DELL Technologies

Whitepaper

Cyber-security and cyber-resiliency for AI

Author:

Deepak Rangaraj
Product Management

Artificial Intelligence (AI) technology is finally mainstream. Recent AI model advancements along with the explosion of data, and our greatly improved ability to collect, store and process this data efficiently has led to major breakthroughs in AI. Use cases like GenAI, have made AI technology relevant and accessible to everyone. The age of AI has truly begun, and the upcoming decades are going to be defined by what can be achieved with this technology.

The field of cybersecurity with its ever-growing complexity and constant change has the potential to be revolutionized by AI. Researchers are looking for an edge in defending against cyber-attacks through innovative use of AI in anomaly detection, automated remediation, threat intelligence platforms, automated threat hunting and forensic analysis of attacks.

While AI promises to help with defense, cyber threat actors are also weaponizing it for more advanced attacks and evading detection. Gen AI capabilities can be used to create more realistic phishing emails, social engineering attacks and even generate malicious code snippets. AI tools can also be used to probe for vulnerabilities and conduct large scale automated attacks while constantly adapting to avoid detection. No doubt, AI is a double-edge sword in cybersecurity.

No enterprise wants to be left behind in the AI race and they are all rapidly moving to integrate AI into their businesses to gain an advantage. This highly complex and rapidly evolving technology has moved from niche application space and become an integral part of mission-critical workloads for enterprises. Securing AI systems has become crucial for the survival of today's enterprises and for the end users.

Most discussions on AI and cybersecurity are focused on the defense and attack aspects of using AI tools. However, the proliferation of AI workloads and their increased attack surface has made AI systems themselves a prime target for attacks. The notion of AI being the target of attacks is just starting to gain a lot more attention.

Figure 1: Facets of AI security discussions

How AI workloads differ from normal software and why securing AI systems is challenging?

AI workloads are in a sense, software running on systems. So, all the challenges of securing software workloads also apply to AI workloads. At the same, AI workloads have some unique characteristics that make them different from normal software and this adds a new set of security challenges.

In normal software the input and output can be logically mapped with an understanding of the algorithms used. AI models on the other hand are like magic black boxes that are trained on enormous amounts of data to learn the model parameters. This trained model is then fed inference data in the real world to generate the output.

The inner workings of the model (like why the parameters are a certain value etc.) cannot be fully deciphered or explained. When the model generates unanticipated results, we cannot debug it logically like traditional software. The unexpected result could be due to problems in training inputs, incorrect model parameters, malicious tampering or even due to our lack of understanding of the solution space.

The risks involved with AI systems can be roughly split into the following 5 categories.

1. Input data - Training and Inference
   There's no AI without data, and this makes their data supply chain one of the biggest sources of risks. Data could be sourced from data brokers, third party suppliers and even competitors. It is often stored across geographically distributed locations. Confidentiality, integrity, and privacy of the data is a big concern. Attackers can also manipulate or poison the training data for controlling outputs, evasion (avoid fraud or anomaly detection) or introducing systemic bias into the AI models.
2. Model data
   The AI model itself is highly a valuable business and/or mission critical IP that needs to be protected from theft and tampering. Attackers with access to the AI model can reverse engineer sensitive input data (like personal or financial information) or learn how to manipulate it for their own financial gains, evasion, or destruction.
3. Infrastructure
   The scale of infrastructure needed for AI systems, both for compute as well as data storage, is massive. Most AI systems use specialized hardware and infrastructure that is spread across cloud, on-prem, co-location and edge. These hybrid and non-homogenous environments present challenges in managing supply chain and device integrity as well as configuration and lifecycle management. Most AI inferencing happens at the edge where physically securing the devices is a challenge.
4. Human
   Human involvement in designing, managing, and running these systems introduces added risks like insider threats, inadvertent configuration and operational errors, algorithmic bias etc.
5. Compliance
   Governments and regulatory bodies are sensitive to the copious amounts of user data used by AI systems, as well as the lack of transparency and traceability of the internal operations of these systems. AI systems are also managing critical tasks like self-driving cars, healthcare, finance, utilities etc. This has resulted in several evolving regulations and guidelines to protect users' safety & data, establish user privacy and transparency rights, as well as eliminate discrimination and systemic bias in AI. Complying with these regulations is a challenge for enterprises.

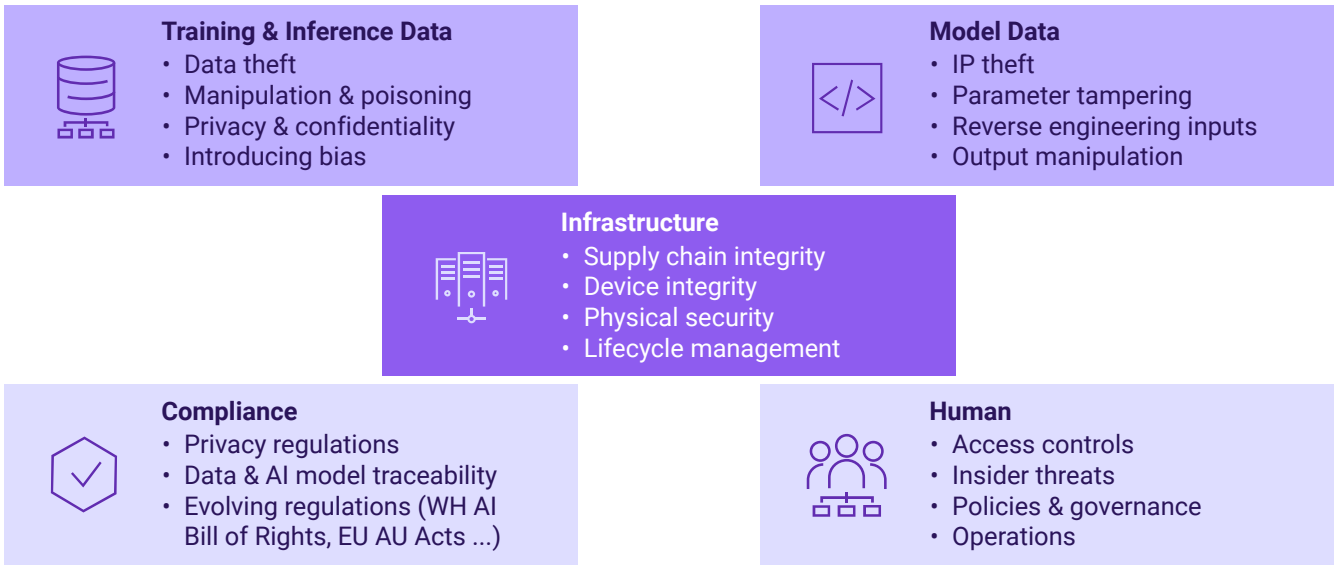
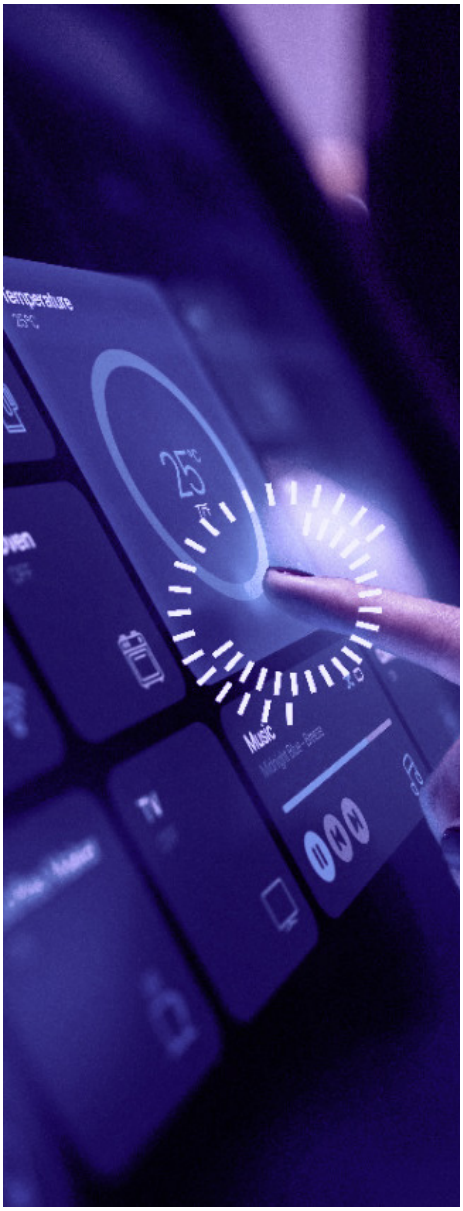


Figure 2: AI system security risks

How to secure AI systems and workloads?

Infrastructure security - Securing AI systems starts with securing their underlying infrastructure, the foundation on which these systems are built. Given the massive scale and hybrid environments of these AI systems, using a Zero Trust approach offers the best chance of success. This includes the ability to uniquely identify and attest the integrity of every component of the infrastructure throughout the lifecycle of the system and authenticating both users and devices before granting access. The top things to consider for securing the infrastructure are as follows.

- Ensure infrastructure components are Secure by Design – with security built into them from the very beginning of the design process.
- Supply chain is secure – Protection from tampering, insertion of malware, counterfeit components, or malicious implants in products all the way from sourcing, manufacturing to delivery.
- Integrity of devices is maintained – security is anchored to an immutable hardware-based Root of Trust (RoT), cryptographic verification of all firmware/software, cryptographic device identity and attestation.
- Physical security, especially for edge inference – ability to detect intrusion, generate alerts, erase data, and remotely shut down or power cycle devices.
- Asset tracking and vulnerability management - Track hardware and software assets in the datacenter using bill of materials (HBOM & SBOM) and monitor reported vulnerabilities to identify at-risk components.
- Stay current with patches and updates – Attackers exploit older vulnerabilities on unpatched systems more frequently than recently found vulnerabilities. So, timely patching of systems is the most critical action in reducing the risk of attacks.



Data Protection - The next step in securing the AI systems is to secure the algorithm (model parameters, hyperparameters, and other IP) and data (training and inference). This relies heavily on the data protection capabilities of the infrastructure, as well as the enterprise's data governance and access policies. The top things to consider for securing data are as follows.

- Data supply chain integrity – Whether data is generated and collected internally or sourced externally from data brokers, 3rd party public repositories etc., it is critical to ensure that the data is trustworthy and has not been tampered with. Like supply chain of physical products, the data supply chain integrity must be ensured through data supplier qualification, quality assessment, protection, and access controls.
- Data storage protection – All data must be encrypted, regularly backed up and stored on drives that are FIPS certified. The keys used to encrypt the data must be stored on external key management servers. So even if the disks are physically stolen from the servers (e.g., inference devices at the Edge), the keys will not go with the stolen disks. Using hardware-based encryption like Self-encrypting drives (SED) will minimize performance impacts.
- Trusted Execution Environments (TEE) for data usage – Training/ inference data and AI model data/IP are most vulnerable when being actively used. It is critical to establish Trusted Execution Environments on the servers and use the data only within these TEEs. Confidential compute technologies from CPU and GPU vendors provide high-level of security and isolation needed for these TEEs. They help maintain the confidentiality, privacy, and integrity of the data in use. Whether the AI workload is running in shared cloud environments or being trained on data in external data vendor servers or AI inference is run on end user systems, these confidential compute technologies provide isolation between the workload and lower layers of the infrastructure including VM, Hypervisor, OS etc.
- Data in flight protection – AI systems are often run in distributed environments leading to massive amounts of data transfer over the network. All such data traffic and network communications must be encrypted using standards like TLS 1.3 and HTTPS along with link level encryption in the servers.

Management and Monitoring

The final step in securing the AI systems is to create and enforce the right set of policies to govern everything from access to the system and stored data, compliance, operations, orchestration, infrastructure management, to retiring or decommissioning the systems. Frameworks like the NIST AI Risk Management Framework and the accompanying playbook provide excellent guidance to enterprises in creating these policies.

Once the policies have been identified they need to be enforced. Infrastructure capabilities, software controls and application layer tools (SIEM, XDR etc.) are all critical in setting up and enforcing these policies. Infrastructure management tools like iDRAC, OpenManage Enterprise and APEX AIOps, provide enterprises with a flexible set of controls, telemetry, and automation at the infrastructure layer. They enable enterprises to manage and operate AI system infrastructure according to their security sensitivity and risk tolerance. Some of the main capabilities include

- Continuous monitoring and recovery – No systems can be 100% secured, so cyber resiliency is critical. Infrastructure must have capabilities to lock down configurations and settings, track compliance, continuously monitor for any drift, detect anomalies and ongoing attacks, and automatically recover to a known good state.
- Data governance processes and policies – Needed to manage risk and ensure regulatory compliance. They must be in line with Zero Trust principles of least privilege access and always verify, to protect from both outside attacks and insider threats.
- User Identity and Access Management – Strong Multi-factor Authentication (MFA), Single sign-on (SSO) and Directory management capabilities are all critical to avoid phishing attacks, privilege escalations and lateral movement of threats.
- Fine-grained telemetry and observability – The right level of visibility into the system signals and events is essential for automated monitoring and recovery.
- Secure retirement and decommissioning – Sanitization of systems through secure erasure of all drives, configuration and other sensitive data is essential to confidently retire or repurpose systems.

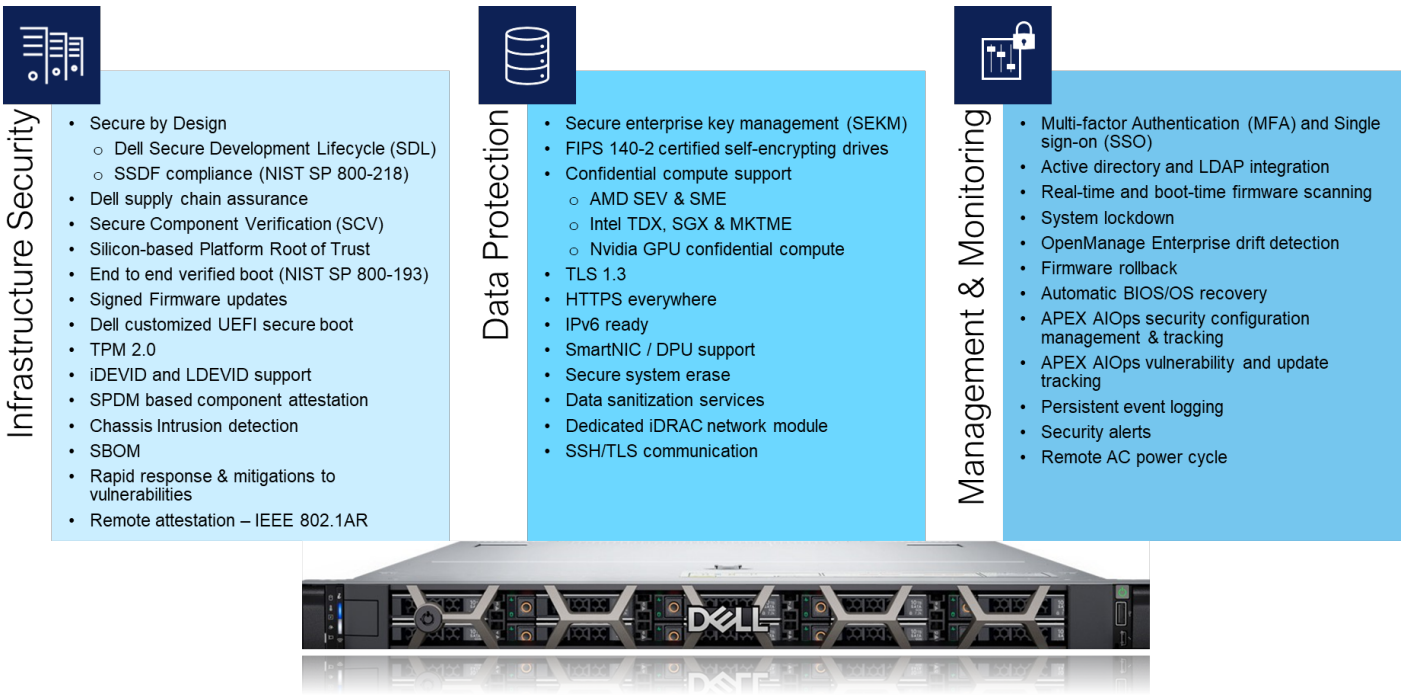


Figure 3: Securing AI systems with Dell PowerEdge

Dell Technologies and PowerEdge advantage

AI systems are complicated and introduce their own set of challenges to enterprises, like dealing with enormous amounts of data, compliance requirements, lack of transparency and full understanding of the internal workings of the AI models etc. At Dell Technologies, we believe that customers shouldn't have to add securing AI system infrastructure to their growing list of challenges and complexity.

We strive to make it easy for our customers to securely run AI workloads on our infrastructure platforms like PowerEdge servers. We have simple, standardized tools and smart automation built into our products to help you on every step of your journey whether you are deploying, configuring, managing, or retiring your servers.

Our PowerEdge servers have all the capabilities needed to secure your infrastructure, protect your data, and secure your operations and lifecycle management. Learn more about PowerEdge server security and cyber-resiliency capabilities and simplify your AI infrastructure security journey.

Learn more about PowerEdge security

Learn more about Dell solutions

Contact a Dell Technologies Expert

View more resources

Join the conversation with #PowerEdge