RESEARCH PAPER

Innovation Ecosystems and Geo Technological Competition: Prospects for U.S.–Kazakhstan Cooperation in Artificial Intelligence

Academician Bakhtay Sekerbayev

September 2026

Innovation Ecosystems • AI Infrastructure • Central Asia – U.S. Technology Policy

Abstract

The intensification of U.S.–China technological competition is restructuring the global architecture of innovation, compelling regional economies to define their position within an increasingly bipolar technological order. This paper examines the strategic logic of deepened United States–Kazakhstan cooperation in artificial intelligence, situating it within the broader contest for influence over digital infrastructure, computing standards, and critical technology supply chains in Central Asia. It argues that Kazakhstan's transition from an aspirant to an active builder of AI infrastructure — marked by the establishment of a dedicated Ministry of Artificial Intelligence and Digital Development, the enactment of the region's first comprehensive AI legislation, the launch of the Data Center Valley project in Ekibastuz, and a parallel commitment to nuclear generation capacity sufficient to power that infrastructure — creates a concrete, time-bound opportunity for U.S. engagement rather than a speculative one. The paper further examines the technical trajectory of AI

infrastructure itself, including the emerging generation of open interconnect standards (UALink, Ultra Ethernet, CXL 4.0) that will shape how compute clusters are built through 2030, and situates ongoing memorycentric computing research — including a joint initiative with the TGI-Amiron group — within that trajectory. It concludes with a set of concrete cooperation pathways and a candid assessment of the risks that could foreclose the window now open.

$10B	**100,000 NVIDIA GPUs**	**2.4 GW**	**1,000,000 people**

Data Center Valley Kazakhstan nuclear capacity AI workforce training target investment (Ekibastuz) Planned GPU cluster scale target by 2035

Sources: Firebird/NVIDIA–Kazakhstan agreement (June 2026); Kazakhstan Atomic Energy Agency; Ministry of Artificial Intelligence and Digital Development, National AI Platform targets.

1. Innovation Ecosystems in an Era of Geo Technological Competition

Under contemporary conditions, the formation of an innovation-driven economy has become one of the strategic priorities of socio-economic development for states and regions alike. Against a backdrop of deepening geopolitical tension, intensifying trade and technological competition, and accelerating digital transformation, the capacity of national economies to build resilient innovation ecosystems has become a determinant of technological sovereignty, competitiveness, and access to advanced scientific and technological solutions.

The United States retains a leading position across most key technological sectors of the global economy. The American model of innovation development — largely grounded in the principles of open innovation — rests on three interconnected pillars: the development of institutional and infrastructural environments; the stimulation of innovation activity and the reduction of barriers to entrepreneurship; and government support for priority, strategically significant technologies. The consistent transformation of scientific ideas into commercially viable solutions has been a defining prerequisite for both the U.S. National Innovation System (NIS) and its constituent Regional Innovation Systems (RIS), and the U.S. experience remains instructive for regulatory design, multi-level strategic planning, and mechanisms for eliciting active private-sector participation in innovation governance.

Artificial intelligence now occupies a central position within this system of international technological competition. While no definitive winner has yet emerged in the global AI race, the technological system is becoming increasingly polarized around two centers of power — the United States and China — creating additional pressure on regional economies that previously maintained strategic neutrality. Under this pressure, such economies are increasingly compelled either to declare a technological orientation or to actively manage a balance between the two poles. Kazakhstan's own posture, examined in Section 4, is a case study in the latter strategy.

2. The Bipolar Technological Order: Five Fronts of Competition

Contemporary geo technological confrontation is developing across five interconnected fronts: energy; rareearth minerals and microelectronics; digital and physical infrastructure; the knowledge economy; and the regulatory and legal governance of technology. It is along these fronts that U.S.–China competition is most clearly manifested, and it is here that new geographic and institutional contours of the global technological hierarchy are being drawn.

One of the principal U.S. competitive advantages in computing infrastructure is its leadership in highperformance microelectronics and specialized computing accelerators. American technology corporations — NVIDIA and AMD foremost among them — are expanding into Southeast and Central Asian markets, where demand for localized AI capability is rising steadily. In parallel, the populations and businesses of these regions are becoming further integrated into American digital platforms across social media, cloud services, and enterprise software, deepening the region's dependence on the American technology stack — a dependence with commercial upside for U.S. firms and strategic significance for Washington.

China, for its part, is pursuing a parallel strategy of promoting its own technological standards and digitaleconomy regulatory models, with the explicit aim of reducing Central Asian reliance on Western digital solutions. Beijing is actively supporting the digitalization of tourism, industry, and agriculture through major Chinese IT companies, and the propagation of compatible Chinese digital standards, platforms, and regulatory approaches could meaningfully expand its influence over regional digital processes while narrowing the space available to Western alternatives. Chinese investment in the IT sector, however, frequently carries conditions — particularly around cybersecurity, data processing, and cross-border data transfer — that can constrain the autonomy of local companies in selecting their own digital infrastructure.

Export controls over strategically important technologies, materials, and components have intensified in parallel on both sides. China's policy toward rare-earth elements illustrates Beijing's willingness to use control over critical supply chains as an instrument of technological and economic policy: licensing requirements attached to Chinese-origin materials or equipment used in the processing of rare metals or the production of magnets can affect transactions in which no Chinese company is directly involved. This dynamic increases uncertainty across global production chains and reinforces the strategic case for diversifying the sourcing of both critical resources and critical technologies — a case in which Kazakhstan, as a major producer of uranium and other critical minerals, has direct standing.

3. Contested Middle Ground: ASEAN and the Digital Alignment Contest

Cooperation between the United States and ASEAN illustrates the broader dynamic now unfolding in Central Asia. Washington has an interest in consolidating political, economic, and technological ties with the region, while ASEAN states generally regard the U.S. presence as a useful counterweight to China's growing influence — a configuration that gives the United States room to expand its technological and economic footprint in Southeast Asia. The situation surrounding TikTok and the disposition of its U.S. operations is instructive here: it demonstrates that digital platforms have become instruments of geopolitical and technological influence in their own right, not merely commercial products, and that this dynamic is intensifying competition between American and Chinese models of the digital ecosystem more broadly.

The same logic — platforms and standards as instruments of influence, not merely commerce — is now visible in Central Asia, where the stakes are arguably higher because the underlying digital and energy infrastructure is still being built rather than retrofitted.

4. Kazakhstan's AI Ecosystem: From Strategy to Infrastructure

Kazakhstan has moved from AI policy aspiration to institutional and physical build-out on a compressed timeline. In September 2025, the government created a dedicated Ministry of Artificial Intelligence and Digital Development — elevating AI policy from a sub-department to a central pillar of national governance, led by Minister Jaslan Mädiev. This ministry now serves as the coordinating node for the Concept for the Development of Artificial Intelligence for 2024–2029, a framework organized around three priorities: infrastructure, human capital, and regulation. Central Asia's first comprehensive AI legislation entered into force in January 2026, and President Kassym-Jomart Tokayev has designated 2026 the “Year of Digitalization and Artificial Intelligence,” anchoring a government-wide implementation plan spanning eight priority blocks.

The infrastructure component of this strategy has already attracted large-scale foreign investment. In June 2026, Kazakhstan signed a $10 billion package of agreements with Firebird and NVIDIA to build the Data Center Valley in Ekibastuz, encompassing a planned cluster of 100,000 NVIDIA GPUs — among the larger single AI infrastructure commitments announced globally to date, and a direct, present-tense analogue to the kind of American technology-stack expansion the original draft described only in the ASEAN context. The government has reinforced this with a VAT and customs-duty exemption on imported AI computing equipment, intended to accelerate private-sector build-out of high-performance computing capacity.

Human-capital and research infrastructure are being developed in parallel. The Alem.AI International Center for Artificial Intelligence, housed within Astana Hub, anchors a national plan to train one million people in AI skills, and is guided by a Presidential AI Development Council that includes international participation — notably UAE AI Minister Omar Sultan Al Olama — alongside existing partnerships with UAE firms Presight AI (developing Kazakhstan's first domestically hosted supercomputer) and AIQ (energy-sector AI applications). The Qazaqstan Venture Group has committed up to $1 billion in venture capital targeting AI, big data, and cloud computing projects, and the government has separately identified 53 investment projects worth roughly $4.2 billion under its Smart City and digital-economy programs.

For a paper assessing the strategic case for U.S. engagement, the operative point is this: Kazakhstan is not requesting American partnership to build an AI ecosystem from a blank sheet — it is requesting partnership to shape the trajectory of one that is already under construction, with capital, legislation, and physical infrastructure already committed. That materially changes both the urgency and the terms of the opportunity for American firms and policymakers.

5. The Energy Foundation: Nuclear Power and the Physics of AI Growth

Following an October 2024 national referendum in which more than 71 percent of voters approved the construction of nuclear generation capacity, Kazakhstan has moved rapidly to develop a multi-plant nuclear program. Russia's Rosatom was selected in June 2025 to lead construction of the country's first plant — the Balkhash NPP, near Ulken in the Almaty region, using VVER-1200 reactor technology, with an intergovernmental agreement signed in May 2026 and active construction targeted for 2027. China's CNNC was subsequently selected to lead a second plant, sited in the Zhambyl district adjacent to the first, with siting approved in early 2026. In April 2026, the President approved a national nuclear industry strategy extending to 2050, envisioning at least three operating plants by mid-century and a target of at least 2.4 gigawatts of nuclear capacity by 2035.

Kazakhstani officials have linked this buildout explicitly to rising electricity demand from mining, metals processing, hydrogen production, industrial manufacturing, and — notably — AI infrastructure. The dualvendor structure of the program, splitting the first two plants between Russia and China, is itself a data point for this paper's central argument: even as Kazakhstan pursues deep U.S. technology-sector partnership in AI, it is simultaneously managing its core energy dependencies through its two largest neighboring powers. This is the balancing strategy described in Section 1 made concrete in physical infrastructure, and it should inform how the United States frames cooperation — not as an exclusive alignment, but as one leg of a deliberately diversified technological and energy portfolio Kazakhstan intends to keep diversified.

6. Technical Architecture of Next-Generation AI Infrastructure

AI infrastructure is being reshaped by the exponential growth in model size, training and inference data volumes, and the resulting need to scale compute clusters far beyond the limits of earlier architectures. By 2030, AI infrastructure will likely be characterized by high compute density, widespread liquid cooling, optical interconnects, and memory-centric architectures designed to maximize data-processing efficiency — and the standards governing how accelerators communicate within and across racks are being actively contested and rewritten now, not in the abstract future.

6.1 The Interconnect Standards Race

NVIDIA's proprietary NVLink fabric currently dominates production AI training deployments and remains, as of 2026, the only production-ready option for the largest clusters, supporting up to 576 GPUs in a single coherent domain. A coalition of vendors — including AMD, Intel, Google, Microsoft, Meta, Broadcom, Cisco, HPE, and AWS — has organized around the open UALink standard as a direct alternative: UALink 1.0, published in April 2025, specifies scaling to 1,024 accelerators per fabric at 200 GT/s per lane, with UALink 2.0 published in April 2026; production hardware from AMD, Intel, and Astera Labs is expected in late 2026, with meaningful deployment volume likely extending into 2027. In parallel, the Ultra Ethernet Consortium finalized its 1.0.1 specification in mid-2025, re-architecting Ethernet from the physical layer up to solve the multipathing, congestion-control, and security limitations inherited from earlier RDMA protocols such as RoCE v2.

On the memory side, the CXL Consortium released CXL 4.0 in November 2025, doubling bandwidth to 128 GT/s and enabling multi-rack memory pooling exceeding 100 terabytes — a direct response to memory bandwidth having become, alongside raw compute, the primary bottleneck constraining GPU cluster performance for models exceeding 100 billion parameters. Consolidation is already underway in the switching silicon that will implement these standards: Marvell's $540 million acquisition of XConn Technologies in January 2026 signaled the emergence of CXL and UALink switching as a distinct, strategically important layer of the AI hardware stack. Despite differences among vendor software ecosystems — CUDA, ROCm, and oneAPI foremost among them

— convergence at the interconnect and memory layer is the clearer trend, and it is this convergence, not any single vendor's roadmap, that will determine how portable and how open the next generation of AI infrastructure actually is.

Standard	Governing Body	Current Status (2026)	Function
UALink 1.0 / 2.0	UALink Consortium	2.0 published Apr 2026; production hardware late 2026	Open GPU-to-GPU scale-up interconnect, up to 1,024 accelerators
Ultra Ethernet 1.0.1	Ultra Ethernet Consortium	Finalized mid-2025; shipping	Re-architected Ethernet fabric for AI cluster networking
CXL 4.0	CXL Consortium	Released Nov 2025; multirack pooling by 2027	Memory expansion and disaggregation, 128 GT/s
NVLink	NVIDIA (proprietary)	Production-dominant	Current incumbent scaleup fabric, up to 576 GPUs

Table 1. State of open AI interconnect standards as of August 2026.

6.2 Memory-Centric and Data-Intensive Computing

Particular research and commercial interest are now concentrated on architectures optimized for dataintensive workloads — accelerated database filtering and aggregation, recommendation systems, graph processing, and the vector operations underlying modern retrieval-augmented and inference systems. Within this direction, a project underway jointly with the TGI Amiron group is pursuing near-memory processing: moving simple operations directly into memory to reduce data movement between memory and the central processing unit, relieve CPU load, and improve overall computing efficiency. This work sits squarely within the broader industry shift — evidenced by CXL 4.0's memory-pooling capability and by dedicated switching-silicon investment such as Marvell's XConn acquisition — toward treating memory bandwidth, not raw compute, as the binding constraint on next-generation AI systems.

Optical interconnects are becoming correspondingly important for GPU-to-GPU connectivity in both AI and high-performance computing systems, offering higher bandwidth and lower latency than traditional electrical interconnects and enabling more efficient communication across large-scale clusters — a capability that will matter directly for any facility built at the scale contemplated in Ekibastuz.

7. Pathways for U.S.–Kazakhstan Cooperation

Kazakhstan's demonstrated ability to build a full-cycle innovation ecosystem — integrating digital infrastructure, educational programs, professional training, and mechanisms for commercializing technological solutions — gives the U.S.–Kazakhstan relationship a foundation beyond bilateral technology transfer. It positions Kazakhstan as a potential point of attraction for the wider international AI community and a platform connecting American technological and investment capital with the broader computing needs of Central Asian markets. The following pathways follow directly from the evidence in Sections 4–6:

- Compute infrastructure co-investment: engage U.S. hyperscale’s and chipmakers as strategic partners — alongside Firebird and NVIDIA — in the buildout and long-term operation of the Ekibastuz Data Center Valley and successor sites, with attention to standards portability (Section 6.1) so that Kazakhstan's infrastructure is no single-vendor locked.
- Workforce and research partnership: connect the Alem. AI center and the national one-million-person AI training target with U.S. university and industry training programs, building the talent base that any large compute cluster requires to operate at utilization.
- Energy-technology cooperation as a parallel track: given the dual-vendor (Rosatom/CNNC) structure of Kazakhstan's nuclear buildout, U.S. engagement should not assume energy exclusivity; grid modernization, efficiency, and cooling-technology cooperation (Section 6.2) offer a complementary channel that does not require displacing existing nuclear partners.
- Standards alignment: encourage Kazakhstan's infrastructure planners to specify UALink/Ultra Ethernet/CXL-compatible architecture in new procurement, both to avoid vendor lock-in and to keep the resulting infrastructure interoperable with U.S. cloud and enterprise systems.
- Regulatory dialogue: given that Kazakhstan's AI law entered into force in January 2026 — the first in the region — U.S. engagement on AI governance now can shape a framework still being implemented, rather than one already locked in around Chinese or purely domestic models.

8. Risk Factors and Strategic Considerations

A forward-looking assessment should state plainly what could foreclose this opportunity rather than assume it stays open by default.

- Balancing, not alignment: Kazakhstan's simultaneous nuclear partnerships with Russia and China, and its active courtship of Chinese digital standards alongside American ones, indicate that Kazakhstan will continue to hedge rather than align exclusively with any single technological bloc. U.S. strategy should be calibrated to a multi-aligned partner, not an exclusive one.
- Execution risk on a compressed timeline: the Ministry of AI and Digital Development, the AI law, and the Data Center Valley agreement were all established within roughly an eighteen-month window (2025–2026). Institutional capacity to execute at this pace, particularly around grid interconnection and skilled-labor supply, is unproven and merits close monitoring rather than assumption of smooth delivery.
- Standards fragmentation: the interconnect race described in Section 6.1 is unresolved. Infrastructure committed today on a single-vendor (NVLink) basis carries different long-term interoperability risk than infrastructure built to open standards not yet in volume production.
- Export-control exposure: as Section 2 notes, Chinese and U.S. export-control regimes over semiconductors and rare-earth-derived components can each reach transactions in which the controlling country is not a direct party. Kazakhstan-based projects using components sourced across both blocs should be assessed for this exposure specifically.

9. Conclusion: Toward a Regional AI Platform

The formation of a modern AI ecosystem requires the simultaneous development of computing infrastructure, human capital, regulatory frameworks, international technological cooperation, and mechanisms for commercializing innovation. Kazakhstan's progress against each of these dimensions since 2025 — a dedicated ministry, enacted legislation, a $10 billion compute infrastructure commitment, a nuclear power program explicitly tied to industrial and AI demand, and a national workforce-training target — has moved the country from a plausible candidate for regional AI leadership to an active builder of it. For the United States, the strategic choice is no longer whether Central Asia merits AI-sector engagement, but how quickly and on what terms to engage with a build-out that is already underway, being financed, and being shaped — with or without American participation — by other partners in parallel.

Selected Sources

Construction Review Online, “Kazakhstan Advances Plans for Second Nuclear Power Plant as Russia's Rosatom Eyes Role,” June 2026.

World Nuclear Association / World Nuclear News, Kazakhstan country profile and nuclear strategy updates, March–July 2026.

Rosatom State Atomic Energy Corporation, press releases on the Balkhash NPP, May 2026.

Qazinform News Agency, “How Kazakhstan Plans to Become an AI Powerhouse” and “Kazakhstan to Establish International Technology Hub for AI and Startups,” May–July 2026.

The Diplomat, “6 Months into Kazakhstan's Year of AI,” June 2026.

Ministry of Artificial Intelligence and Digital Development of the Republic of Kazakhstan, agency overview.

Global FDI Reports, “Building a Digital Nation,” April 2026.

The Astana Times, “From Smart Cities to AI Education: Kazakhstan's 2026 Digitalization Projects” and “Kazakhstan's Smart City Strategy Enters New Phase Through Global AI Partnerships,” 2026.

Introl, “UALink and CXL 4.0” and “Marvell's $540M XConn Acquisition Signals AI Interconnect Consolidation,” January–February 2026.

Tom's Hardware, “Ultra Ethernet: The Data-Center Interconnection of Tomorrow Detailed,” February 2026.

Wikipedia / UALink Consortium, UALink specification history, retrieved May 2026. ChipDocket,

“CXL vs UALink vs NVLink: The Scale-Up Fabric,” April 2026.