POSTDOCTORAL TECHNICAL WHITE PAPER

BEYOND MODEL SCALE

Why Frontier AI Providers Should Evaluate the Predictive Tensor Control Plane (PTCP)

A non-proprietary, capability-level analysis for the Kimi K3 era

PERFORMANCE	ASSURANCE	SOVEREIGNTY
More accepted output per accelerator-dollar	Independent, bounded runtime control	Portable policy across heterogeneous infrastructure

Tensor Networks, Inc.

440 N Wolfe Rd, Sunnyvale, California 94085 | www.ptcp.ai

Evidence cutoff: July 24, 2026 | Revised non-proprietary edition: July 26, 2026

NON-PROPRIETARY DISCLOSURE

This edition intentionally omits algorithms, equations, source code, scoring logic, model parameters, feature engineering, optimization methods, compilation details, and other trade-secret implementation information.

Executive Abstract

Frontier-model competition is becoming a systems contest. Model quality remains necessary, but capable challengers increasingly combine sparse scale, long context, agentic products, multimodal operation, open or deployable weights, and aggressive API economics. This changes the buyer's comparison from model benchmark position alone to the delivered outcome: useful work per dollar, tail latency, tool reliability, evidence-grade control, recoverability, and deployment sovereignty.

Moonshot AI's Kimi K3 illustrates the shift. At this paper's July 24, 2026 evidence cutoff, Moonshot reported a hosted K3 release with a 2.8-trillion-parameter sparse mixture-of-experts architecture, one-million-token context, native multimodality, long-horizon tool use, and published API prices of $3 per million uncached input tokens and $15 per million output tokens. Moonshot stated that full weights, the final license, and a technical report were scheduled for July 27. These were vendor disclosures rather than a fully independently reproduced model card at the cutoff. [1]-[3]

RECOMMENDATION Frontier AI providers should evaluate Tensor Networks' PTCP as a model-agnostic control overlay around the serving, infrastructure, agent, and assurance fabric. Evaluation should begin in observation-only mode and progress to bounded production control only after predefined proof, non-regression, rollback, and governance gates are met.

PTCP's strategic proposition is not that it changes model weights or makes a model intrinsically more intelligent. It seeks to improve the operating envelope around the model: cross-layer situational awareness, anticipatory resource coordination, low-latency enforcement of already authorized actions, independent agent containment, and a defensible evidence trail. These capabilities can matter when the benchmark gap narrows but the cost and risk of delivering accepted outcomes remain highly differentiated.

This revised edition is intentionally non-enabling. It describes capability categories, trust boundaries, interfaces, expected outcomes, evaluation methods, and adoption controls. It does not reveal proprietary algorithms, tensor structures, topology measures, scoring functions, feature sets, parameterization, training or update procedures, optimization objectives, source code, policy-compilation logic, or device-specific recipes.

Decision question	White-paper answer	Required evidence before scale
Why now?	K3-class challengers compress model-only differentiation while increasing infrastructure and agentic stress.	Current workload traces, buyer-loss analysis, and representative serving profiles.
Why PTCP?	PTCP targets service efficiency, runtime assurance, recovery, and deployment reach without retraining the model.	Controlled comparison against the incumbent scheduler, traffic engineering, identity, and security stack.
How should it be adopted?	Instrument, observe, advise, canary reversible actions, and then decide.	Signed rollback, zero critical safety events, protected non-regression, and uncertainty-bounded benefit.

The conclusion is conditional. PTCP is strategically relevant because it addresses the control deficit exposed by next-generation AI services. Its value must be earned through evidence; it should not be inferred from terminology, patents, partnerships, or architectural elegance.

Document Status, Claim Discipline, and Disclosure Boundary

TEMPORAL STATUS Kimi K3 is treated as a newly launched hosted challenger at the July 24 evidence cutoff, with a promised open-weight release and technical report still pending at that cutoff. [1]-[3]

This paper is a technical and strategic argument for controlled evaluation. It does not claim that PTCP has already improved a frontier-model provider's production environment, and it does not imply endorsement by Moonshot AI, a hardware vendor, a standards body, or a government agency. Tensor-authored sources are used to identify the proposed architecture and validation agenda; they are not independent proof of production benefit. [14]-[16]

Evidence class	Treated as established	Still to be proven
Primary third-party	K3 launch disclosures; government agentic-AI guidance; public Linux, NVIDIA, and SONiC implementation surfaces.	Incremental PTCP benefit over mature schedulers, congestion controls, and security tooling.
Patent and public record	Issued public patent lineage and a public notice concerning licensing activity.	Counsel-confirmed claim scope and license rights for each shipped product capability.
Tensor-authored	A coherent capability model, safety posture, and falsifiable validation program.	Production-scale performance, containment accuracy, interoperability, certification, and economic outcomes.

Protected information excluded from this edition

- Mathematical formulations, proprietary scoring functions, optimization objectives, and decision thresholds internal to PTCP.
- Feature-selection, feature-engineering, tensor construction, model architecture, training, calibration, and update procedures.
- Topology-computation methods, anomaly logic, geodesic calculations, graph transformations, and internal causal inference methods.
- Source code, pseudocode, data structures, schemas detailed enough to reproduce internals, policy compiler logic, and target-specific implementation recipes.
- Confidential benchmark corpora, customer data, unpublished performance results, unreleased product roadmaps, key-management internals, and operational secrets.

The paper retains enough information for a buyer, researcher, investor, or infrastructure leader to understand the problem, evaluate the claimed outcomes, design a safe proof of value, and conduct procurement diligence without enabling replication of Tensor Networks' proprietary technology.

Contents

1. The Strategic Thesis: Competition Moves Below the Model

As frontier capability diffuses, the durable unit of competition becomes the complete model-delivery system: weights, inference runtime, fabric, storage, agent harness, policy, evidence, and recovery. Providers still need excellent models, but the commercial outcome is increasingly determined by how reliably and economically those models are operated.

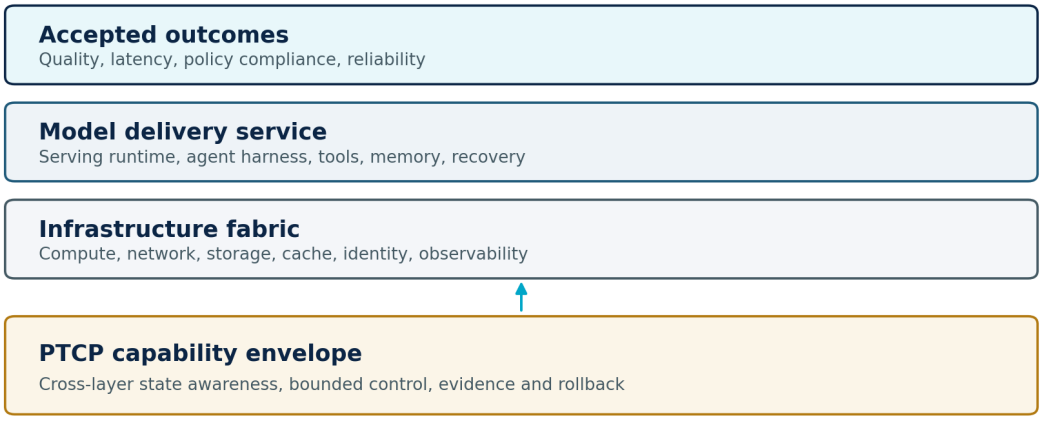


Figure 1. Competitive advantage increasingly depends on the system around the model. PTCP is presented as a capability envelope, not as a model-weight modification.

1.1 Why model-only moats decay

- **Sparse scale changes the cost frontier.** Large sparse systems shift pressure toward placement, communication, cache locality, and coordination in addition to arithmetic throughput.
- **Long context makes state movement a product constraint.** Extended sessions magnify prefill bursts, cache footprints, transfer decisions, storage pressure, and tail sensitivity.
- **Long-horizon agents extend risk beyond inference.** The consequential event may be a tool call, credential use, delegated subagent, data movement, or external side effect rather than the token that preceded it.
- **Price transparency changes procurement.** As list prices fall, buyers pay greater attention to accepted completion cost, reliability, control, and deployment constraints.
- **Sovereign demand rewards portability.** Public-sector, regulated, critical-infrastructure, and customer-controlled deployments require local authority, transparent controls, and bounded disconnected operation.

1.2 The control-plane opportunity

Modern inference architectures already validate the systems premise. NVIDIA Dynamo, for example, emphasizes dynamic scheduling, disaggregated serving functions, request routing aware of cached state, and accelerated state transfer. NVIDIA also argues that high-frequency fabric telemetry supports proactive AI-factory operations. These references do not validate PTCP, but they show that orchestration below the model materially affects utilization, latency, and service reliability. [7], [8]

STRATEGIC TEST If two models produce comparable accepted outcomes, the winning provider is likely to be the one that delivers those outcomes with lower tail risk, lower infrastructure waste, stronger evidence, faster recovery, and fewer deployment constraints.

2. Kimi K3 Is the Trigger, Not the Target

K3 matters less as a single rival than as evidence that advanced capability, low price, long context, multimodality, and agentic ambition can arrive together. The appropriate response is not to imitate a competitor's private architecture. It is to make the provider's complete service more economical, governable, recoverable, and deployable.

2.1 What was known at the evidence cutoff

Dimension	Moonshot disclosure at July 24, 2026	Analytical status
Availability	Hosted model, Kimi products, coding surface, and API reported launched July 16.	Vendor availability established; independent SLA analysis outside scope.
Scale	2.8T total parameters; 16 of 896 experts reportedly selected per token.	Do not infer active-parameter count or operating efficiency without the technical report.
Context and modes	One-million-token context; native text, image, and video; tool and structured-output support.	Product claim; quality and reliability remain task dependent.
Deployment	Vendor recommended a 64-or-more-accelerator supernode for local deployment.	Signals fabric intensity; exact sizing and economics require released artifacts.
Economics	$0.30 cache-hit input, $3 uncached input, and $15 output per million tokens.	Published list price at the evidence cutoff.
Open release	Full weights, final license, and technical report promised for July 27.	Pending at the cutoff; earlier licenses do not establish K3's final terms.

Sources: Moonshot AI's Kimi K3 technical announcement, quickstart documentation, and pricing materials. [1]-[3]

2.2 Systems stresses implied by K3-class services

K3-class attribute	Operational stress	PTCP capability hypothesis	Proof metric
Sparse, distributed serving	Communication hot spots, placement sensitivity, and collective stalls.	Cross-layer state awareness can identify avoidable contention and recommend bounded corrective action.	Accepted outcomes per accelerator-hour; tail stall time; queue and retransmit behavior.
Million-token context	Prefill bursts, large cached state, transfer and recomputation penalties.	Phase and locality awareness can reduce avoidable state movement and burst amplification.	Tail time to first token; cache reuse; recomputation; bytes moved.
Long-horizon tool use	Permission drift, unplanned egress, lateral movement, and cascading actions.	Identity-linked behavior control can detect or constrain unauthorized communication expansion.	Scenario recall; false containment; time to interrupt; audit completeness.
Aggressive API pricing	Margin pressure and buyer focus on accepted-output economics.	Operational optimization can improve useful work per loaded infrastructure dollar.	Accepted completion cost; retry rate; SLO penalties; energy per accepted task.
Deployable weights	Sovereign, private, hybrid, and heterogeneous deployment expectations.	A model-agnostic policy and evidence layer can travel across infrastructure domains.	Deployment lead time; policy parity; portability defects; disconnected continuity.

2.3 What K3 does not prove

K3's launch does not prove superiority over every proprietary frontier model, lower cost for every workload, or stronger safety for enterprise agency. Vendor-identified limitations concerning agent behavior and harness discipline reinforce the systems thesis: capable models need a capable execution envelope. [1]

K3 is not a threat model for PTCP. It is a forcing function. The relevant question is whether a provider can operate any K3-class model, including its own future systems, with demonstrably better efficiency, control, and deployment flexibility than competitors. PTCP must be tested against that question, not against model benchmarks it cannot influence.

3. The Frontier Model Is a Distributed Cyber-Physical System

A production frontier-model service is a feedback system spanning accelerators, networks, caches, storage, schedulers, identities, tools, data services, and human authority. Failures propagate across these layers faster than conventional dashboards can explain them.

3.1 From inference endpoint to coupled service

A request may be admitted by a gateway, use a prompt cache, enter a prefill pool, transfer state to a decode pool, invoke retrieval, dispatch tools, spawn subagents, stream output, persist memory, and produce auditable side effects. Each transition changes physical state, including queues, memory, bandwidth, thermal conditions, and failure-domain exposure, as well as logical state, including identity, privilege, data boundary, obligation, and intent.

A controller that reacts to averages can destabilize the tail. A scheduler that optimizes accelerator occupancy can overload a network or storage path. A security gate without task context can miss lateral movement or interrupt legitimate work. The gap is not simply a missing dashboard; it is the absence of a cross-layer state, authority, and evidence discipline.

3.2 Four coupled failure loops

Loop	Typical trigger	Propagation	Business consequence
Congestion	Burst, placement imbalance, or path concentration.	Queueing, retransmission, stalled collective, longer occupancy, more queueing.	Tail latency, lost throughput, unstable scaling, SLO loss.
Cache and storage	Poor locality, stale route, or state-transfer mismatch.	Miss, recompute or transfer, bandwidth pressure, eviction, additional misses.	Higher cost per completion and variable response time.
Agent	Ambiguous objective, hostile content, or compromised tool.	Unexpected action, new identity or data edge, delegated action, cascading side effects.	Security incident, policy breach, or manual recovery.
Control	Stale state, delayed feedback, or competing actuator.	Conflicting change, oscillation, misleading telemetry, stronger correction.	Outage risk and loss of operator trust.

3.3 Government guidance converges on runtime control

The April 2026 multinational guidance led by NSA and CISA treats agentic AI as a composite service containing models, tools, data, memory, planning, and sometimes subagents. It identifies cascading failures, weak segmentation, lateral movement, resource exhaustion, and audit complexity. Recommended controls include unique cryptographic identity, per-action policy, segmentation, approval gates, rate limits, live interruption, rollback, behavioral baselines, independent monitoring, and unified logs. [4]

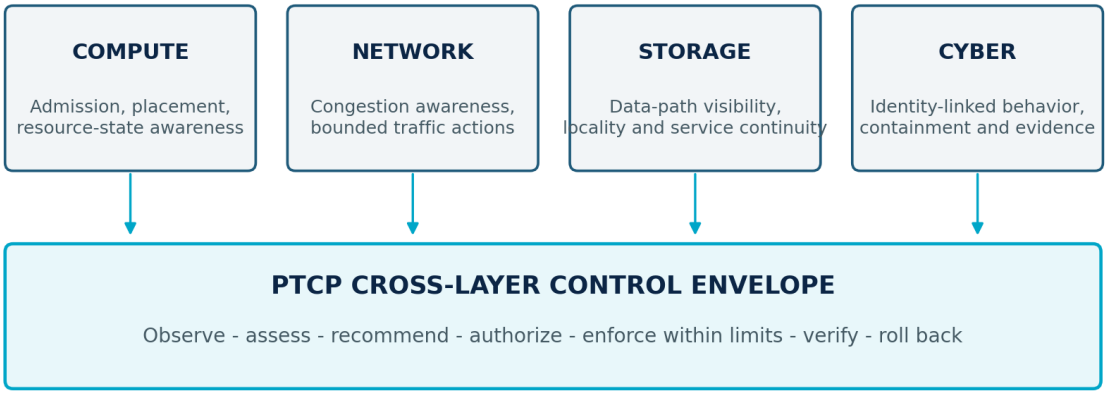
NIST's Generative AI Profile similarly emphasizes continuous monitoring, fallback and deactivation procedures, incident response, real-time safety metrics, logging, and test, evaluation, verification, and validation. It warns that laboratory benchmarks do not guarantee validity in operational contexts. PTCP aligns conceptually with these objectives, but alignment is not compliance. Each deployment

must map implemented controls to the applicable authorization boundary and sector obligations. [5], [6]

DESIGN IMPLICATION The control plane should observe model-serving conditions and agent authorization context together while preserving separation of duties among inference, recommendation, approval, enforcement, and recovery.

4. PTCP Capability Model Without Proprietary Disclosure

For evaluation purposes, PTCP is best understood as a model-agnostic predictive control overlay that observes cross-layer conditions, identifies material deviations, recommends bounded actions, enforces only authorized controls, and records an independent evidence chain. This definition is deliberately functional. It does not disclose the proprietary mathematics or implementation techniques used to produce predictions, scores, or recommendations.



Non-enabling disclosure: capability categories are shown; proprietary analytics and implementation internals are omitted.

Figure 2. Capability-level PTCP envelope across compute, network, storage, and cybersecurity. Proprietary analytics and implementation internals are intentionally omitted.

4.1 Six observable capability domains

Capability domain	Externally observable function	What a customer can test
Cross-layer visibility	Correlates infrastructure, workload, service, identity, and policy context without requiring model-weight changes.	Coverage, clock quality, identity binding, tenant separation, overhead, and data minimization.
Predictive state awareness	Provides forward-looking or early-warning operational signals with declared uncertainty.	Lead time, calibration, precision, stability, drift response, and value beyond simple baselines.
Topology-aware assurance	Identifies material changes in physical, logical, dependency, or authorization relationships.	Scenario precision and recall, false containment, explanation quality, and context sensitivity.
Bounded control	Applies preapproved, scoped, expiring, rate-limited, and reversible actions through supported interfaces.	Signature validation, ceilings, expiry, target compatibility, safe mode, and rollback.

Capability domain	Externally observable function	What a customer can test
Evidence and recovery	Links observed condition, recommendation, authorization, action, outcome, and recovery record.	Completeness, tamper evidence, causal reconstruction, export, retention, and incident usability.
Deployment portability	Carries policy intent and evidence across supported cloud, host, DPU, switch, scheduler, proxy, and isolated domains.	Capability parity, explicit degradation, disconnected operation, removal, and data portability.

4.2 Compute, network, storage, and cybersecurity benefits

- **Compute.** PTCP can contribute phase-aware resource visibility, placement context, capacity-risk warning, and policy-bounded admission or workload steering. The buyer should test accepted output per accelerator-hour rather than peak utilization alone.
- **Network.** PTCP can correlate workload demand with queue, path, and service state, recommend congestion-avoidance actions, and use low-latency enforcement surfaces where appropriate. The buyer should test tail latency, loss, retransmission, fairness, and stability.
- **Storage and data movement.** PTCP can expose locality, transfer pressure, cache or retrieval dependencies, and storage-service degradation as part of the same operating context. The buyer should test bytes moved, avoidable recomputation, retrieval reliability, and service continuity.
- **Cybersecurity.** PTCP can associate identity, task, tool, data classification, destination, and behavior changes, then support observation, shaping, redirection, isolation, or revocation within authorized limits. The buyer should test scenario recall, legitimate-traffic impact, time to interrupt, and evidence completeness.

4.3 Slow-path and fast-path separation

PTCP's public design can be described safely as two operational paths. A higher-latency analytical path performs observation, assessment, simulation, recommendation, and approval. A bounded low-latency path applies only previously authorized actions and reports outcomes. The exact analytics, scores, model structures, thresholds, and compilation logic are proprietary and outside this paper.

Property	Analytical and governance path	Bounded enforcement path
Purpose	Observe, assess, compare, simulate, recommend, authorize.	Match an authorized condition, apply a bounded action, count, report, expire, and roll back.
Authority	May propose; does not independently bypass the policy gate.	Enforces only signed, scoped, compatible policy within hard ceilings.
Failure behavior	Freezes promotion, preserves evidence, and falls back to the declared incumbent or safe state.	Uses a named safe policy or enters an explicitly defined conservative mode.
Customer evidence	Version, input lineage, recommendation, approval, replay result, and uncertainty.	Policy identity, target, action, counters, expiry, outcome, and rollback status.

4.4 Functional substitution is a credibility requirement

A defensible architecture should remain useful if a proprietary analytical component is replaced or disabled. Buyers should contract for observable input and output behavior, interface stability, safety properties, and evidence, rather than require disclosure of confidential internals. Tensor should demonstrate that each module can be benchmarked against strong baselines and that no single analytical model can write directly to an enforcement target.

NON-ENABLING BOUNDARY This paper makes no representation of the internal tensor representation, graph method, scoring equation, optimization criterion, model-update logic, thresholds, or policy compiler. Those items should be handled under appropriate confidentiality, evaluation, licensing, and intellectual-property controls.

4.5 Patent, licensing, and evidence boundary

Public records identify relevant Pattern-of-Life patent lineage and a Federal Register licensing notice. Those records can establish the existence of prior public technical work and licensing activity, but they do not by themselves establish the scope, ownership, enforceability, freedom to operate, or production performance of every PTCP capability. A procurement or investment process should require counsel-reviewed claim mapping, chain-of-title and license diligence, confidentiality controls, and a clear separation between public disclosures and proprietary know-how. [12], [13]

5. Competitive Mechanism I: Useful Output per Dollar

The economically relevant denominator is not raw tokens or peak accelerator throughput. It is the fully loaded cost of an accepted, timely, policy-compliant outcome. A system can improve tokens per second while worsening accepted-task cost if it increases retries, state transfer, cache misses, tool failures, or tail SLO violations.

Measurement layer	Primary measures	Why a narrow metric fails
Physical	Accelerator idle and stall time; memory pressure; link utilization; retransmits; queue depth; power.	High average utilization can conceal synchronized stalls and long tails.
Serving	Time to first token; inter-token latency; output rate; cache reuse; recomputation; preemption.	More generated tokens can be waste if tasks are rejected, repeated, or abandoned.
Product	Accepted completion; tool success; grounded result; human correction; abandonment.	A fast answer can still be wrong, unsafe, incomplete, or unusable.
Business	Accepted outcomes per loaded dollar; SLO loss; incident exposure; deployment time.	List price excludes integration, reliability, assurance, and opportunity cost.

5.1 Capability-level mechanisms to test

1. Workload-phase awareness that distinguishes compute, memory, network, storage, and tool-service conditions before recommending admission or placement changes.
2. Locality-aware decisions that account for the cost and risk of moving or recomputing state rather than selecting a superficially short queue.
3. Early identification of path concentration or emerging hot spots before loss and retransmission amplify into collective or service stalls.
4. Approved service degradation under constraint, such as alternate pools or bounded priority treatment, instead of uncontrolled queue growth.
5. Closed-loop evidence that records whether the predicted operational outcome followed the authorized change and whether rollback was required.

5.2 Falsifiable hypotheses

Hypothesis	Comparator	Primary endpoint	Failure condition
State-aware control preserves locality.	Current request router, scheduler, and cache policy.	Recomputation or transfer cost per accepted task.	No improvement or worse tail latency after workload stratification.
Cross-layer diagnosis reduces congestion tail.	Existing traffic engineering and congestion controls.	Tail network-stall contribution in targeted scenarios.	Benefit disappears under blinded replay or introduces instability.
Bounded actuation increases useful capacity.	Baseline admission and capacity policy.	Accepted outcomes per accelerator-hour.	Gain is explained by quality loss, retries, or SLO regression.

Hypothesis	Comparator	Primary endpoint	Failure condition
Evidence shortens recovery.	Current incident and rollback process.	Time to identify, interrupt, restore, and explain.	Operators cannot reproduce the action and outcome chain.

COMPETITIVE IMPLICATION If K3-class pricing resets buyer expectations, a provider cannot defend margin with benchmark quality alone. It needs an evidence-backed account of accepted-output economics and a control system that can improve that account without adding unacceptable risk.

6. Competitive Mechanism II: Long-Horizon Agent Assurance

Long-horizon agents need a control boundary that persists across prompts, tools, subagents, retries, model switches, and infrastructure domains. PTCP's strongest strategic opportunity may be to make that boundary independently observable and interruptible.

6.1 Why prompt-layer safety is insufficient

A model can comply with a language policy while operating inside an overprivileged harness. Conversely, a capable model can be destabilized by hostile retrieved content, a compromised tool, an incorrect memory, or loss of task context. The assurance target is therefore the end-to-end action and communication graph, not only the model's text output.

Material actions should carry a verifiable principal, task, policy version, data classification, destination, purpose, and approval state. PTCP can contribute an independent enforcement and evidence plane only when integrated with identity, zero trust, secrets management, data-loss prevention, application authorization, and incident response. [4]-[6]

6.2 Topological containment at a capability level

A topological containment policy expresses which principals may communicate with which tools, services, data stores, or external destinations for a defined task. It can also express customer-approved limits related to rate, scope, data class, geography, and time. PTCP can treat an unauthorized relationship or material graph expansion as a reviewable state change. The internal method used to detect or score that change is proprietary and is not disclosed here.

Risk class	Default action	Authority	Evidence requirement
Observe	Log and enrich; no service-path change.	Automated.	Identity, task, flow, tool, destination, and policy context.
Shape	Mark, pace, or rate-limit inside a preapproved envelope.	Automated under signed ceiling.	Reason, duration, counters, expiry, and observed impact.
Redirect	Use an approved proxy, sandbox, alternate pool, or inspection path.	Policy gate; human approval for sensitive data.	Destination equivalence, protected invariants, and rollback path.
Isolate	Block a new relationship or quarantine a principal, session, or workload.	Automated for explicit deny; otherwise approved.	Reason code, preserved state, scope, appeal, and release path.
Terminate or revoke	Stop task, revoke credential, or disable a tool binding.	Named high-risk authority.	Complete action chain, incident record, recovery and reauthorization plan.

6.3 Safety properties to require

- Default-deny expansion for protected zones, with explicit, signed, time-limited exceptions.
- Identity continuity across model, tool, service, subagent, data, and destination boundaries.

- An independent interruption path that does not depend on the model or agent runtime cooperating.
- Idempotent bounded actions, minimum dwell, update-rate ceilings, and rollback tested under telemetry loss and partial partition.
- Tamper-evident evidence linking observation, recommendation, approval, enforcement, outcome, and release.
- Human override with accountable quarantine release; no irreversible automated action in the initial proof of value.

6.4 The false-containment problem

Containment is not free. A false block can interrupt customer work, corrupt a transaction, or violate an availability obligation. Security recall must therefore be paired with legitimate-traffic false-containment rate, time to restore, and downstream impact. The safest initial deployment is observation-only. Canary control should begin with low-consequence, reversible measures and a named recovery objective.

ASSURANCE ADVANTAGE A provider can differentiate by proving that its agents operate inside an independently observable and interruptible envelope even when the underlying model is proactive, replaced, fine-tuned, or hosted in a customer-controlled environment.

7. Competitive Mechanism III: Sovereign and Heterogeneous Deployment

Deployable challengers raise buyer expectations for locality and control. Frontier providers need a policy and evidence layer that can operate across public cloud, colocation, private cloud, sovereign facilities, edge sites, and disconnected environments.

7.1 Sovereignty is more than data residency

A sovereign AI service must answer who can observe, decide, authorize, actuate, update, interrupt, and audit, not only where bytes reside. Local enforcement must continue safely if a central analytics service is unavailable. Policy artifacts should be inspectable, authorized within the customer domain, and scoped to infrastructure, tenant, workload, and time.

Sovereignty dimension	Required capability	PTCP design test
Data	Local minimization, classification, retention, and redaction.	Can the deployment operate without exporting sensitive payload content?
Operational	Local approval, rollback, incident response, and failover.	Does central disconnection preserve a declared safe service state?
Technological	Portable interfaces across supported host, DPU, switch, scheduler, identity, and model stacks.	Does each intended control compile exactly or fail explicitly?
Legal and organizational	Named controller, processor, operator, approving authority, and jurisdiction.	Are actions attributable and contractually constrained?
Evidentiary	Locally verifiable records and reproducible decisions.	Can an independent assessor reconstruct why an action occurred?

7.2 Portability without magical abstraction

Public interfaces such as Linux XDP and AF_XDP, disaggregated network operating systems such as SONiC, and DPU or SmartNIC platforms make portable enforcement plausible. They are not semantically identical. Driver behavior, offload limits, verifier constraints, queue semantics, hardware capabilities, and operational ownership vary. PTCP should publish a target capability matrix and explicitly reject or degrade unsupported behavior rather than silently approximate a protected security invariant. [9]-[11]

7.3 Deployment modes

Mode	State and authority	Enforcement posture	Best fit
Provider-managed	Centralized with tenant partitioning and provider authority.	Host, fabric, and service controls under provider operations.	Public API and managed enterprise service.
Customer-controlled	Customer domain; optional constrained summary exchange.	Customer-controlled host, DPU, switch, identity, and proxy surfaces.	Regulated private cloud or colocation.

Mode	State and authority	Enforcement posture	Best fit
Disconnected or isolated	Local-only observation, retained evidence, and bounded local authority.	Cached approved controls, local approval, explicit rollback.	Defense, critical infrastructure, air-gapped operations.
Hybrid multi-domain	Domain-local state with constrained inter-domain exchange.	Each domain enforces its own authorized control set.	Multinational and supply-chain ecosystems.

COMMERCIAL OPTION VALUE Even before an efficiency gain is proven, a portable control and evidence layer may expand addressable market by reducing the engineering burden of sovereign and customer-controlled delivery. That option value must be measured against integration and support cost.

8. Non-Proprietary Reference Deployment Pattern

A credible deployment separates observation, analysis, authorization, enforcement, and evidence into independently testable trust zones. The following pattern describes functional responsibilities and acceptance properties only. It does not reveal proprietary models, feature sets, scores, data structures, algorithms, or device-specific implementation steps.

Functional layer	External responsibility	Minimum customer-verifiable property
Observation	Collect authorized infrastructure, workload, service, identity, and control health signals.	Clock and identity quality, schema governance, backpressure, minimization, tenant separation.
Context and state	Associate observations with workload phase, service objective, topology, authority, and uncertainty.	Lineage, versioning, missing-data behavior, drift visibility, resource budget.
Recommendation	Produce candidate operational or assurance actions for review.	No direct enforcement authority; bounded recommendation set; reason and expected outcome.
Replay and validation	Evaluate proposed changes against recorded or simulated conditions.	Reproducibility, scenario coverage, protected-invariant checks, declared fidelity limits.
Authorization	Apply risk class, separation of duties, scope, signature, expiry, and promotion workflow.	Unsigned or stale policy rejected; revocation and exception ownership tested.
Target adaptation	Validate whether an authorized intent is supported by a particular target.	Exact compatibility or explicit failure; version and resource ceilings recorded.
Enforcement	Apply only authorized bounded actions and return counters and status.	Small trusted surface, watchdog, safe mode, expiry, atomic or equivalent rollback.
Evidence and recovery	Link observation, recommendation, authorization, action, outcome, and restoration.	Tamper evidence, export, retention, incident correlation, removal and recovery testing.

8.1 Control-plane safety and availability

- No analytical component writes directly to an enforcement target; all actions traverse authorization and compatibility validation.
- Observation and actuation channels are independently authenticated; compromised telemetry cannot become control authority.
- Every target has resource and update-rate ceilings; exceeding a ceiling rejects promotion and raises a visible control-health condition.
- Every authorized change is versioned, scoped, expiring, attributable, and associated with a named rollback state.
- Clock degradation, stale context, partition, or disagreement enters a declared safe mode and preserves incumbent service control.
- The incumbent scheduler, routing, congestion, zero-trust, and incident systems remain authoritative until ownership is explicitly transferred.

8.2 Threat model

PTCP is itself a privileged control system and therefore a high-value target. The assessment should cover forged or delayed telemetry, replayed or stale authorization, compromised adapters, malicious recommendation inputs, policy tampering, privilege escalation, cross-tenant leakage, host or DPU compromise, evidence deletion, and denial of analytical services. The deployment should test roots of trust, identity lifecycle, key rotation, artifact signing, build provenance, least privilege, recovery, and independent interruption. MITRE ATLAS can complement conventional enterprise and infrastructure threat frameworks. [18]

REFERENCE RULE PTCP must reduce total control risk. An opaque privileged controller that improves one performance metric while weakening auditability, stability, or recovery is a net regression.

9. Quantitative Business Case

The business case should be built from measured reclaimed capacity, avoided SLO loss, reduced incident consequence, recovery improvement, and deployment option value, net of licensing, integration, infrastructure, assurance, and operating cost.

9.1 Capacity-value sensitivity

The table below is illustrative arithmetic, not a forecast. It assumes 10,000 accelerators operating 8,760 annual hours. Productive reclaim means measured capacity that produces accepted outcomes without violating quality, safety, or SLO guardrails.

Loaded accelerator cost	1% productive reclaim	3% productive reclaim	5% productive reclaim
$2 per hour	$1.75M	$5.26M	$8.76M
$4 per hour	$3.50M	$10.51M	$17.52M
$6 per hour	$5.26M	$15.77M	$26.28M

Replace fleet size, loaded cost, utilization, demand, reclaim, and realization assumptions with finance-approved values. Reclaimed technical capacity has value only if it is realizable as revenue, avoided capital, faster customer admission, or resilience headroom.

9.2 Full cost inventory

Cost category	Examples	Diligence question
Software and support	License, telemetry, storage, observability, support obligations.	Is pricing tied to accelerators, sites, throughput, users, or measured outcomes?
Integration	Adapters, target compatibility, identity, CI/CD, replay environment.	Which interfaces are productized versus customer engineering?
Infrastructure	Collectors, state services, replay capacity, DPU resources, evidence retention.	What is steady-state and burst overhead by workload class?
Operations	On-call, review, key lifecycle, upgrades, incident and rollback drills.	Who owns the controller when it conflicts with the network, scheduler, or security stack?
Assurance	Threat modeling, red team, independent assessment, compliance evidence.	Which proof artifacts are reusable across sites and jurisdictions?
Exit and reversibility	Policy export, evidence export, removal, fallback, staff retraining.	Can the provider return to incumbent controls without service loss?

9.3 Decision model

Use downside, base, and upside cases. The downside assumes no measurable performance gain but preserves bounded integration learning and removal. The base case requires a small, reproducible operational gain plus stronger evidence or recovery. The upside combines capacity, SLO, assurance, and sovereign revenue benefits without double counting. Production expansion

should occur only when the downside is bounded by reversibility and the measured base case clears the organization's risk-adjusted hurdle.

FINANCE DISCIPLINE Do not monetize every detected anomaly or blocked red-team action. Value only realized outcomes with an accountable owner, a defensible counterfactual, and documented causality.

10. Sixteen-Week Proof-of-Value Program

A 16-week program can determine whether PTCP deserves production investment without granting premature authority over critical traffic or agents. Select one infrastructure-intensive workload and one agentic workflow with material but reversible tool use. Exclude life-safety, destructive, irreversible financial, and mission-critical production actions from automated canary control.

Figure 3. Evidence-gated PTCP adoption sequence. Timing may be extended when seasonality, safety, or workload coverage is insufficient.

Stage	Weeks	Activities	Exit evidence
0 - Baseline	1-2	Instrumentation contract; workload strata; incumbent KPIs; threat model; acceptance criteria; rollback drill.	Required-event coverage and identity/clock quality pass; baseline is reproducible.
1 - Shadow	3-5	Observation and recommendations only; replay congestion and attack scenarios; measure overhead and calibration.	No service actuation; bounded overhead; measurable incremental signal beyond named baselines.
2 - Advisory	6-8	Operator-visible recommendations; blinded accept/reject logging; replay validation; failure injection.	Stable recommendations; explanations usable; no unresolved high-severity control defect.
3 - Canary	9-12	Signed low-risk actions on tightly scoped traffic or isolated agents; automatic expiry; switchback and rollback drills.	Protected non-regression; successful recovery; statistically credible operational signal.
4 - Decision	13-16	Expanded red team; independent review; finance model; operating model; procurement and exit terms.	Go, hold, or stop decision with evidence package and named residual risks.

10.1 Workstreams and accountability

Workstream	Accountable owner	Critical deliverable
Serving performance	Inference platform lead	Stratified baseline, treatment design, accepted-output metric.
Fabric and enforcement	Network or DPU lead	Target capability matrix, resource ceilings, recovery test.

Workstream	Accountable owner	Critical deliverable
Agent assurance	Product security lead	Identity continuity, permitted communication policy, scenario corpus, interruption path.
Data and evaluation	Independent evaluation lead	Pre-registration, blinded analysis, reproducibility package.
Safety and risk	AI risk officer or delegate	Non-regression gates, exception process, residual-risk acceptance.
Economics	Finance partner	Realization-adjusted ROI and sensitivity model.
Vendor and legal	Procurement and counsel	IP and license diligence, confidentiality, support, audit, data, and exit terms.

10.2 Mandatory stop conditions

- Any unbounded, unsigned, or unaudited actuation path, or any ability to bypass the authorization gate.
- Critical cross-tenant exposure, privilege escalation, or inability to revoke PTCP authority.
- Rollback cannot restore the named prior state within the tested recovery objective.
- Observation or enforcement overhead violates the pre-registered non-inferiority margin.
- Control oscillation, route instability, or conflict with incumbent controls cannot be explained and bounded.
- Primary benefits exist only in aggregate and disappear under workload-stratified or switchback analysis.

11. Experimental Design and Decision Metrics

Competitive claims should be evaluated as causal hypotheses with protected non-regression endpoints, workload stratification, uncertainty intervals, and reproducible traces. The objective is not to prove that every PTCP component is universally superior; it is to determine where the integrated capability adds incremental value without unacceptable risk.

11.1 Proposed primary endpoints

Domain	Primary endpoint	Proposed canary-to-scale gate	Guardrail
Economics	Accepted outcomes per accelerator-hour and loaded dollar.	At least 3% improvement with a 95% interval excluding zero in one prespecified stratum.	No quality, retry, abandonment, energy, or SLO regression beyond margin.
Latency	Tail time to first token or task completion.	Non-inferior within a pre-registered margin; improvement secondary unless prespecified.	Median and p95 latency, jitter, timeout, and credit impact.
Fabric	Tail network or collective stall contribution.	Credible reduction in the targeted congestion scenario.	No oscillation, fairness loss, or shifted retransmission.
Agent security	Detection or containment of curated unauthorized communication expansion.	At least 95% scenario recall in the defined corpus.	No more than one false containment per 10,000 legitimate material actions.
Recovery	Time to identify, interrupt, restore, and reconstruct.	Low-risk rollback completed within the approved objective.	Zero lost evidence and verified service-state restoration.
Overhead	Fast-path latency and CPU cost; analytical resource use.	Within target-specific pre-registered budget.	No tail amplification when a collector or analytical service fails.

Thresholds are proposed decision criteria, not reported PTCP results. Adjust them only before data are unblinded and document the rationale.

11.2 Analysis plan

6. Pre-register endpoints, workload strata, exclusions, non-inferiority margins, assignment, minimum sample, and stop conditions.
7. Estimate effects within strata and report central tendency and tails, not only means. Account for time dependence and shared infrastructure.
8. Use paired switchbacks, stepped canaries, or other designs appropriate to interference from network and cluster-level changes.
9. Control multiplicity for secondary measures. A dashboard with many moving metrics is not a decision rule.
10. Publish effect sizes, uncertainty, missingness, configuration, workload mix, negative results, and raw-trace references under the applicable data policy.

11.3 Security evaluation

The scenario matrix should cover hostile retrieved content, compromised tools, credential misuse, unauthorized destinations, data-class boundaries, subagent fan-out, tool-loop resource exhaustion, path or route anomalies, denial of telemetry, stale authorization, forged identity, and control-plane compromise. Separate detection, decision, enforcement, recovery, and evidence. A blocked scenario without an attributable policy and recovery chain is incomplete evidence.

11.4 Comparative baselines

PTCP must compete against the actual incumbent stack rather than a deliberately weak baseline. Performance comparisons should include native request routing, scheduler and cache policy, traffic engineering, congestion control, and observability. Security comparisons should include identity gateways, service mesh or proxy policy, endpoint controls, data-loss prevention, and existing anomaly detection. The evaluation should identify incremental value and interaction cost.

EVIDENCE STANDARD A credible result is not that a metric improved after PTCP was installed. It is a reproducible, uncertainty-bounded treatment effect with workload context, protected non-regression, and a documented action-to-outcome chain.

12. Risks, Limitations, and Explicit Non-Claims

PTCP introduces a privileged control capability. Its value depends as much on disciplined limits, interface maturity, governance, and recovery as on predictive performance.

Risk	Why it matters	Required mitigation or evidence
Incomplete or skewed telemetry	A predictive system can become confidently wrong when clocks, identity, sampling, or context drift.	Data-quality SLO, uncertainty visibility, missingness tests, safe-mode trigger.
Control interaction	PTCP may conflict with schedulers, routing, congestion, autoscaling, or security gates.	Explicit ownership map, staged authority, dwell limits, joint failure injection.
Actuation instability	Rapid changes can oscillate routes, queues, admission, or placement.	Update ceilings, stability tests, canary, named rollback, control-health SLO.
False topology signal	Legitimate deployment or workflow changes may resemble abnormal relationship expansion.	Context, corroboration, scenario precision and recall, human review.
Overhead and state growth	High-frequency visibility and cross-layer context consume resources.	Fixed budgets, sampling sensitivity, degradation mode, independent overhead measurement.
Privileged-plane compromise	Forged authorization or a compromised adapter can affect many services quickly.	Rooted identity, signing, separation of duties, reproducible builds, revocation, red team.
Vendor and interface maturity	A public architecture may still require bespoke integration and scarce expertise.	Capability matrix, acceptance tests, support SLO, productized adapters, references.
IP or licensing ambiguity	Public lineage may be broader or narrower than rights needed for shipped capabilities.	Counsel-reviewed claim and license mapping, confidentiality, indemnity, freedom-to-operate diligence.
Privacy and governance	Cross-layer traces can reveal tenant, user, code, data, or workload behavior.	Minimization, local processing, field-level controls, retention, deletion, access audit.
Economic non-realization	Technical gains may not convert to revenue, avoided cost, or risk reduction.	Finance realization factor, scenario model, and stop after a failed base case.

12.1 Explicit non-claims

- PTCP does not improve model weights, reasoning quality, factuality, or alignment by itself.
- No public evidence reviewed for this edition establishes that PTCP has outperformed a frontier provider's production control stack.
- PTCP does not replace routing protocols, congestion control, network operating systems, inference schedulers, service meshes, zero trust, data-loss prevention, SIEM, or human authorization.
- The presence of a proprietary analytical method does not guarantee forecast accuracy, causal diagnosis, stability, or security.
- A patent, public licensing notice, partnership, or available implementation interface is not a benchmark, certification, government approval, or vendor endorsement.
- The illustrative economic sensitivity is not a price quote, forecast, valuation, or representation of Tensor customer results.

- Kimi K3 specifications and limitations are vendor disclosures current to the stated evidence cutoff; later released artifacts may change the assessment.

12.2 Conditions under which PTCP should not be adopted

Do not adopt PTCP when workload scale cannot justify the integration burden; when incumbent controls already meet performance and assurance requirements at lower complexity; when required target semantics cannot be enforced exactly; when telemetry cannot be collected lawfully and reliably; when control ownership is unresolved; or when Tensor cannot provide auditable interfaces, safe rollback, support, and evidence commensurate with the requested authority.

GOVERNANCE RULE Architecture creates permission to test, not permission to trust. Authority should expand only as evidence, governance, and reversibility expand.

13. Strategic Recommendations

13.1 For frontier-model executives

11. Define competition at the service level. Use a board-visible scorecard for accepted outcomes per dollar, tail SLOs, agent-control incidents, recovery, and sovereign deployment lead time.
12. Sponsor a bounded PTCP proof of value. Pair one infrastructure-efficiency workload with one agent-assurance workflow and preserve incumbent authority until evidence is independently reviewed.
13. Create a control-ownership council spanning inference, network, storage, product security, SRE, risk, procurement, and data governance.
14. Pre-negotiate exit. Require policy and evidence export, removal testing, interoperability commitments, support obligations, and no service loss when PTCP is disabled.

13.2 For platform and infrastructure leaders

- Instrument workload phase, locality, topology, identity, service objectives, and accepted-output state with explicit data-quality ownership.
- Publish the incumbent baseline and protect its safety mechanisms; PTCP earns incremental authority rather than inheriting it.
- Require signed intent, target compatibility validation, expiry, and explicit rejection of unsupported semantics.
- Design interruption, rollback, partition behavior, and service restoration before optimizing throughput.
- Budget overhead by component and workload; treat control-plane health as a production SLO.

13.3 For Tensor Networks

Tensor should publish a non-proprietary interface and assurance package appropriate for a privileged infrastructure product: product scope, trust boundaries, target capability matrix, resource ceilings, failure-mode summary, signature and key-lifecycle requirements, reproducible evaluation harness, baseline-selection rules, independent red-team results, deployment references, support model, and a precise counsel-reviewed account of patent and license rights.

Tensor should protect trade secrets through disciplined disclosure segmentation. Public materials should explain customer-visible capabilities and falsifiable outcomes. Confidential diligence may expose additional test hooks, interface details, and assurance artifacts under agreement. Source code, algorithms, parameters, proprietary scores, feature engineering, and device-specific recipes should remain protected unless a deliberate licensing or escrow decision requires otherwise.

13.4 Go, hold, or stop

Decision	Evidence condition	Action
Go	Primary endpoint passes; all guardrails pass; rollback and evidence complete; base-case economics clear.	Expand one scope boundary at a time; retain switchback and recurring red-team gates.
Hold	Promising signal but uncertainty crosses zero, seasonality is incomplete, or bounded operational defects remain.	Extend observation or advisory work; close evidence gaps; do not widen authority.
Stop	Safety stop, unexplained instability, no incremental benefit, nonportable semantics, or uneconomic integration.	Remove actuation, export evidence, document the result, retain only independently useful instrumentation.

Conclusion

Kimi K3 is a warning against treating frontier status as a permanent property of model weights. Capability can diffuse quickly; operating discipline compounds. Providers must compete on the complete service, not only the model.

PTCP addresses the right strategic layer. Its non-proprietary value proposition combines cross-domain visibility, predictive state awareness, topology-aware assurance, bounded low-latency control, independent evidence, recovery, and deployment portability. These capabilities offer a plausible path to more useful output per dollar, stronger long-horizon agent control, and broader sovereign deployment without retraining the underlying model.

Plausibility is not proof. Cross-layer analytics may fail to beat simple baselines. Signals may correlate without diagnosing cause. New control authority can destabilize mature systems. Integration and governance may cost more than reclaimed capacity. Patent lineage and partnerships cannot settle those questions.

The appropriate course is disciplined adoption: observe first, compare against strong incumbents, preserve protected non-regression, expand only through signed and reversible actions, and require independent review of benefit and risk. Passing that test could create a durable systems moat around models whose benchmark lead may be temporary.

Glossary

Term	Definition
Accepted outcome	A task result that meets defined quality, policy, latency, and completion criteria; stricter than a generated token.
Actuation	A control-plane change applied to traffic, admission, placement, identity, tool access, or service state.
AF_XDP	A Linux interface for high-performance packet access between an XDP program and user space.
AI factory	The integrated compute, network, storage, runtime, data, and operating system that produces AI services.
Bounded control	An authorized action constrained by scope, target, duration, resource ceiling, approval, and rollback.
Control plane	The layer that observes state and authorizes policy, distinct from the data or execution path that applies it.
DPU	A data processing unit used to offload and isolate networking, storage, security, and infrastructure functions.
Evidence chain	The attributable sequence from observation and recommendation through approval, action, outcome, and recovery.
Fast path	The bounded low-latency enforcement path applying preapproved policy.
Mixture of experts	A model architecture that selects a subset of expert subnetworks for a token or routing unit.
Pattern of Life	A representation of expected multidimensional behavior over time, context, and entity; internal PTCP implementation is not disclosed here.
PTCP	Predictive Tensor Control Plane, Tensor Networks' proposed model-agnostic infrastructure control overlay.
Slow path	The higher-latency path for observation, assessment, simulation, recommendation, authorization, and evidence.
SONiC	Software for Open Networking in the Cloud, a disaggregated network operating system.
Topological containment	A policy-controlled communication or dependency envelope used to detect or restrict unauthorized relationship expansion.
TTFT	Time to first token, commonly used to characterize interactive inference latency.
Useful output per dollar	Accepted outcomes divided by fully loaded delivery cost, including retries, SLO effects, and operating cost.
XDP	eXpress Data Path, a programmable Linux packet-processing hook.

References

[1] Moonshot AI. "Kimi K3: Technical Introduction." July 16, 2026. Product and architecture announcement; accessed July 24, 2026. Source

[2] Moonshot AI. Kimi K3 quickstart and API documentation; accessed July 24, 2026. Source

[3] Moonshot AI. Kimi API pricing materials; accessed July 24, 2026. Source

[4] NSA, CISA, and international partners. "Careful Adoption of Agentic AI Services." April 30, 2026. Source

[5] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1, July 2024. Source

[6] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1, January 2023. Source

[7] NVIDIA. Dynamo design and architecture documentation; accessed July 24, 2026. Source

[8] NVIDIA. AI-factory and Spectrum-X telemetry materials; accessed July 24, 2026. Source

[9] The Linux Kernel Documentation. AF_XDP. Source

[10] The Linux Kernel Documentation. XDP redirect. Source

[11] SONiC Project. Architecture and project documentation. Source

[12] United States Patent and Trademark Office. US 11,308,384 B1, "System and method for pattern of life anomaly detection." Issued April 19, 2022. Source

[13] Federal Register. "Notice of Intent To Grant an Exclusive License With a Joint Ownership Agreement." May 3, 2023. Source

[14] Tensor Networks, Inc. PTCP product site; accessed July 24, 2026. Vendor-authored. Source

[15] Tensor Networks, Inc. "PTCP for the AI Factory." Technical white paper, July 2026. Vendor-authored proposed architecture. Source

[16] Tensor Networks, Inc. "The Empirical Tensor." Technical overview, May 2026. Vendor-authored validation agenda. Source

[17] Carahsoft Technology Corp. Tensor Networks partnership announcement, 2025. Channel evidence, not technical certification. Source

[18] MITRE. Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS). Source

Appendix A. Evidence and Claim-Status Matrix

This matrix separates public fact, architectural inference, vendor claim, and proposed test so that readers can distinguish evidence from advocacy.

Claim or proposition	Status	Basis	Required next evidence
Kimi K3 was available as a hosted product and API at the cutoff.	Public fact at cutoff	Moonshot launch and API sources [1]-[3].	Independent service testing and SLA review.
K3 full weights, final license, and technical report were available by July 24.	Not established	Moonshot scheduled release for July 27 [1].	Inspect released artifacts after publication.
Modern inference benefits from dynamic scheduling, cached-state awareness, and proactive telemetry.	Established design direction	NVIDIA architecture and telemetry sources [7], [8].	Workload-specific comparative measurement.
Agentic services require identity, per-action policy, segmentation, interruption, and rollback.	Authoritative guidance	NSA/CISA and NIST [4]-[6].	Implementation mapping and independent control assessment.
Public Linux, SONiC, and DPU-class interfaces can support bounded enforcement.	Established implementation surface	Linux and SONiC documentation [9]-[11].	Target-specific compatibility, performance, and failure tests.
Public patent and licensing records establish relevant lineage.	Public record	USPTO and Federal Register [12], [13].	Counsel-reviewed claim scope, chain of title, license rights, and freedom to operate.
PTCP improves frontier-model economics and assurance.	Unproven vendor hypothesis	Tensor-authored capability and validation agenda [14]-[16].	Independent comparative proof of value.
PTCP can create a competitive systems moat.	Strategic inference	Mechanism analysis in this paper.	Repeatable gains, portable deployment, customer willingness to pay, and operational references.

Appendix B. Pilot Acceptance Specification

The checklist converts the recommendation into a procurement-ready acceptance package. Thresholds remain illustrative until approved and pre-registered by the deploying organization.

Artifact or test	Owner	Acceptance condition	Evidence retained
System context and control ownership	Architecture council	Observation, recommendation, approval, enforcement, rollback, and incident authorities named.	Architecture, RACI, target inventory, exception log.
Threat model and trust boundaries	Product security	Observation, authorization, adapters, targets, tenant, and evidence threats assessed.	Threat model, ATLAS mapping, abuse cases, residual risks.
Target capability matrix	Network and platform	Every intended control is supported exactly or rejected explicitly; versions and ceilings recorded.	Compatibility suite, verifier or hardware reports, target inventory.
Data-quality contract	Data and evaluation	Field coverage, identity binding, clock error, late event, and sampling limits pass.	Schema, lineage, quality report, rejected-event corpus.
Baseline and trace corpus	Evaluation lead	Incumbent metrics reproducible by workload stratum; seasonality and incidents represented.	Immutable trace references, workload manifest, analysis package.
Shadow recommendations	Tensor and evaluation	Signals stable, calibrated, within overhead budget, and compared with strong baselines.	Model card at capability level, intervals, negative results; no proprietary algorithm disclosure.
Replay and failure validation	SRE and platform	Candidate outcomes reproducible; protected invariants checked; fidelity limits documented.	Scenario library, replay hash, counterfactual report.
Signing, expiry, and revocation	Security engineering	Unsigned, stale, or replayed authorization rejected; rotation and emergency revoke pass.	Test logs, key hierarchy summary, audit events.
Canary non-regression	Evaluation and risk	Primary benefit and all guardrails satisfy pre-registration; no critical safety event.	Assignment log, blinded result, uncertainty intervals, approvals.
Rollback and partition drill	SRE	Named prior state restored within objective; central loss enters declared safe mode.	Timed drill record, counters, service verification, review.
Economics and realization	Finance	Measured base case clears hurdle with conservative realization factor.	Cost inventory, sensitivity model, owner sign-off.
Exit and portability	Procurement and legal	Evidence export, removal, and fallback demonstrated; IP, support, audit, and data terms accepted.	Export package, removal test, contract schedule, legal diligence.

Acceptance authority should be divided. At minimum, an independent evaluation lead validates benefit, product security validates control safety, SRE validates recovery, finance validates realizability, legal validates rights and confidentiality, and the accountable risk owner accepts residual risk.

Appendix C. Non-Proprietary Disclosure-Control Checklist

Use this checklist before distributing the paper, creating derivative marketing collateral, or briefing a third party.

Control	Distribution test
Customer-visible outcomes only	The document explains what PTCP enables and how a customer tests it, without revealing how proprietary calculations are performed.
No enabling mathematics	No internal equations, objective functions, weights, thresholds, transformations, or derivations are included.
No proprietary model internals	No feature construction, tensor dimensions, factorization layout, model parameters, rank logic, training, calibration, or update procedure is disclosed.
No topology algorithm disclosure	No graph metric, score construction, curvature method, transport formulation, path computation, or internal causal method is specified.
No reproducible implementation recipe	No source code, pseudocode, exact schemas, compiler logic, rule maps, device recipes, privileged commands, or production configuration is included.
No confidential evidence	No customer names, traces, benchmarks, security incidents, unpublished results, or unreleased product commitments appear without authorization.
Claims remain qualified	Vendor claims are labeled; proposed thresholds are not reported results; patents and partnerships are not presented as performance proof.
Diligence route is defined	Confidential evaluation occurs under appropriate NDA, data-use, export-control, security, licensing, and clean-room provisions.

FINAL DISTRIBUTION RULE When a detail is necessary for a controlled evaluation but unnecessary for a public understanding of capability, place it in a separately governed confidential diligence package rather than this white paper.

