DATA SHEET

Securing AI Infrastructure with Terniion™

The Problem: AI Moves Fast. Its Attack Surface Moves Faster.

Your AI systems now touch everything—training data streaming in from distributed sources, GPU clusters running for weeks at a time, inference endpoints serving thousands of queries per minute, and autonomous agents calling tools with elevated privileges. Every stage of the AI lifecycle creates a new attack vector that traditional security was never built to contain.

The constraints are real:

- Training data aggregated from partners, sensors, and public repositories can be poisoned at the source, silently corrupting models before they ship.
- Inference endpoints sit directly in the blast radius of user traffic, where every query is a reconnaissance opportunity for model extraction and data exfiltration.
- Multi-agent systems share memory and communication channels; one compromised agent can cascade through the entire stack via session smuggling and context poisoning.
- MLOps pipelines carry 20+ known vulnerability classes, from model-format code execution to unauthenticated platform APIs and GPU cluster compromise.

Perimeter tools, API gateways, and VPNs were not built for this. Xiid's Terniion is.

Terniion eliminates the attack surface across the AI lifecycle—from data collection through training, inference, and agent-to-agent communication—without rewriting pipelines or replacing infrastructure.

Business Outcomes

Reduced Model and Data Risk

Process-level isolation and closed inbound ports keep training data, model weights, and RAG knowledge bases out-of-range of attackers, protecting the intellectual property at the core of your AI program against extraction, inversion, and poisoning attacks.

Reduced Operational Complexity

One overlay replaces the stack of VPNs, bastion hosts, API gateways, and firewall exceptions most AI teams stitch together to connect notebooks, training clusters, model registries, and inference endpoints across cloud, on-prem, and hybrid environments.

Lower Total Cost of Ownership

Fewer remote-access tools, fewer credentials to rotate, and fewer exposed services mean fewer hours spent on incident response, patch firefighting, and audit preparation for the same AI footprint.

Modern Security Without Downtime

Terniion overlays self-hosted MLOps and CI/CD platforms like Kubeflow, MLflow, GitLab, Jenkins, Gitea, and Bitbucket Data Center, without rip-and-replace. Data scientists keep working, models keep shipping, and pipelines keep running while the attack surface shrinks to near zero.

5 Best Practices for for Securing AI Infrastructure

Best Practice	Key Actions	Xiid Terniion Solution
Secure Data Collection and Ingestion	Protect training data in transit from edge devices, partners, sensors, and public repositories. Prevent poisoning at the point of ingestion.	SealedTunnel establishes outbound-only, triple-encrypted tunnels from data sources to central repositories. No open inbound ports on source systems. Gateway deployments aggregate data from multiple sources across hostile or unmanaged networks.
Protect Training Clusters and Compute	Keep GPU clusters, distributed training nodes, and AI accelerators off the public internet while preserving cluster coordination.	Training nodes run without public IP addresses. Outbound-only coordination enables distributed training across regions and clouds without security compromises. Process-level microsegmentation prevents lateral movement even if oneprocess is compromised.
Harden Inference Endpoints	Stop reconnaissance, model extraction, prompt injection, and data exfiltration against production AI services.	All inbound ports can be closed on inference infrastructure. Access is process-to-process, scoped to authenticated applications rather than open AI APIs. Triple-layer, post-quantum encryption protects every session end-to-end.
Isolate Agent-to-Agent Communication	Prevent session smuggling, context poisoning, and cross-agent contamination in multi-agent systems.	Each agent operates behind closed inbound ports. Agent memory stores and communication channels are reachable only through process-isolated SealedTunnel connections, so a compromised agent cannot eavesdrop on or inject into its peers.
Quantum-Secure the AI Lifecycle	Assume adversaries are harvesting encrypted training data, model weights, and inference traffic today to decrypt later.	Triple-layer, post-quantum encryption (Kyber key encapsulation, Dilithium signatures) protects data ingestion, training coordination, inference queries, and agent communications against harvest-now, decrypt-later strategies.

How Terniion Secures AI Infrastructure

Terniion is a secure network access control platform designed as an overlay. It does not require you to rip and replace networks, firewalls, or MLOps platforms.

What This Looks Like in AI Environments

Training Infrastructure and MLOps

- GPU clusters, distributed training nodes, and MLOps platforms run with all inbound ports closed.
- Data scientists, Kubeflow pipelines, and MLflow runs connect through outbound-only tunnels, scoped to specific processes rather than entire networks.
- A compromised notebook, CI/CD runner, or container cannot pivot into the training cluster, code repository, or model registry.

Inference Endpoints and Production Models

- Inference servers become non-routable from the internet — they cannot be scanned, probed, or fingerprinted by external attackers.
- Authorized client applications reach specific inference processes through process-to-process tunnels, never the underlying host, subnet, or adjacent services.
- Model weights stay protected against extraction and inversion attacks, and query data remains end-to-end encrypted throughout the inference workflow.

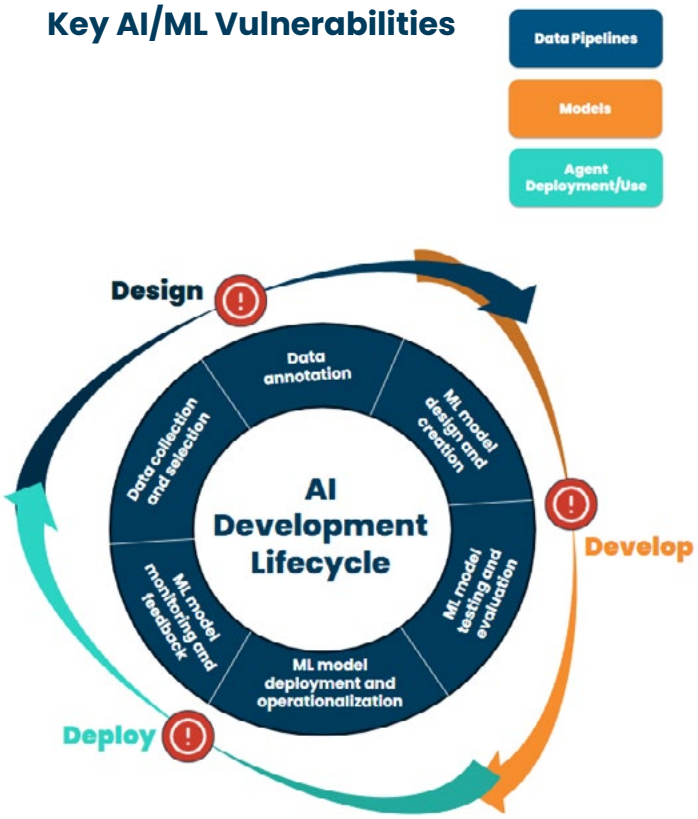
Agentic AI and Multi-Agent Systems

- Each agent runs with closed inbound ports and authenticates peers through SealedTunnel connections.
- RAG knowledge bases, vector stores (Pinecone, Weaviate, Qdrant), and document repositories are reachable only through process-isolated tunnels, preventing knowledge base poisoning.
- Agent memory and context cannot be contaminated across workloads, and every connection carries cryptographic proof of authenticity for SIEM correlation and audit.

Key Components

- **SealedTunnel**
  Executed via STLink, a lightweight software agent on communicating endpoints, these are outbound-only, triple-encrypted, process-to-process tunnels that perform reliably even over unstable networks, ideal for distributed training clusters, inference endpoints, and agent workloads.
- **Commander**
  The control brain coordinating secure connections across cloud, on-prem, and hybrid environments. It centralizes policy so “who can talk to what” becomes maintainable rules instead of 10,000 firewall exceptions.
- **Connector**
  Joins two outbound-only communication halves without ever decrypting traffic. It serves as a secure rendezvous where quantum-encrypted messages are dropped off and picked up by authorized STLinks, keeping critical infrastructure undiscoverable.
- **Aclave™**
  Credential-less authentication that lets staff and partners access systems without usernames, passwords, or long-lived credentials.

Key AI/ML Vulnerabilities



FAQS

Will this disrupt training jobs or require production downtime?

No. Terniion is designed as an overlay. It wraps existing training clusters, model registries, and inference services without firmware changes or pipeline redesigns. Rollout can be phased; start with the highest-risk assets like inference endpoints and model registries, then expand across training, MLOps, and agent infrastructure.

Do we need to replace GitLab, Kubeflow, MLflow, or our training infrastructure?

No. Terniion integrates with the self-hosted MLOps stack you already run such as Kubeflow, MLflow, GitLab, Jenkins, Gitea, Bitbucket Data Center and with your on-prem, cloud, and hybrid training infrastructure. It reduces reliance on VPNs and bastion hosts for AI workloads without forcing a wholesale rebuild. (Note: Terniion protects self-hosted platforms you control; it is not compatible with GitHub or other publicly-reachable SaaS code hosts.)

How does this help with regulators, auditors, and insurers?

Regulators and insurers increasingly focus on attack surface, lateral movement, and data protection. Terniion closes inbound ports, hardens access to sensitive training data and inference endpoints, and applies post-quantum cryptographic controls, making it easier to demonstrate reduced attack surface, containment of incidents, and alignment with modern frameworks including NIST CSF 2.0 and CNSA 2.0.

The Bottom Line

AI will be central to your business for the next decade and beyond. Your security model needs to move with it.

Terniion gives you a way to:

- Eliminate the attack surface across data collection, training, inference, and agent communications.
- Protect model intellectual property against extraction, inversion, and harvest-now, decrypt-later attacks.
- Replace a tangled stack of VPNs, gateways, and firewall rules with one overlay.
- Deploy in under 90 minutes, with no rip-and-replace of existing infrastructure.

Secure the AI systems shaping your business against today's adversaries and tomorrow's quantum-capable ones.

Ready to harden AI infrastructure without slowing down your models?
Request a briefing | See a live demo