**AI Security: Threat Landscape, Attack Surfaces, Frameworks, and Regulations**

**The AI-Specific Threat Landscape**

AI systems face a range of novel cyber threats that differ from traditional IT risks. Common attack types include **data poisoning** (tampering with training data to alter model behavior) and **backdoor injections**, which can stealthily implant malicious triggers into models[ibm.com](https://www.ibm.com/think/topics/data-poisoning#:~:text=Clean). **Adversarial evasion** is another classic threat: carefully crafted inputs cause misclassification or unsafe outputs. In generative AI, **prompt injection** (malicious user inputs) can force models to reveal sensitive information or execute unintended actions[ibm.com](https://www.ibm.com/think/topics/data-poisoning#:~:text=While%20data%20poisoning%20and%20prompt,data%2C%20spreading%20misinformation%20or%20worse). Attackers may also perform **model inversion or extraction**, recovering private training data or cloning models. OWASP notes that **model theft** (unauthorized copying of a proprietary model) is a key risk (their “LLM10” entry)[owasp.org](https://owasp.org/www-project-top-10-for-large-language-model-applications/#:~:text=LLM10%3A%20Model%20Theft). Other concerns include **membership inference** (determining if a specific datum was in the training set) and **bias/amplification attacks**. Collectively, these AI-specific threats can compromise confidentiality, integrity, or availability of AI-powered services, and must be addressed through secure design, monitoring, and validation.

**The AI Attack Surface: Data, Model, and Deployment Risks**

AI extends the attack surface at every layer – from data to deployment. A 2025 industry survey found **94%** of organizations expect AI will *worsen* their attack surface, citing risks like sensitive data leakage, ungoverned “shadow AI,” and complex compliance issues[trendmicro.com](https://www.trendmicro.com/en_us/research/25/g/ai-cyber-risks.html#:~:text=However%2C%20as%20optimistic%20as%20IT,5%20years). Key exposed areas include:

* **Training & Operational Data:** All data used to train or serve AI (personal profiles, proprietary info, etc.) can be exfiltrated or manipulated. For example, poisoned or tampered training data can systematically skew model behavior[ibm.com](https://www.ibm.com/think/topics/data-poisoning#:~:text=Clean). Leaked training data (through inversion attacks) can breach privacy. This layer also includes logs and model outputs that may inadvertently reveal secrets.
* **Models & APIs:** Proprietary models themselves become targets. Attackers may steal model weights or replicate model behavior (compromising intellectual property and model confidentiality[owasp.org](https://owasp.org/www-project-top-10-for-large-language-model-applications/#:~:text=LLM10%3A%20Model%20Theft)). Exposed inference APIs invite **adversarial and prompt attacks**: an attacker can craft inputs that bypass filters or trigger unsafe outputs. OWASP’s LLM Top-10 explicitly highlights prompt injection and insecure output handling as critical API-level threats[owasp.org](https://owasp.org/www-project-top-10-for-large-language-model-applications/#:~:text=OWASP%20Top%2010%20for%20Large,1).
* **Deployment Infrastructure:** AI services often run in complex environments (cloud platforms, edge devices, container ecosystems). This adds new vulnerabilities: researchers have documented attacks like **malicious cloud compute hijacking** and **API misuse**. For instance, attackers can submit heavy workloads to exhaust AI resources, or exploit misconfigured cloud identity controls. Vectra notes examples in AWS Bedrock (a cloud AI service) where *malicious prompts* and *compute hijacking* were observed[vectra.ai](https://www.vectra.ai/blog/ai-is-now-the-attack-surface-why-your-security-stack-must-adapt-fast#:~:text=In%20our%20recent%20blog%20on,adversary%20behavior%20against%20AI%20systems). In short, every component – from data pipelines to compute clusters – must be hardened.

Together, these factors mean that AI projects must secure not just traditional IT assets but also protect data integrity, model confidentiality, and AI service endpoints. Integrating AI security into the broader attack-surface management program is now essential[trendmicro.com](https://www.trendmicro.com/en_us/research/25/g/ai-cyber-risks.html#:~:text=However%2C%20as%20optimistic%20as%20IT,5%20years)[ibm.com](https://www.ibm.com/think/topics/data-poisoning#:~:text=While%20data%20poisoning%20and%20prompt,data%2C%20spreading%20misinformation%20or%20worse).

**Adversarial ML Frameworks: MITRE ATLAS**

MITRE ATLAS (Adversarial Threat Landscape for AI Systems) is a dedicated framework cataloging AI-specific attack tactics and techniques[nightfall.ai](https://www.nightfall.ai/ai-security-101/mitre-atlas#:~:text=MITRE%20ATLAS%20,in%20AI%20and%20LLM%20security). Organized much like MITRE ATT&CK, ATLAS documents known adversarial actions (e.g. data poisoning, model evasion, stealing model outputs) along with real-world case studies. Security teams use ATLAS to **threat-model AI systems**: by reviewing the listed tactics and techniques, defenders can anticipate how attackers might compromise AI pipelines[nightfall.ai](https://www.nightfall.ai/ai-security-101/mitre-atlas#:~:text=MITRE%20ATLAS%20,in%20AI%20and%20LLM%20security). In practice, ATLAS helps analysts map their AI deployment to potential adversary behaviors, ensuring the organization considers AI-focused exploits in its threat assessments.

**Distinguishing Threats: Attacks *on* AI vs. Attacks *using* AI**

It’s important to separate attacks **targeting AI components** from attacks **leveraging AI as a tool**. U.S. policy explicitly makes this distinction. “Attacks Using AI” refers to adversaries *using AI capabilities* (e.g. machine-learning-driven phishing, automated vulnerability discovery, AI-generated disinformation) to amplify their impact. In contrast, “Attacks Targeting AI Systems” means strikes *against the AI itself* – such as corrupting training data or hacking ML models[dhs.gov](https://www.dhs.gov/archive/ai/promoting-ai-safety-and-security#:~:text=,that%20affect%20critical%20infrastructure%20operations). For example, poisoning a training set is an attack *on* AI, whereas using AI to generate more convincing fake emails is an attack *with* AI. DHS guidance and surveys underscore this duality: AI represents a “multi-sided threat” – it is both a new target and a powerful new weapon[dhs.gov](https://www.dhs.gov/archive/ai/promoting-ai-safety-and-security#:~:text=,that%20affect%20critical%20infrastructure%20operations)[trendmicro.com](https://www.trendmicro.com/en_us/research/25/g/ai-cyber-risks.html#:~:text=AI%20represents%20a%20multi,approach%20to%20cyber%20risk%20management). Indeed, over half of security executives expect AI to enable more complex attacks even as it introduces new vulnerabilities in AI systems[trendmicro.com](https://www.trendmicro.com/en_us/research/25/g/ai-cyber-risks.html#:~:text=AI%20represents%20a%20multi,approach%20to%20cyber%20risk%20management). Understanding this helps organizations defend appropriately on both fronts (e.g. securing AI pipelines *and* hardening defenses against AI-driven attackers).

**Key Industry Frameworks and Standards**

Several frameworks have emerged to guide AI security and governance across industries:

* **MITRE ATLAS** (see above) – a comprehensive adversarial AI tactics matrix[nightfall.ai](https://www.nightfall.ai/ai-security-101/mitre-atlas#:~:text=MITRE%20ATLAS%20,in%20AI%20and%20LLM%20security).
* **NIST AI RMF (Risk Management Framework)** – a U.S. government standard for AI risk management. It emphasizes *trustworthiness* concepts (fairness, transparency, robustness) and organizes guidance into four core functions: **Map, Measure, Manage, Govern**[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=,Map%2C%20measure%2C%20manage%2C%20and%20govern). Organizations use it to embed risk assessments and controls into the AI lifecycle.
* **OWASP GenAI/LLM Top-10** – community-driven security guidelines for generative AI. The OWASP Top-10 for LLM Applications (2025) identifies critical vulnerabilities (prompt injection, insecure outputs, data poisoning, model DoS, supply-chain risks, etc.) to address during development[owasp.org](https://owasp.org/www-project-top-10-for-large-language-model-applications/#:~:text=OWASP%20Top%2010%20for%20Large,1). It provides concrete examples and mitigations for each.
* **ISO/IEC 42001** – an international management-system standard (released 2023) for AI governance[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=ISO%2FIEC%2042001%2C%20the%20international%20management,technique%20to%20meet%20those%20expectations). Analogous to ISO 27001 for cybersecurity, it specifies requirements for an “AI Management System” covering risk assessment, control implementation, monitoring, and continuous improvement[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=ISO%2FIEC%2042001%2C%20the%20international%20management,technique%20to%20meet%20those%20expectations)[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=,data%20protection%2C%20and%20adversarial%20threats).
* **CSA AI Controls Matrix** – a new framework by the Cloud Security Alliance. It defines 243 security controls across 18 domains (audit, application security, cryptography, data protection, model security, etc.), explicitly mapped to ISO 42001 and the NIST AI RMF[tenable.com](https://www.tenable.com/blog/cybersecurity-snapshot-ai-security-framework-ai-security-controls-07-11-2025#:~:text=The%20%E2%80%9CArtificial%20Intelligence%20Controls%20Matrix%E2%80%9D,1). This vendor-neutral matrix helps organizations systematically protect AI systems from design through operation.
* **SANS/OWASP AI Security Controls** – a collaborative project (SANS Institute and OWASP) to produce an actionable AI security control set, grouped into categories like Access, Data, Deployment, Inference, Monitoring, and Governance[tenable.com](https://www.tenable.com/blog/cybersecurity-snapshot-ai-security-framework-ai-security-controls-07-11-2025#:~:text=The%20organizations%20will%20mash%20up,teams%20can%20adopt%20right%20away). These efforts align existing knowledge (OWASP AI Exchange) into practical guidance for defenders.

Together, these and other standards (e.g. IEEE P7000-series ethical AI standards, sector-specific guidelines) provide structured approaches to AI security and governance. They ensure organizations have checklists and best practices for secure model development, testing, and operations.

**NIST AI Risk Management Framework (AI RMF)**

The **NIST AI Risk Management Framework (AI RMF)**, published January 2023, is a foundational U.S. standard for AI. It was developed collaboratively (public comments, workshops) and is **voluntary**[nist.gov](https://www.nist.gov/itl/ai-risk-management-framework#:~:text=In%20collaboration%20with%20the%20private,AI%20products%2C%20services%2C%20and%20systems). The RMF’s goal is to help organizations “incorporate trustworthiness considerations into the design, development, use, and evaluation of AI”[nist.gov](https://www.nist.gov/itl/ai-risk-management-framework#:~:text=In%20collaboration%20with%20the%20private,AI%20products%2C%20services%2C%20and%20systems). It introduces AI-tailored risk concepts like explainability, fairness, and robustness. Crucially, it structures guidance into four functions[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=,Map%2C%20measure%2C%20manage%2C%20and%20govern):

* **Map:** Identify the context in which an AI system operates (intended use, stakeholders, regulatory requirements).
* **Measure:** Evaluate the AI system’s attributes (performance, reliability, privacy, security) using metrics and tests.
* **Manage:** Implement safeguards and processes to mitigate identified risks (e.g. bias mitigation, security patches).
* **Govern:** Establish oversight, policies, and accountability (roles, audits, external review) for AI systems.

These functions are iterative and cross-cutting. In addition to RMF 1.0, NIST released (March 2023) a *Generative AI Profile* (NIST-AI-600-1) to address risks unique to LLMs. Organizations adopt the AI RMF to complement broader risk programs (integrating with ISO 31000 or enterprise RM). The AI RMF Playbook and crosswalks to other standards are available for implementation guidance. In sum, NIST AI RMF provides a *trusted blueprint* for embedding security and ethics into AI projects[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=,Map%2C%20measure%2C%20manage%2C%20and%20govern)[nist.gov](https://www.nist.gov/itl/ai-risk-management-framework#:~:text=In%20collaboration%20with%20the%20private,AI%20products%2C%20services%2C%20and%20systems).

**OWASP Top 10 for LLMs**

OWASP’s **GenAI Security Project** maintains a “Top 10” list of the most critical vulnerabilities in LLM-driven applications. The 2025 list (v1.1) includes:

* **LLM01: Prompt Injection** – adversaries craft inputs to make LLMs reveal hidden instructions or data[owasp.org](https://owasp.org/www-project-top-10-for-large-language-model-applications/#:~:text=OWASP%20Top%2010%20for%20Large,1).
* **LLM02: Insecure Output Handling** – failing to sanitize or validate AI outputs, allowing downstream exploits (e.g. code execution).
* **LLM03: Training Data Poisoning** – maliciously tampering with training datasets to bias or subvert the model.
* **LLM04: Model Denial of Service** – resource exhaustion attacks (overloading LLMs with prompts) to disrupt availability.
* **LLM05: Supply Chain Vulnerabilities** – compromised third-party components (libraries, pre-trained models, plugins) introducing weaknesses.
* **LLM06: Sensitive Information Disclosure** – accidental leaking of confidential data in AI responses.
* **LLM07: Insecure Plugin Design** – poorly secured LLM plugins (e.g. allowing RCE).
* **LLM08: Excessive Agency** – giving LLMs too much autonomy (e.g. chain-of-thought tasks) without proper constraints.
* **LLM09: Overreliance** – users failing to critically evaluate LLM outputs, leading to flawed decisions.
* **LLM10: Model Theft** – unauthorized copying or distribution of the LLM itself[owasp.org](https://owasp.org/www-project-top-10-for-large-language-model-applications/#:~:text=LLM10%3A%20Model%20Theft).

Each entry is documented with real-world examples and mitigation strategies. Developers of AI-powered apps should consult this list (and the accompanying guidance) to harden their systems. For instance, to counter prompt injection (LLM01), OWASP recommends strict input filtering and output verification. By following the OWASP LLM Top-10, teams ensure they address the most prevalent GenAI risks identified by the community[owasp.org](https://owasp.org/www-project-top-10-for-large-language-model-applications/#:~:text=OWASP%20Top%2010%20for%20Large,1)[owasp.org](https://owasp.org/www-project-top-10-for-large-language-model-applications/#:~:text=LLM10%3A%20Model%20Theft).

**ISO/IEC Standards for AI Management (ISO/IEC 42001)**

The **ISO/IEC 42001:2023** standard establishes an *AI Management System* (AIMS) – a formal governance framework for AI. Analogous to ISO 27001 for security or ISO 9001 for quality, 42001 prescribes high-level requirements for organizations to use AI responsibly. Key points of ISO 42001 include:

* **Lifecycle Governance:** It requires applying governance processes at each stage (from inception to retirement) of the AI lifecycle[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=ISO%2FIEC%2042001%2C%20the%20international%20management,technique%20to%20meet%20those%20expectations). This includes planning, risk assessment, validation, monitoring, and decommissioning.
* **Risk Management:** Organizations must identify AI-specific risks (security, privacy, ethical) and implement controls to mitigate them[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=ISO%2FIEC%2042001%2C%20the%20international%20management,technique%20to%20meet%20those%20expectations). It mandates risk assessments and continuous improvement, much like other ISO management standards.
* **Alignment with Other Standards:** ISO 42001 references ISO/IEC 22989 (AI concepts) and integrates with general risk frameworks (e.g. ISO 31000)[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=%2A%20ISO%2031000%3A%20A%20general,ERM%29%20programs). For example, AWS notes that organizations might use ISO 31000 or the NIST AI RMF in parallel with 42001 to fulfill its risk assessment requirements[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=,Map%2C%20measure%2C%20manage%2C%20and%20govern).
* **Structured Process:** The standard encourages threat modeling (e.g. STRIDE or OWASP for ML) and impact assessments throughout development[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=,data%20protection%2C%20and%20adversarial%20threats). Clause 6 and 8 of 42001 focus on risk assessment and control implementation; clauses 9 and 10 require monitoring and review.
* **Accountability & Ethics:** It emphasizes documenting decisions (governance policies, audit trails) and ensuring accountability for AI outcomes. High-risk “AI impact assessments” (akin to GDPR DPIAs) are explicitly recommended for ethically or legally sensitive applications[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=After%20an%20organization%20has%20identified,DPIAs%20are%20specifically%20designed%20to).

By certifying to ISO/IEC 42001, organizations can demonstrate they have a systematic AI governance program. In practice, this means formalizing roles (e.g. AI risk officers), defining processes for model validation, and embedding controls for explainability and fairness. The standard is intended for any organization using AI – from enterprises to governments – and supports international consistency in AI risk management.

**The Global Regulatory Environment**

AI is a global phenomenon, and governments worldwide are rushing to regulate it – but approaches differ widely. Common themes are emerging, yet implementations vary:

* **Varied Definitions and Scope:** Countries often define “AI” differently. For example, the EU’s AI Act provides a broad definition of an “AI system” (based on OECD’s definition), whereas some jurisdictions (UK, Israel, Japan, etc.) rely on sector-specific interpretations or none at all[whitecase.com](https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china#:~:text=1.%20,not%20currently%20provide%20a%20comprehensive). There is no single global definition, so multinational firms may need to adopt the strictest definition that applies to any market[whitecase.com](https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china#:~:text=1.%20,not%20currently%20provide%20a%20comprehensive).
* **Different Legal Forms:** Some governments legislate new AI laws (e.g. South Korea’s AI Framework Act passed Jan 2025 for enforcement in 2026), while others issue executive orders or guidelines[whitecase.com](https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china#:~:text=2,state%20initiatives%2C%20and%20actions%20by). The EU’s AI Act is a binding **Regulation** (directly enforceable in all member states), but the UK opted for non-binding principles enforced by existing regulators. The US has used a mix of EOs and amendments to existing laws. As a result, compliance obligations can look quite different from one country to the next[whitecase.com](https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china#:~:text=2,state%20initiatives%2C%20and%20actions%20by).
* **Binding vs. Advisory:** Some rules are mandatory (the EU Act, China’s pending rules) while others are voluntary or advisory (OECD Guidelines, industry codes of practice). For example, the OECD and UNESCO have published AI principles that encourage transparency and safety, but lack enforcement power. In contrast, the EU Act and evolving national AI laws carry penalties for non-compliance. This patchwork means companies must track both enforceable laws and best-practice frameworks[whitecase.com](https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china#:~:text=3,the%20disadvantage%20that%20different%20UK).
* **Fragmentation & Overlap:** Many jurisdictions overlap AI regulation with related areas. Data protection, safety, and security laws already apply to AI (e.g. GDPR for personal data in AI), and new AI rules often reference these existing frameworks. Therefore, a company might face parallel obligations under AI law and, say, consumer protection or export control rules[whitecase.com](https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china#:~:text=5,some%20cases%20imposes%20special%20compliance). Emerging AI regulations also often have extra-territorial reach, meaning an AI tool available globally must meet the strictest applicable standard (a “highest common denominator” compliance strategy)[whitecase.com](https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china#:~:text=1.%20,not%20currently%20provide%20a%20comprehensive).

In practice, this global environment forces organizations to build flexible compliance programs. They must identify applicable AI regulations (e.g. by country, by sector) and often implement the most stringent requirements uniformly (for example, meeting EU Act standards even outside the EU). International coordination efforts (G7 AI Safety Summit, OECD’s AI Policy Observatory) aim to harmonize principles, but for now businesses must navigate a complex, evolving patchwork.

**The European Union AI Act**

The **EU AI Act** (adopted June 2024) is the first comprehensive AI regulation. It establishes a **risk-based framework** for AI systems:

* **Banned (“Unacceptable”) AI:** Certain applications are outright prohibited. These include AI that manipulates people via subliminal techniques, social scoring that profiles individuals by behavior, and real-time biometric identification in public spaces (e.g. live facial recognition)[europarl.europa.eu](https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=Unacceptable%20risk). Some narrow exceptions exist (e.g. serious crime investigation under strict oversight).
* **High-Risk AI:** Any system that can significantly impact safety or fundamental rights is “high-risk.” This covers two categories: (1) AI embedded in products already regulated by EU safety laws (automotive, medical devices, etc.), and (2) stand-alone AI used in critical domains (e.g. managing critical infrastructure, education, employment, essential services, law enforcement, border control)[europarl.europa.eu](https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=High%20risk). All high-risk systems must undergo **rigorous conformity assessments** before deployment and be monitored throughout their lifecycle. Providers of high-risk AI must meet transparency, human oversight, and robustness requirements. Users also have obligations (e.g. logging, human check).
* **Transparency Obligations:** General-purpose generative AI (like advanced LLMs) and certain AI outputs fall under transparency rules. For example, content produced by AI must be clearly labeled as AI-generated, and models should be designed to prevent illicit outputs[europarl.europa.eu](https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=Generative%20AI%2C%20like%20ChatGPT%2C%20will,requirements%20and%20EU%20copyright%20law). Providers of very large language models must register with the EU and may need to publish summaries of copyrighted data in their training sets[europarl.europa.eu](https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=Generative%20AI%2C%20like%20ChatGPT%2C%20will,requirements%20and%20EU%20copyright%20law).
* **Timeline & Scope:** The Act became largely applicable in late 2024, with immediate bans on unacceptable uses (effective Feb 2025) and phased compliance deadlines for others (full compliance by 2027 for most high-risk systems)[europarl.europa.eu](https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=In%20June%202024%2C%20the%20EU,parts%20will%20be%20applicable%20sooner). Crucially, it has **extraterritorial scope**: any company (inside or outside the EU) supplying AI products or services to the EU market must comply.

In effect, the EU Act creates the world’s first unified AI regulatory environment, setting a global benchmark. Companies deploying AI into EU states must classify their systems by risk and implement the mandated controls. For example, an AI-driven network monitoring tool from Ciena used in EU critical infrastructure would likely be considered high-risk and require certification under this law.

**U.S. White House Executive Orders on AI**

In the U.S., AI policy has so far been driven by Executive Orders rather than new legislation. Two recent EOs are most notable:

* **EO 14110 (Oct 30, 2023)**: Titled *“Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,”* this order made AI safety and ethics a national priority. It directed NIST (with DHS and DOE) to **develop AI security and privacy guidelines** and accelerate the creation of standards and best practices for AI safety[federalregister.gov](https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence#:~:text=4,Commerce%20may%20deem%20appropriate%2C%20shall). For example, it called for (within 270 days) companion resources to the AI RMF for generative AI, and a Secure Software Development Framework extension for AI[federalregister.gov](https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence#:~:text=4,Commerce%20may%20deem%20appropriate%2C%20shall). Crucially, it mandated that companies building *“potentially dangerous”* or *“frontier”* AI models share information with the government. Within 90 days, the Secretary of Commerce was to require firms training advanced (dual-use) models to report on their development activities, cybersecurity controls, and even results of red-team safety tests[federalregister.gov](https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence#:~:text=4,Secretary%20of%20Commerce%20shall%20require). The order also tasked agencies to define thresholds (e.g. compute usage beyond 10^26 operations) that trigger these reporting rules[federalregister.gov](https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence#:~:text=,power%20available%20in%20each%20cluster). In summary, EO 14110 aimed to impose strong oversight: if you are training cutting-edge AI, you must engage with the government on safety.
* **EO 14179 (Jan 23, 2025)**: Titled *“Removing Barriers to American Leadership in Artificial Intelligence,”* this order rescinded much of EO 14110. Signed by President Trump, it revoked that prior order’s provisions and instead emphasizes **accelerating AI innovation**[whitehouse.gov](https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/#:~:text=Section%201,global%20leadership%20in%20artificial%20intelligence)[whitehouse.gov](https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/#:~:text=the%20heads%20of%20all%20agencies,propose%20suspending%2C%20revising%2C%20or%20rescinding). It instructs the White House’s science and security advisors to develop an *AI action plan* to maintain U.S. leadership, with an explicit focus on ensuring no “ideological bias” impedes innovation[whitehouse.gov](https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/#:~:text=Section%201,global%20leadership%20in%20artificial%20intelligence). The order directs a sweeping review (within 180 days) of all federal AI policies (including parts of EO 14110) to identify and remove regulations seen as hindering innovation[whitehouse.gov](https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/#:~:text=the%20heads%20of%20all%20agencies,propose%20suspending%2C%20revising%2C%20or%20rescinding). In practice, this means the mandatory disclosure and red-teaming requirements of 14110 are being rolled back. The new U.S. policy stresses free-market development and coordination, whereas the previous administration’s order stressed government-led safety controls.

Taken together, these shifts illustrate how U.S. AI governance is currently in flux. Organizations using AI must watch for changing policies. Under EO 14110, firms were gearing up for mandatory reporting and adherence to NIST guidelines. Now, they may see those obligations loosened, but still can expect voluntary federal guidance (NIST is continuing AI standards work) and potential future rules (Congress and agencies are also studying AI regulation). In any case, both EOs underscore U.S. intent to balance AI leadership with security: even under the new order, ensuring robust cybersecurity for AI infrastructure remains a stated priority[federalregister.gov](https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence#:~:text=4,Commerce%20may%20deem%20appropriate%2C%20shall)[whitehouse.gov](https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/#:~:text=Section%201,global%20leadership%20in%20artificial%20intelligence).

**Implications for a Global Company like Ciena**

For a multinational technology company (e.g. Ciena, a global network infrastructure firm), these developments have concrete consequences. Key implications include:

* **Adopting AI Risk Governance:** Ciena should integrate AI risk management into its existing cybersecurity governance. This means aligning with standards like ISO/IEC 42001 for responsible AI governance[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=ISO%2FIEC%2042001%2C%20the%20international%20management,technique%20to%20meet%20those%20expectations) and using NIST’s AI RMF functions to assess and mitigate risk in its AI-enabled products[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=,Map%2C%20measure%2C%20manage%2C%20and%20govern). For example, during development of an AI-driven network monitoring tool, Ciena could perform ISO 42001–style risk assessments (documenting purposes, stakeholders, risks) and apply threat modeling (STRIDE or MITRE ATLAS) to identify potential exploits at each stage[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=,trustworthy%20and%20responsible%20AI%20development).
* **Implementing Industry Controls:** Ciena can leverage the industry control frameworks to harden its AI systems. Using the CSA AI Controls Matrix would ensure comprehensive coverage (from data encryption to model integrity)[tenable.com](https://www.tenable.com/blog/cybersecurity-snapshot-ai-security-framework-ai-security-controls-07-11-2025#:~:text=The%20%E2%80%9CArtificial%20Intelligence%20Controls%20Matrix%E2%80%9D,1). The OWASP GenAI Top-10 provides immediate guidance on common pitfalls (e.g. vet user prompts, secure LLM APIs)[owasp.org](https://owasp.org/www-project-top-10-for-large-language-model-applications/#:~:text=OWASP%20Top%2010%20for%20Large,1). Embedding these controls into product design and CI/CD pipelines helps Ciena meet international best practices even before formal regulation.
* **Regulatory Compliance Across Markets:** Ciena must prepare for differing legal requirements. In the EU, if any of its AI features count as “high-risk” (for instance, AI managing critical telecom services), Ciena would need to demonstrate conformity under the AI Act – including detailed documentation, risk assessments, and possibly third-party audits[europarl.europa.eu](https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=High%20risk). It must also label AI-generated content and adhere to transparency rules for any generative AI components[europarl.europa.eu](https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=Generative%20AI%2C%20like%20ChatGPT%2C%20will,requirements%20and%20EU%20copyright%20law). In the U.S., while mandatory requirements are less prescriptive, Ciena should be ready to follow NIST guidelines and possibly respond to federal inquiries (e.g. share model security measures) if EO 14110-like rules were reinstated. Globally, differing definitions of “AI” (and when AI laws apply) mean Ciena must likely design to the most stringent standard it faces[whitecase.com](https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china#:~:text=1.%20,not%20currently%20provide%20a%20comprehensive). Practically, this could involve auditing its AI supply chain (to ensure no forbidden components under any country’s law) and keeping thorough records of model training data to answer regulators.
* **Ongoing Monitoring and Innovation:** AI security is evolving rapidly. Ciena should establish processes for continuous monitoring of AI risks (as ISO 42001 advises) and stay engaged with emerging standards. This includes staying informed on things like OWASP updates, NIST playbooks, and global regulatory changes. Participation in industry consortia (e.g. IETF or ETSI working groups on AI) can help Ciena influence and anticipate future rules.

In summary, Ciena (like any global tech firm) must embed AI safety into its corporate risk program. By proactively applying frameworks (NIST RMF, ISO 42001, CSA Controls) and adopting security best practices (MITRE ATLAS, OWASP Top-10, robust software development), Ciena will both protect its AI-driven products and streamline compliance with EU, US, and other global regulations[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=ISO%2FIEC%2042001%2C%20the%20international%20management,technique%20to%20meet%20those%20expectations)[tenable.com](https://www.tenable.com/blog/cybersecurity-snapshot-ai-security-framework-ai-security-controls-07-11-2025#:~:text=The%20%E2%80%9CArtificial%20Intelligence%20Controls%20Matrix%E2%80%9D,1). Such efforts not only reduce legal and security risks but also build customer trust by demonstrating that Ciena takes AI security and responsibility seriously across all markets.

**Sources:** Authoritative industry and academic reports and official documents were used, including technical surveys on AI threats[ibm.com](https://www.ibm.com/think/topics/data-poisoning#:~:text=Clean)[owasp.org](https://owasp.org/www-project-top-10-for-large-language-model-applications/#:~:text=OWASP%20Top%2010%20for%20Large,1), cybersecurity vendor analyses[trendmicro.com](https://www.trendmicro.com/en_us/research/25/g/ai-cyber-risks.html#:~:text=However%2C%20as%20optimistic%20as%20IT,5%20years)[tenable.com](https://www.tenable.com/blog/cybersecurity-snapshot-ai-security-framework-ai-security-controls-07-11-2025#:~:text=The%20%E2%80%9CArtificial%20Intelligence%20Controls%20Matrix%E2%80%9D,1), and policy frameworks from NIST, OWASP, and ISO[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=,Map%2C%20measure%2C%20manage%2C%20and%20govern)[aws.amazon.com](https://aws.amazon.com/blogs/security/ai-lifecycle-risk-management-iso-iec-420012023-for-ai-governance/#:~:text=ISO%2FIEC%2042001%2C%20the%20international%20management,technique%20to%20meet%20those%20expectations). The EU AI Act and U.S. Executive Orders are cited directly[europarl.europa.eu](https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=Unacceptable%20risk)[federalregister.gov](https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence#:~:text=4,Commerce%20may%20deem%20appropriate%2C%20shall) to reflect current regulatory details.