPUs, cloud platforms and managed AI services each require different financing structures.

3. Kazakhstan's credible pathway is phased and demand led. Sovereign cloud, regulated workloads, GPU-as-a-Service and inference are stronger first-stage opportunities than frontier-scale training.

4. AIFC provides a familiar legal and institutional platform for international investors. Its common-law framework, independent court and arbitration centre, and regulated financial ecosystem provide established rules for investment and dispute resolution.

From capital requirements to financeable assets

5.1

AI infrastructure requires substantial capital before it can generate revenue. Land must be secured, grid capacity reserved, facilities constructed, cooling and fibre installed, and computing equipment procured before customers can consume cloud or AI services. The financing challenge is therefore not simply to mobilise capital, but to convert a proposed facility into an asset with sufficiently predictable cash flows to support debt and long-term investment.

This requires a distinction between technical capacity and financeable capacity. Technical capacity describes what a site may eventually support. Financeable capacity is the portion supported by deliverable power, secured land and permits, an executable construction plan, credible operators and contracted or otherwise demonstrable demand. A site with a long-term target of several hundred megawatts may initially have only a small financeable phase if the remaining power, customers and capital have not been secured.

AI infrastructure should also not be treated as a single asset class. It combines four layers with different economic lives, sources of revenue and risk profiles.

1. The first layer comprises land, grid connections, substations, fibre routes and other enabling infrastructure. These assets generally have long operating lives, but their value depends on site rights, permits, connection schedules and the availability of customers. A grid connection that cannot be delivered on time, or land without the necessary power and fibre, has limited value to an AI-infrastructure investor.

2. The second layer is the data-centre facility, including the building, electrical systems, backup power, cooling, physical security and internal connectivity. These assets can support long-term financing when their capacity is leased or reserved under contracts with credible tenants. Their principal risks are construction cost, completion timing, power availability, operating performance and tenant concentration.

3. The third layer consists of GPUs, servers, networking and storage. This equipment has a shorter commercial life than the facility in which it operates. Its value is exposed to utilisation, technological obsolescence, equipment availability, export compliance and uncertain residual value. Financing for this layer therefore generally requires shorter tenors, faster amortisation or dedicated equipment-refresh provisions.

4. The fourth layer consists of cloud platforms and managed AI services. Their value depends on software, cybersecurity, data governance, customer integration, technical support and service performance. Unlike a building or server, much of this layer may not provide conventional physical collateral. Its financing therefore depends primarily on recurring customer revenue and the operating capability of the service provider.

The financing structure should match these differences. Using long-term facility debt to finance rapidly depreciating computing equipment may create a maturity mismatch. Conversely, financing an entire data-centre campus with short-term equipment debt can expose a long-life asset to unnecessary refinancing risk. Where commercially and legally practical, the facility and computing equipment can be financed through separate packages or ring-fenced project vehicles while remaining operationally integrated.

The capital requirement should also be divided into three stages. Development capital covers feasibility studies, grid applications, design, permitting, customer development and vendor engagement. Construction capital funds the facility, power and cooling systems. Technology and refresh capital finances the initial computing equipment and its subsequent replacement. Each stage should have a defined source of funding rather than assuming that one financing facility will cover the project from development through operation.

For lenders, the quality of revenue is as important as its projected amount. A forecast based on general growth in AI adoption is not equivalent to contracted revenue. More financeable structures include long-term leases, minimum-capacity commitments, take-or-pay provisions, reserved GPU clusters and service contracts with minimum payments.

The central financing principle is therefore to match each asset layer with an appropriate source, tenor and security package.

How AI infrastructure is financed globally

5.2

AI infrastructure is rarely financed through a single instrument. A project may use sponsor capital during development, institutional equity and construction debt during the build phase, equipment finance for GPUs, and bonds or other capital-market instruments for subsequent expansion or refinancing.

The scale of the global buildout is increasing the importance of external capital. Goldman Sachs Research estimates that major technology companies could undertake more than USD 5 trillion of capital expenditure on AI and data centres between 2025 and 2030. Financing requirements of this magnitude may eventually encounter concentration limits in public credit markets, particularly where a small number of hyperscalers account for a growing share of corporate borrowing. Private infrastructure, real estate and credit markets are therefore expected to play a larger role alongside hyperscaler balance sheets.[98]

The financing method ultimately depends on what gives investors confidence that they will be repaid: the sponsor's financial strength, customer contracts, project-level revenue, equipment collateral, public risk-sharing or the performance of an operating portfolio.

[98] Goldman Sachs — Private markets expected to have a growing role in data-centre financing

01

Sponsor balance-sheet and corporate financing

Large technology companies can finance infrastructure through operating cash flows, corporate borrowing and finance leases. Under this model, investors and lenders rely primarily on the sponsor's overall financial strength rather than on the cash flows of an individual facility. The sponsor retains most construction, utilisation and technology risk but benefits from greater control over the timing and design of capacity.

Microsoft illustrates the scale of this model. In the fourth quarter of fiscal year 2026, the company reported USD 41 billion of capital expenditure, approximately two-thirds of which related to shorter-lived assets, primarily CPUs and GPUs. The company also reported adding 88 data centres during the fiscal year.[99] The distinction between shorter-lived computing equipment and longer-lived data-centre assets reinforces the need to finance different parts of the infrastructure stack on different terms.

For Kazakhstan, hyperscaler balance-sheet investment remains primarily an investment-attraction objective rather than an immediately available domestic financing channel. Participation is more likely where a project offers deliverable power, sufficient scale, credible operations, connectivity, regulatory certainty and identifiable regional demand.

[99] Microsoft — FY2026 fourth-quarter earnings

02

Contract-backed debt and private credit

Bank loans, private credit and asset-backed structures are increasingly being used to finance data-centre and computing capacity supported by customer contracts. Repayment may be secured by long-term leases, capacity-purchase agreements, computing equipment, project accounts and contractual guarantees. The Bank for International Settlements has documented the increasing use of both on-balance-sheet borrowing and separate financing vehicles for AI infrastructure.[100]

CoreWeave demonstrates how contracted demand can support financing for computing equipment. Following an agreement of up to USD 11.9 billion with OpenAI in March 2025, the company closed a USD 2.6 billion delayed-draw term loan linked to the equipment, hardware and cloud systems required to perform that contract.[101]

The model expanded during 2026 through additional facilities backed by high-performance-computing infrastructure and customer commitments. These transactions included an investment-grade facility supported by an associated customer contract, as well as subsequent facilities rated below investment grade. In August 2026, CoreWeave closed a further USD 2.6 billion facility with an approximate five-year maturity, while the underlying customer contracts averaged approximately three years. The structure therefore required lenders to accept renewal and re-leasing risk after the initial contracts expire.[102]

The transactions show both the potential and the limitations of contract-backed compute finance. Customer commitments and equipment collateral can increase debt capacity, but they do not eliminate exposure to customer concentration, contract renewal, GPU demand, technological obsolescence and residual value. Credit quality depends on the strength and duration of the contracts, the underlying customers, the collateral package and the operator's ability to recontract capacity.

For Kazakhstan, this channel would require enforceable customer contracts, clearly identified collateral, lender step-in rights and transparent arrangements for equipment maintenance, replacement and remarketing. Currency risk must also be addressed where equipment and financing costs are denominated in foreign currency but customer revenue is primarily received in tenge.

[100] Bank for International Settlements — AI infrastructure financing and financial stability

[101] CoreWeave — $2.6 billion secured debt financing facility

[102] CoreWeave — $2.6 billion loan facility expands financing flexibility for AI infrastructure

03

Project finance and institutional infrastructure capital

Project finance is most suitable where repayment can be linked to cash flows generated by a defined facility or project company. Financeable structures generally require secured land and permits, deliverable power, an executable construction plan, experienced sponsors and operators, and long-term commitments from credible tenants. Infrastructure and real-estate investors may provide equity alongside construction or project debt, particularly for the longer-lived facility layer.

The Meta–BlackRock El Paso venture announced in July 2026 illustrates this approach. Funds managed by BlackRock hold an 80% interest and Meta retains 20% in a venture developing a 1 GW data-centre campus with approximately USD 14 billion of development costs. The structure includes USD 12.5 billion of debt financing, while Meta will lease the full campus, initially under four-year leases with extension options, and provide residual-value guarantees with an initial aggregate threshold of approximately USD 13 billion.[103]

[103] Meta — New venture with BlackRock to develop El Paso data centre

The transaction separates ownership and financing of the long-lived facility from the hyperscaler's use and management of the asset. However, the institutional capital and debt remain strongly supported by Meta's tenancy and residual-value protection. The transferable lesson for Kazakhstan is therefore the alignment of asset ownership, customer contracts, debt maturity and residual-value risk, not the assumption that an uncontracted local project could obtain comparable leverage or pricing.

04

Equipment leasing and GPU finance

GPU clusters have a different financing profile from data-centre buildings. GPU and AI accelerator hardware depreciates faster than any previous IT asset class, with major generational refresh cycles occurring every 2 to 3 years.[104] Unlike a data-centre shell, a GPU cluster can lose economic value quickly if utilisation is weak, newer chips become available, export restrictions change, or maintenance and networking costs are underestimated. This makes equipment leasing, vendor finance, asset-backed lending and private credit particularly relevant, especially where financing is linked to reserved-capacity contracts or long-term customer commitments.

Equipment leasing can align financing terms with the shorter commercial life of GPUs and servers. Under residual-value lease structures, the lessor retains ownership and may assume responsibility for remarketing or disposing of equipment at the end of the lease term. This can reduce the lessee's exposure to uncertain residual value and facilitate periodic hardware refresh, although the allocation of risk depends on return conditions, residual-value guarantees and other contractual provisions. Lease accounting, tax treatment and balance-sheet effects also depend on the applicable accounting framework and should not be inferred from the commercial form of the arrangement.

For Kazakhstan, GPU financing should be based on realistic utilisation and pricing assumptions rather than installed capacity alone. Foreign-currency exposure is particularly important because equipment is generally imported and priced in foreign currency, while part of the customer base may generate revenue in tenge. Financing structures may therefore require foreign-currency contracts, indexation, hedging or other mechanisms to manage this mismatch.

[104] CHG-MERIDIAN — AI infrastructure financing

05

Green and sustainable finance

Operational data centres and digital-infrastructure portfolios may obtain longer-term funding through bonds, sukuk, private placements or refinancing facilities once construction, commissioning and initial utilisation risks have declined.

Sustainable-finance structures may be used in both construction and operating phases and may broaden the pool of lenders and investors where the financed assets satisfy defined environmental criteria. In July 2026, AirTrunk secured a USD 2.325 billion syndicated green loan to support the development of its JHB2 hyperscale campus in Johor Bahru, Malaysia.

The transaction involved 30 financial institutions. AirTrunk states that the campus targets a design PUE of 1.37 and will incorporate water-efficient cooling technology.[105]

Green or sustainability-labelled financing requires defined eligibility criteria, appropriate use-of-proceeds controls and ongoing reporting. For Kazakhstan-based projects, relevant indicators may include energy and water efficiency, electricity-generation mix, emissions intensity, renewable-energy arrangements and effects on the electricity grid. Efficient cooling or a low design PUE would not, by itself, justify characterising electricity-intensive data-centre capacity as green.

[105] Crédit Agricole CIB — Green project financing for AirTrunk's JHB2 data centre

Islamic-finance instruments may also be relevant. Guidance published in 2024 by ICMA, the Islamic Development Bank and the London Stock Exchange Group explains how green, social and sustainability sukuk can be aligned with the ICMA Principles.[106] Sukuk could in principle finance eligible digital or supporting energy infrastructure, but the guidance addresses sustainability labelling rather than project bankability. Applicability would still depend on the underlying asset structure, cash-flow quality, legal and tax treatment, and investor demand.

[106] ICMA, IsDB and LSEG — Guidance on Green, Social and Sustainability Sukuk

Relevant Case: KT-Telecom

AIX adds a capital-market dimension to the financing landscape. As of August 2026, KT-Telecom, a Kazakhtelecom subsidiary developing the Data Center Valley project, had three active bond issues on AIX. Two USD-denominated issues carry 6% coupons: USD 41.537 million maturing in June 2036 and USD 71.040 million maturing in June 2036. In July 2026, KT-Telecom added a CNY 594.103 million issue carrying a 4.5% coupon and maturing in July 2031. All three instruments are listed for professional clients.[107] The issues demonstrate KT-Telecom's access to long-term capital-market funding in multiple currencies and provide an issuer-level precedent for financing digital infrastructure through AIX.

Taken together, these financing channels show that AI infrastructure is funded by matching each asset layer with the appropriate capital provider, tenor, security package and allocation of risk. For Kazakhstan, the common requirements are deliverable power, phased construction, identifiable demand, credible operators, enforceable contracts and transparent allocation of utilisation, technology and residual-value risks.

The immediate objective is therefore not to maximise the number of financing instruments, but to develop projects that can satisfy these conditions and progress from sponsor-funded development to construction finance, institutional investment and subsequent capital-market refinancing.

[107] Astana International Exchange — Official List

Public-private partnerships and government-backed demand

5.3

Public-private partnerships provide a framework through which the public and private sectors can jointly develop infrastructure or deliver public services over an extended period.

The private partner may finance, design, construct, operate and maintain an asset or service, while the public partner defines the required outcomes and may provide land, co-financing, long-term payments or other forms of support. The defining feature of a PPP is therefore not a particular financing instrument, but the contractual allocation of responsibilities, risks and returns over the project's life cycle.

A PPP may combine private equity, commercial or development-bank lending, institutional investment, public co-financing and, where appropriate, capital-market instruments. The private partner may recover its investment through user charges, availability payments, contracted service purchases, compensation of eligible investment or operating costs, or a combination of these mechanisms. This makes PPP potentially relevant where infrastructure serves an identifiable public requirement but cannot initially be supported by commercial demand alone.

Kazakhstan's PPP framework

5.3.1

Kazakhstan has used PPP and concession arrangements since 2005. The current framework is governed principally by the Law on Public–Private Partnership of 31 October 2015 (No. 379-V), as amended.[108] The separate Law on Concessions of 7 July 2006 was repealed in 2025, with concession agreements incorporated into the PPP framework.

Article 7 of PPP law distinguishes between institutional and contractual PPP. Contractual PPP may take the form of a concession agreement or a service contract and may incorporate elements of trust management, lease, construction and service provision. Projects may be initiated by the public or private sector and may draw on co-financing, land or existing assets, compensation of eligible investment or operating costs, availability payments and other forms of support. The appropriate structure depends on the public-service requirement, revenue model, allocation of risk and long-term fiscal implications.

Policy responsibility rests with the Ministry of National Economy, while the Kazakhstan PPP Centre performs analytical and methodological functions and maintains the national project register.

[108] Law of the Republic of Kazakhstan on Public–Private Partnerships

Scale and dynamics of the market

5.3.2

As of 1 July 2026, Kazakhstan's active and completed PPP portfolio comprised 1,150 agreements with a combined value of KZT 2.81 trillion. This included 649 active projects worth KZT 2.44 trillion and 501 agreements with expired contractual terms worth KZT 368.2 billion. A further 152 agreements with an original value of KZT 862.5 billion had been terminated and are excluded from these portfolio totals. [109]

[109] Kazakhstan Public–Private Partnership Center — PPP project database

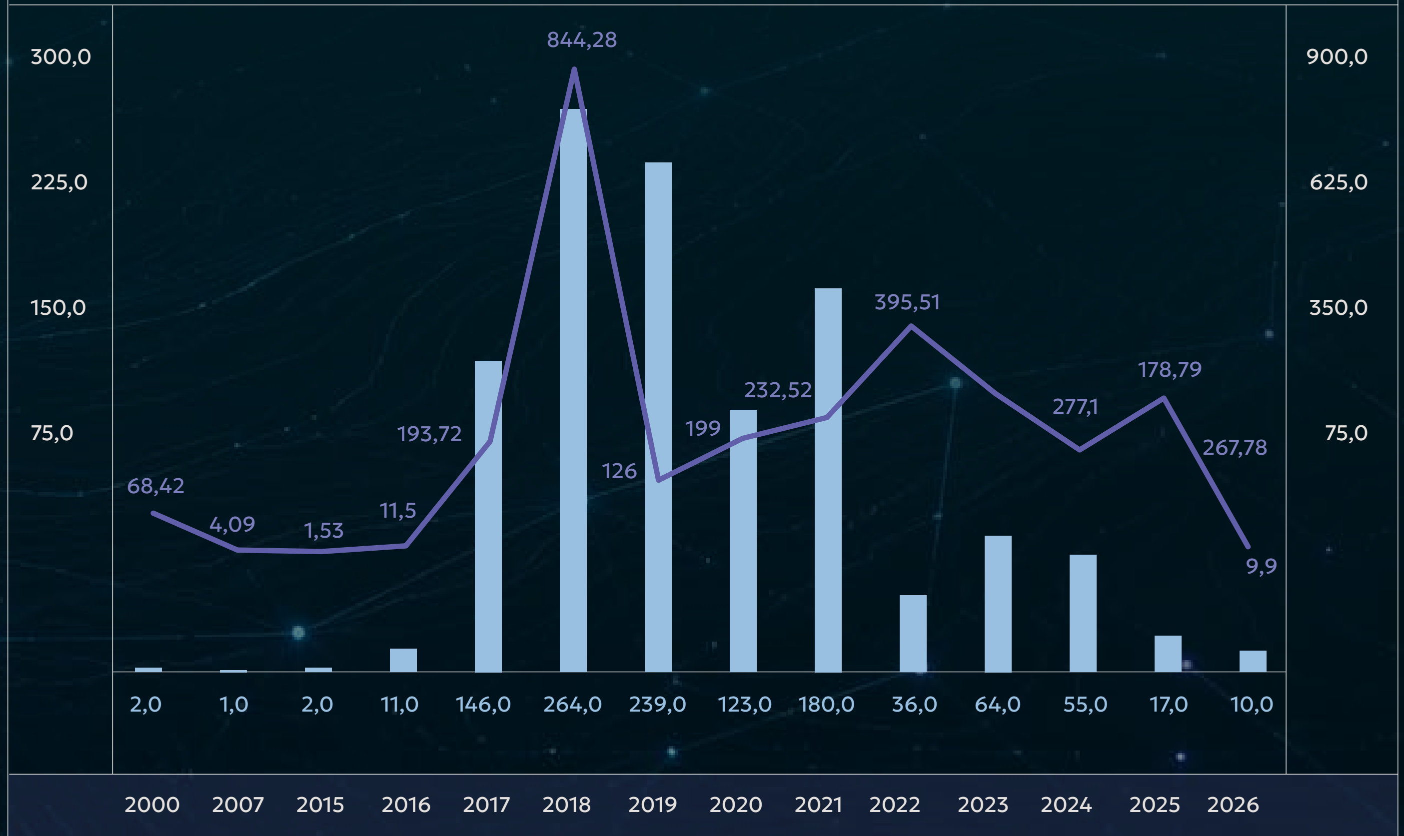
Figure 13 Source: Kazakhstan PPP Centre, as of 1 July 2026; AIFC analysis.

PPP activity by year, 2000 - 2026

Number of projects
Contract value, KZT billion

Note: * 2026 covers the first half of the year. The figure covers 1,150 active and completed agreements and excludes 152 terminated agreements.

The market has evolved from a limited number of infrastructure concessions before 2017 into a broader programme dominated by local social-infrastructure projects. Annual volumes increased rapidly during 2017–2021 and subsequently declined following tighter project-screening and fiscal requirements. At the same time, the average value of newly signed agreements increased, consistent with a shift towards fewer and larger projects.



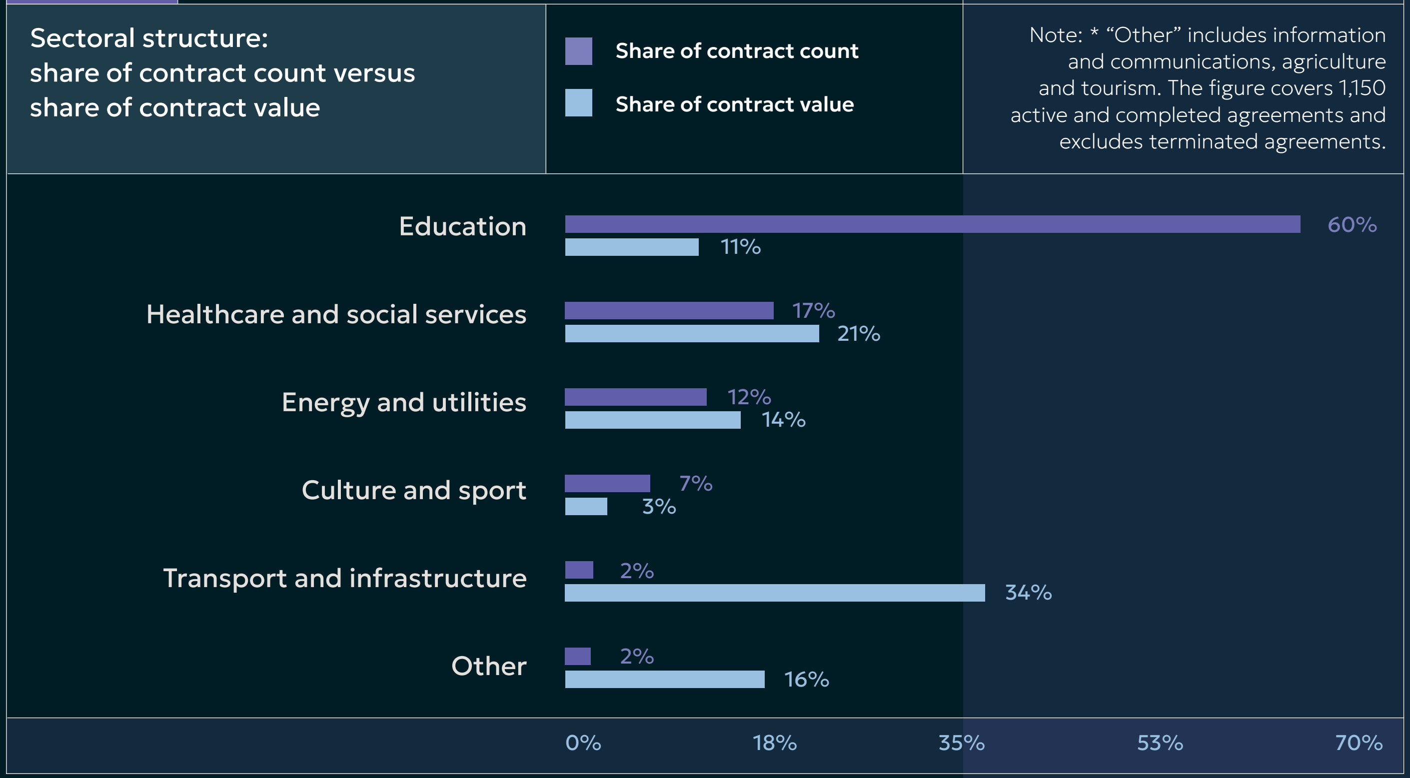
Education, healthcare and social services, and energy and utilities account for almost 89% of agreements by number. The portfolio is substantially more concentrated by value: transport and infrastructure represent only 2.3% of agreements but 34.3% of total value. This distinction matters for AI infrastructure because the more relevant precedents are the smaller group of capital-intensive infrastructure and digital projects rather than the large number of small social-sector agreements.

Figure 14

Sectoral structure: share of contract count versus share of contract value

Source: Kazakhstan PPP Centre, current-status note as of 1 July 2026; AIFC analysis.

Note: * "Other" includes information and communications, agriculture and tourism. The figure covers 1,150 active and completed agreements and excludes terminated agreements.



Relevance for AI and digital infrastructure

5.3.3

Kazakhstan has applied PPP structures to selected digital-infrastructure and public-service projects. The national register includes 17 signed ICT projects with a combined value of approximately KZT 514 billion, representing around 18,3% of cumulative portfolio value despite accounting for approximately 1.5% of agreements.[110]

[110] Kazakhstan Public-Private Partnership Center — PPP project database

The largest digital precedent is the approximately KZT 282 billion national rural broadband programme, structured through long-term service arrangements with telecommunications operators. Other examples include digital public-service platforms and Astana's approximately KZT 93 billion AI-enabled city-management platform.

PPP is likely to be most relevant where the public sector is an identifiable long-term user or where enabling infrastructure has a clear public-service function. Potential applications include sovereign cloud, shared government compute, research infrastructure, secure data environments and selected fibre, power or site infrastructure. Long-term capacity commitments, availability payments or contracted service purchases could provide an anchor revenue stream against which private capital can be raised.

The structure should reflect the different economic lives of the underlying assets. Land, grid connections, fibre and data-centre facilities may support long-duration financing, while GPUs and servers depreciate more quickly and require periodic replacement. Facility and computing assets may therefore require separate financing arrangements or explicit pricing, refresh and residual-value provisions.

For Kazakhstan, PPP should be used selectively for demand-led projects involving sovereign, public-sector or regulated workloads. Merchant data centres serving private or export customers are more likely to rely on conventional corporate, project or private-credit financing. In either case, PPP does not substitute for bankability: deliverable power, credible demand, measurable service levels, capable operators and transparent long-term fiscal commitments remain essential.

Revenue models across the asset stack

5.4

The commercial logic of AI infrastructure varies across both the asset stack and the workload served. Training capacity is most financeable when supported by anchor tenants, dedicated-cluster agreements or long-term capacity reservations. Inference is more closely linked to recurring usage, service levels and customer retention, with revenue generated through GPU-hour pricing, API calls, subscriptions, reserved capacity or managed-service contracts.

For Kazakhstan, the more credible near-term models are sovereign and regulated-sector cloud, GPU-as-a-Service, inference services and anchor-tenant facilities supported by identifiable demand. Larger training-oriented capacity may become viable where firm power, suitable sites, connectivity and credible operators are combined with contracted customers. The immediate financing case should therefore be based on secured utilisation rather than expectations of future AI demand.

Revenue is generated differently at each layer of the infrastructure stack. A data-centre facility, GPU cluster, sovereign cloud environment and managed AI service do not have the same customer base, pricing model or financing profile.

At the physical-infrastructure layer, revenue is typically generated through leases, capacity reservations or availability-based payments. In wholesale colocation, customers pay for committed capacity, often linked to power use, while the operator provides the facility, electricity, cooling, security, redundancy and connectivity. In build-to-suit structures, a facility is developed for an anchor tenant that commits to long-term capacity before or during construction. This model is more financeable because revenue is contracted before the asset is fully operational.

At the compute layer, revenue depends more directly on utilisation. GPU-as-a-Service, reserved clusters, inference capacity and storage can be priced by GPU-hour, subscription, reserved capacity or usage-based fees. This model can convert expensive equipment into recurring income, but it also carries higher operating risk. GPUs depreciate faster than buildings, require maintenance and refresh cycles, and only generate attractive returns when utilisation is high enough to cover power, cooling, support, financing and replacement costs.

At the cloud layer, revenue is linked to secure hosting, storage, cybersecurity, data-management environments and compliance-based services. This is particularly relevant for sovereign cloud and regulated-sector cloud, where customers value data residency, reliability, security and governance. For Kazakhstan, public-sector platforms, financial institutions, telecom operators and large enterprises could provide early demand if workloads are converted into multi-year service arrangements rather than one-off procurement.

At the services layer, AI infrastructure can generate higher-value revenue through model hosting, fine-tuning, inference APIs, data pipelines, governance tools and sector-specific applications. These services are closer to customer workflows and may capture more value than raw compute capacity. However, they also require stronger software capability, cybersecurity, technical support, service-level management and sector expertise.

Table 8 — Source: AIFC Industry Analysis Department.

AI infrastructure revenue model

Note: AI infrastructure generates returns through a layered revenue model, from physical capacity and reserved infrastructure to compute usage, managed AI services and customer-facing applications.

Layer	Product	Revenue model
Physical infrastructure	Land, power, cooling, fibre, data-centre shell	Lease, availability payment, power-linked contract
Capacity products	Colocation, build-to-suit, reserved MW, sovereign cloud	Long-term lease, capacity reservation, service contract
Compute products	GPU-hour, reserved cluster, inference capacity, storage	Usage-based fee, reserved capacity, subscription plus usage
AI services	Model hosting, fine-tuning, APIs, managed AI, sector solutions	Managed-service fee, SLA-backed contract, usage-based pricing
Customer applications	Enterprise workflows, public services, sector applications	Subscription, usage-based, hybrid, outcome-based pricing

The strongest revenue models combine predictable base income with usage-based upside. Long-term capacity commitments, minimum payments and reserved clusters improve bankability because they create visible cash flows. Usage-based pricing then allows the operator to benefit from higher demand. Purely speculative usage models are harder to finance because lenders and investors cannot rely on utilisation that has not yet been contracted or demonstrated.

Commercial sustainability: utilisation, pricing and cost discipline

5.5

The revenue models described above become sustainable only if they cover the full cost of operating and renewing AI infrastructure. Capacity creates the possibility of revenue, but utilisation determines whether that capacity becomes economically productive. For investors, the key question is therefore not only how much infrastructure is built, but how much of it is contracted, used and priced at a level that supports operating costs, debt service and reinvestment.

These simplified formulas are not valuation models. They show the first variables investors will test when assessing whether capacity can generate sufficient gross revenue.

For data-centre assets, the commercial model is usually driven by committed capacity:	**Annual colocation revenue = committed IT load × 1,000 × monthly price per kW × 12**
Where pricing is rack-based, the equivalent simplified formula is:	**Annual rack revenue = contracted racks × monthly rack rental rate × 12**
For GPU clusters, the economics are more utilisation sensitive	**Annual GPU revenue = number of GPUs × GPU-hour price × 8,760 hours × utilisation rate**

These formulas show gross revenue potential, not profitability. A facility or GPU cluster must still cover electricity, cooling, maintenance, staff, cybersecurity, software, financing costs, depreciation and future equipment replacement. This is why the same installed capacity can produce very different financial outcomes depending on utilisation, pricing, power cost and contract quality.

Figure 15

Source: IDC Kazakhstan, Cloud and Data Centre Market Report 2025.

Average monthly rental rate for a rack in Kazakhstan's commercial data-centre market, KZT thousand per month

Local colocation pricing provides one useful benchmark, although it should not be treated as AI-specific pricing. IDC estimates indicate that the average monthly rental rate for a rack in Kazakhstan's commercial data-centre market increased from around KZT 712 thousand in 2022 to KZT 744 thousand in 2025, with forecasts reaching approximately KZT 898 thousand by 2029. This suggests a gradually improving pricing environment for commercial data-centre capacity. However, AI workloads may require higher power density, stronger cooling, specialised networking, greater redundancy and faster hardware refresh. Their economics therefore need to be assessed separately from conventional rack-based hosting.

The most important commercial drivers are power price, time-to-power, construction cost, utilisation, contract tenor, tenant quality and cost pass-through. Power affects operating margins directly. Time-to-power determines how quickly a project can begin generating revenue. Tenant quality and contract duration influence whether lenders can rely on future cash flows. Cost pass-through determines whether changes in electricity, cooling or operating costs can be transferred to customers rather than absorbed by the operator.

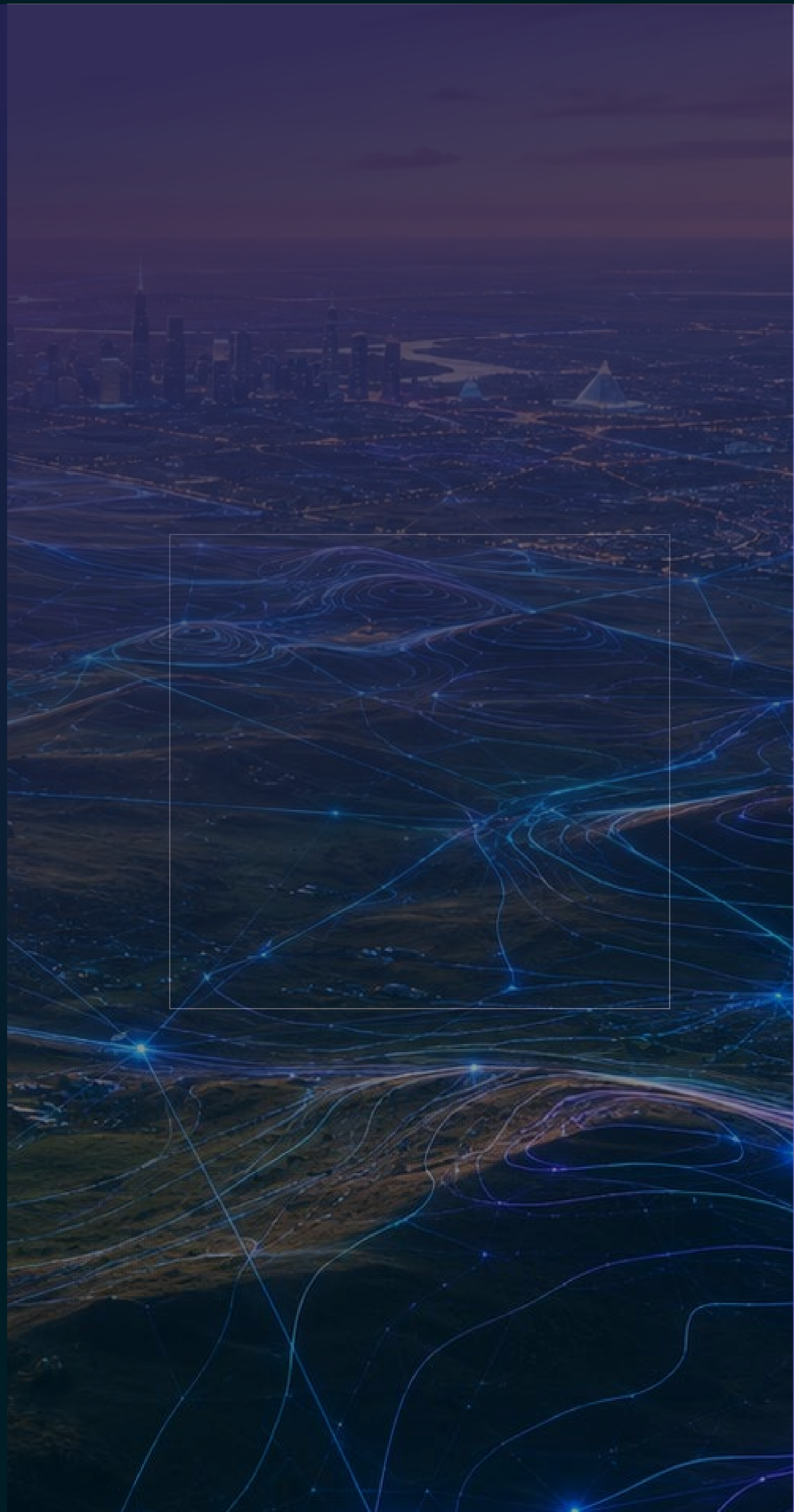


Table 9

Source: AIFC Industry Analysis Department

Key commercial drivers for AI infrastructure

Driver	Why it matters
Power price	Determines a major part of operating cost and pricing competitiveness
Time-to-power	Faster connection improves time-to-revenue and project economics
Capex per MW	Higher construction cost increases required lease rates or utilisation
Utilisation	Converts capacity into revenue
Pricing model	Determines whether revenue scales with usage and covers variable compute, power and service-delivery costs
Tenant quality	Strong counterparties reduce revenue risk, improve lender confidence and may lower financing cost
Contract tenor	Longer contracts support debt financing and refinancing
Cost pass-through	Determines whether power, cooling and operating-cost volatility can be transferred to customers
GPU refresh cycle	Compute equipment may require faster reinvestment than buildings
Cooling efficiency	Lower energy intensity improves margins and sustainability profile
Data governance	Important for public-sector and regulated-sector workloads
Legal enforceability	Required for lenders, investors, vendors and foreign technology partners to rely on contracts, collateral and step-in rights

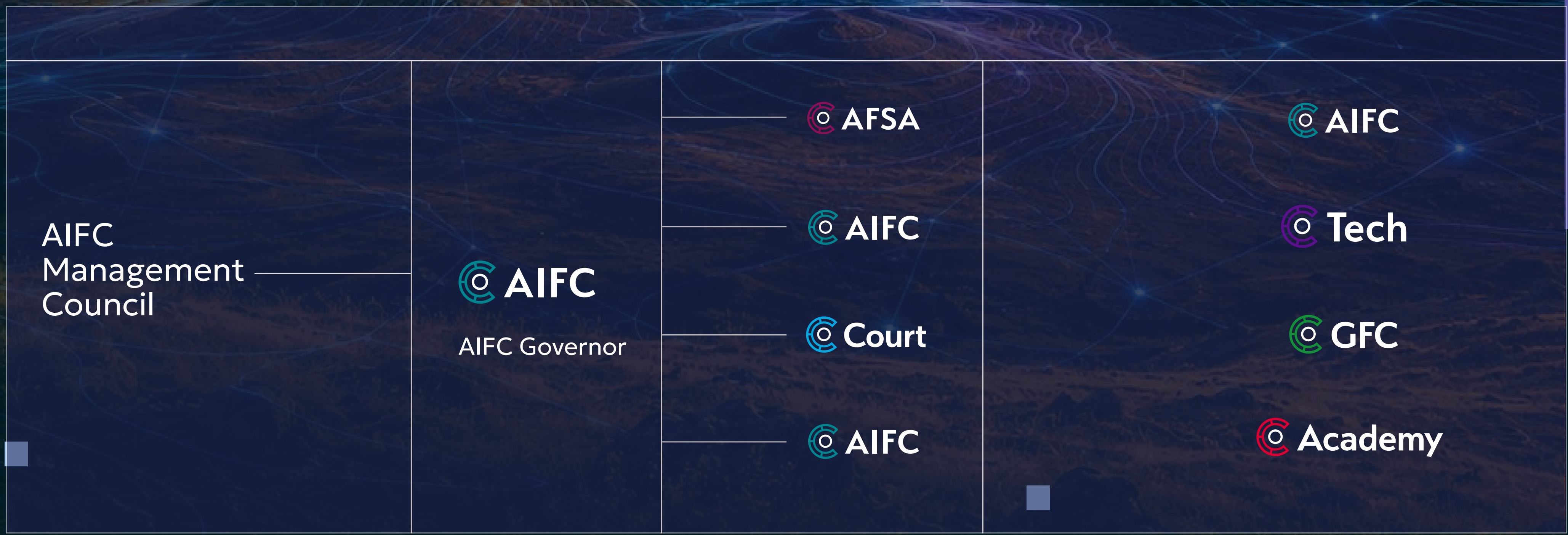
AIFC's role: legal and institutional infrastructure for investment

5.6

The AIFC provides a legal and institutional environment that may be used by sponsors, investors and lenders when structuring AI-infrastructure projects in Kazakhstan. Its relevance lies in the availability of corporate vehicles, a common-law-based legal framework, independent dispute-resolution institutions, capital-market infrastructure and investor-support services.

Since commencing operations in 2018, the AIFC has supported the mobilisation of investment into Kazakhstan. By the end of 2025, cumulative investment raised through the Centre had reached approximately USD 20 billion. Its institutional architecture brings together company registration and financial regulation through the Astana Financial Services Authority, capital-market access through the Astana International Exchange, dispute resolution through the AIFC Court and International Arbitration Centre, and investor-support services through the AIFC Authority and Expat Centre.[111]

[111] AIFC — $6 billion raised through the AIFC in 2025



Corporate establishment and project structuring

The AIFC provides several routes through which an international sponsor may establish a legal presence. Depending on the proposed structure, an investor may incorporate a new AIFC legal entity, register a branch of an existing foreign company or redomicile an existing entity from another jurisdiction. Registration is administered by AFSA and can be completed online through the Digital Resident portal.[112]

Only a subset of AIFC legal forms is likely to be relevant to AI-infrastructure investment. A Private Company[113] may serve as a project holding or joint-venture vehicle, while a Recognised Company allows an existing foreign company to establish an AIFC branch. Partnership structures[114] may be considered for multi-investor arrangements, and an eligible Special Purpose Company[115] may be used for defined conventional or Islamic structured-finance transactions.

These structures can separate project ownership and financing from the operation of the physical asset. For example, an AIFC holding vehicle could hold an interest in a Kazakhstan-based company responsible for developing and operating the data centre. Whether a separate Kazakhstan-law operating entity or branch is required depends on the project's activities, asset location and applicable national regulation.

Registration requires disclosure of the entity's activities, ownership, governance, registered address and ultimate beneficial owners, together with the applicable identification and KYC procedures. AIFC entities must maintain a registered address within the AIFC, while applicable substance and principal-business requirements depend on the selected legal form and permissions granted by the Registrar.

For non-financial activities, an applicant may generally proceed directly to registration. Financial, market and ancillary services require the applicable AFSA authorisation in addition to registration. Registration as an AIFC participant should therefore not be interpreted as permission to undertake regulated financial activities.

[112] Astana Financial Services Authority — Registration

[113] Astana Financial Services Authority — Private Company

[114] Astana Financial Services Authority — Limited Partnership

[115] Astana Financial Services Authority — Special Purpose Company

Contractual governance and investor protection

The AIFC legal framework is based on principles and standards of English common law and includes dedicated rules covering companies, contracts, employment, data protection and insolvency. This can provide international investors and lenders with a more familiar basis for structuring shareholder arrangements, joint ventures, financing documents, security packages and long-term commercial contracts.

Dispute resolution and enforcement

The AIFC Court provides an independent common-law forum for civil and commercial disputes. In addition to disputes falling directly within its jurisdiction, parties that are not registered in the AIFC may agree in writing to submit disputes to the Court.[116] The International Arbitration Centre may similarly be selected through an arbitration agreement. This opt-in mechanism allows project companies, lenders, technology suppliers, power providers and anchor tenants to select a common dispute-resolution forum at the contracting stage. AIFC Court judgments are enforceable in Kazakhstan on the same terms as judgments of Kazakhstan's courts, while IAC awards are subject to the applicable statutory recognition and enforcement framework. Electronic filing and remote hearings can also facilitate participation by parties located outside Kazakhstan.

[116] AIFC Court — AIFC Court Regulations 2017

Financing and capital-market infrastructure

The AIFC ecosystem includes legal vehicles and market infrastructure that project sponsors may use when raising capital. AIFC-based funds, AFSA-authorised credit providers and other financial institutions make their investment and lending decisions independently and on commercial terms. AIX provides infrastructure through which eligible issuers may undertake bond, sukuk and private-placement transactions or refinance operational assets.

Investor navigation and administrative support

For sponsors and investors unfamiliar with the AIFC ecosystem, the AIFC Client Office provides free initial guidance and directs applicants to the relevant AIFC body, service or next step. Where detailed legal, tax or business advice is required, it may refer clients to authorised ancillary service providers; the Client Office itself does not provide legal advice.

The development and initial operation of AI infrastructure may require foreign engineers, data-centre operators, equipment specialists and senior technical personnel. Eligible AIFC participants benefit from a simplified foreign-labour regime, including exemptions from employment quotas and work-permit requirements for qualifying foreign employees engaged in AIFC activities. Work visas may be issued for periods of up to five years.[117]

The AIFC Expat Centre provides a single service point for visa and migration procedures, temporary residence applications, individual identification numbers, electronic digital signatures and related soft-landing services. These services can reduce administrative friction during project establishment and the mobilisation of international personnel.

[117] AIFC — Expat Centre

Regulatory perimeter

The AIFC should be understood as a structuring and financing platform, not as a substitute for the regulation of physical infrastructure in Kazakhstan. An AIFC entity may provide the corporate or financing layer for a project, but a data centre, power connection or telecommunications facility remains subject to applicable national requirements[118] concerning land, construction, grid access, environmental approvals, taxation, cybersecurity, telecommunications and other sector-specific matters.

Similarly, tax incentives within the AIFC are activity-specific and subject to eligibility requirements. Registration as an AIFC participant does not automatically provide a general corporate-income-tax or VAT exemption for data centre or cloud operations.

The AIFC does not alter the underlying economics of an AI-infrastructure project. Power availability, contracted demand, construction readiness and operator capability remain the primary investment tests. Its role is narrower: to provide legal, corporate, market and dispute-resolution infrastructure that sponsors and investors may elect to use. For viable projects, this can reduce selected transaction frictions and provide international counterparties with a more familiar framework for investment in Kazakhstan.

[118] AIFC — Tax Benefits


AIFC

AI as Infrastructure Investment, compute, power and capital formation across Kazakhstan and the Turkic space

Astana International Financial Centre

2026

The Astana International Financial Centre (AIFC) is the #1 Financial Centre in Eastern Europe and Central Asian Region, and a dynamic hub for innovation and growth

World-class regulation, special tax regime, independent judicial system

We help businesses thrive in an environment that helps them attract capital in an enabling environment with regulation based on the best international standards, robust financial framework and an independent judicial system

55/18 Mangilik El, C3.3, Astana, Kazakhstan

Website: **www.aifc.kz**
For registration and licensing: **client@aifc.kz**
For analytics and publications: **industry.analysis@aifc.kz**