THE AI JOURNAL

Home / Topics / AI / AI Data Centers: Securing The Next Great Attack Surface

AI

AI Data Centers: Securing The Next Great Attack Surface

By Jerry Kaner is the founder of Ciphertex® Data Security

AIJ AIJ Thought Leader • 2 days ago 4 minutes read

There are more than 5,400 data centers in the United States, according to Statista, with hundreds housing AI models used to process vast datasets. Often, this includes petabytes, and in a growing number of facilities, exabytes of sensitive information. The value of this data, coupled with the sheer volume, puts a target on these facilities that threat actors cannot resist aiming for.

Despite their heightened risk, there is no existing security standard for AI data centers. They rely on the same frameworks as traditional ones, which do not fully account for heterogeneous accelerators, cross-stack attestation and runtime behaviors unique to AI. It is imperative to establish a pragmatic profile that converts fragmented best practices into enforceable baselines that operators can measure, audit and continuously improve amid mixed fleets and evolving workloads.

The alternative is worsening systemic fragility, with these data centers becoming single points of failure in terms of competitiveness and critical infrastructure. Not taking action will undoubtedly lead to theft of intellectual property, sabotage of services and cascading outages that could have crippling consequences.

Missing Standards, High Stakes

Frontier Models are High-Value Targets

The risk of intellectual property theft carries widespread implications. Companies have invested hundreds of millions, or even billions, in developing frontier model weights, making them among the most valuable IP in the tech sector. Exfiltration collapses competitive advantage overnight, converting capital-intensive capabilities into a commodity rivals can run and fine-tune at marginal cost, eliminating API gatekeeping and undermining pricing power overnight.

The LLaMA leak demonstrated how quickly powerful weights can propagate and seed derivative ecosystems, compressing competitors' time-to-market and eroding differentiation in distribution- and compute-driven races.

Operationally, AI workloads have quickly become embedded in critical business and control workflows. Compromise in one service can ripple outward, triggering outages affecting data integrity, logistics or public safety. The high-density nature of AI computing, paired with increased interconnectivity risks, amplifies the scope and scale of failures, with localized intrusion quickly translating to regional disruption, magnifying impact exponentially.

Automating Criminal Activity with AI

Beyond diminishing corporate equity and hindering operations, model weight theft gives malicious actors access to highly advanced models that can be weaponized to automate and scale criminal operations, posing a significant threat to economic and national security.

Fraud can be conducted at scale by generating deep-fake identities, false documents and deceptive phishing messages that are difficult to detect. For example, AI models have been used to scan networks for valuable information, steal credentials and craft personalized extortion demands that maximize psychological pressure, sometimes demanding hundreds of thousands of dollars in cryptocurrency.

The automation and adaptability of these AI-powered attacks complicate defense efforts and shorten the time criminals need to execute complex fraud and extortion schemes, effectively lowering the skill barrier for large-scale cybercrime operations.

Compromised at the Source

Dependencies on high-risk regions for advanced chip packaging and rare earth materials introduce strategic supply chain risks.
Theft, interdiction or tampering, especially of GPUs, FPGAs and networking hardware, can compromise entire training environments, leading to operational, economic and reputational damage across sectors relying on trusted AI tools. A poisoned dataset introduced at one stage, for instance, may cause a model to behave unpredictably or embed vulnerabilities exploitable in critical applications such as healthcare, finance or national infrastructure.

The absence of formal standards hinders uniform auditing and incident response coordination across supply chain participants, amplifying the difficulty of containment and remediation once a breach occurs. This underscores the need for governance encompassing the full AI lifecycle, from secure development and artifact signing to provenance verification and multi-party response protocols.

Moving Toward a Coordinated, Standards-Driven Future

Securing AI data centers necessitates a unified framework that evolves in phases, starting with enforceable best practices and progressing toward defenses strong enough to deter nation-state adversaries. NIST and related agencies should lead this effort, coordinating with government, industry and academia to align incentives through procurement policies and required incident reporting.

Supply chain security must also be prioritized by verifying provenance, reducing dependencies on high-risk regions and embedding traceability and attestation into certification processes. Mandated intelligence sharing and transparent disclosure are essential to accelerate collective learning and close visibility gaps. Without collaboration, defenders will remain isolated while adversaries continue to advance.

Recommendations for Operators

Regardless of whether a formal standard is established in the near term, immediate and deliberate steps should be taken to safeguard AI data centers. Relying on legacy protocols designed for traditional facilities leaves significant vulnerabilities unaddressed, especially given the heightened sensitivity and strategic importance of AI workloads and data.

Operators must implement practical, measurable controls tailored to the AI environment. This means cataloging all AI model artifacts and verifying the integrity of their hardware, particularly across accelerators and network systems, as well as enforcing strict access protocols. Continuous risk management must be central to operations, including red teaming, telemetry-based monitoring and ongoing R&D focused on hardware and side-channel protections. Formal avenues for incident reporting and near-miss are also critical.

We Cannot Wait to Act

Throughout my career, I have seen the devastating impact of major breaches and data loss across organizations, including government agencies and institutions in the healthcare and finance sectors. I have also spent decades helping to prevent these incidents.

In all of that time, I have *never* seen an opportunity as lucrative for threat actors as that presented by AI data centers. If action is not taken now, they will quickly become an ideal attack surface for adversaries to weaponize the very systems driving the next generation of innovation.