# Ensuring security in the AI and machine learning revolution



AI and machine learning (ML) technologies are at the forefront of a significant revolution across various industries, reshaping business operations and offering capabilities that were previously considered unattainable. Applications such as fraud detection in financial services and diagnostic imaging in healthcare exemplify the profound impacts of these technologies. The evolution of AI/ML, however, brings to light new security challenges that organisations must address as they integrate these systems into their operations.

As highlighted by Diana Kelley, chief information security officer at Protect AI, the rapid adoption of AI technologies introduces a range of novel threats, particularly that of ML model tampering, data leakage, adversarial prompt injection, and AI supply chain attacks. Traditional software security methods are often ill-equipped to counter these emerging risks. To mitigate these vulnerabilities, Kelley advocates for the implementation of Machine Learning Security Operations (MLSecOps), which provides a comprehensive framework designed to secure the AI/ML lifecycle.

The terminology within this field often blurs the lines between artificial intelligence and machine learning, where AI refers to systems that simulate human intelligence and ML, a subset of AI, allows systems to learn independently from data. For instance, AI technologies are employed to monitor transactions for fraudulent activity, while ML models adapt over time to identify new investment patterns. However, any compromise of data inputs jeopardises the reliability of these systems.

The authors of the commentary distinguish between MLOps and DevOps, note that while both practices focus on deployment and maintenance of models, MLOps contends with the fluidities of ML models which are frequently retrained and subject to changes in data that may inadvertently introduce security vulnerabilities. This is contrasted with DevOps, which traditionally addresses static software applications and embeds security through the DevSecOps paradigm throughout the software development lifecycle.

The emergent MLSecOps framework is proposed as an analogous evolution for machine learning, seeking to ensure security is integrated at each stage of the AI/ML process—from data collection and model training to deployment and ongoing monitoring. As digital attacks evolve, the importance of protecting AI systems gains momentum.

Several specific security threats are identified that pertain directly to AI/ML; model serialization attacks involve the manipulation of an ML model during the data compression phase, while data leakage presents significant risks if sensitive information finds its way into public domains. Moreover, adversarial attacks may deceive Generative AI systems into producing erroneous or harmful outputs. Additional danger lies within AI supply chain attacks, which can compromise the foundational data or assets of an ML model before it is operational.

The MLSecOps framework aims to counteract these threats by securing data handling protocols, scanning models for anomalies, and monitoring system behaviours post-deployment. Additionally, collaboration across security teams, ML practitioners, and operational staff is emphasised to create a holistic approach to risk management within these pipelines.

Transitioning to an MLSecOps structure necessitates not only the adoption of new tools but also a cultural and operational realignment within organisations. Chief Information Security Officers (CISOs) are encouraged to foster cooperative environments among security, IT, and ML teams, which are frequently siloed in their operations. Initiatives such as conducting regular AI/ML security audits and establishing robust security controls aligned with MLSecOps principles are recommended first steps. Furthermore, ongoing training and awareness initiatives are critical to sustaining an effective MLSecOps culture as threats continue to evolve.

As AI technologies become increasingly integral to business strategies, the need for robust security practices throughout their lifecycle is paramount. MLSecOps emerges not just as a framework but as an essential progression in securing AI applications against a backdrop of ever-evolving threats, ensuring operational resilience and high performance for organisations adopting these transformative technologies.

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://perception-point.io/guides/ai-security/top-6-ai-security-risks-and-how-to-defend-your-organization/> - This article discusses various AI security risks, including AI-powered cyberattacks, adversarial attacks, data manipulation and data poisoning, model theft, model supply chain attacks, and surveillance and privacy issues, which are all relevant to the security challenges mentioned.
2. <https://www.ey.com/en_us/insights/cybersecurity/ai-and-ml-are-cybersecurity-problems-and-solutions> - This source highlights the increasing use of AI in cyber attacks, such as adversarial ML attacks, data contamination, phishing, data exfiltration, ransomware, and denial-of-service attacks, which aligns with the novel threats introduced by AI technologies.
3. <https://www.helpnetsecurity.com/2024/03/07/ai-security-challenges/> - This article details major AI security challenges, including adversarial machine learning attacks, generative AI system attacks, and supply chain attacks, which are consistent with the security threats identified in the commentary.
4. <https://perception-point.io/guides/ai-security/top-6-ai-security-risks-and-how-to-defend-your-organization/> - This source explains the importance of implementing data handling and validation, zero-trust architecture, and continuous monitoring to secure AI systems, which is in line with the MLSecOps framework proposed.
5. <https://www.ey.com/en_us/insights/cybersecurity/ai-and-ml-are-cybersecurity-problems-and-solutions> - The article emphasizes the need for integrated security measures throughout the AI/ML lifecycle, similar to the MLSecOps framework, to counteract evolving digital attacks.
6. <https://www.helpnetsecurity.com/2024/03/07/ai-security-challenges/> - This source discusses the distinction between traditional software security and the unique challenges posed by AI/ML, highlighting the need for specialized security practices like MLSecOps.
7. <https://perception-point.io/guides/ai-security/top-6-ai-security-risks-and-how-to-defend-your-organization/> - The article explains the differences between MLOps and DevOps, and how MLOps addresses the fluidities and security vulnerabilities of ML models, which is a key aspect of the MLSecOps framework.
8. <https://www.ey.com/en_us/insights/cybersecurity/ai-and-ml-are-cybersecurity-problems-and-solutions> - This source underscores the importance of collaboration across security teams, ML practitioners, and operational staff to create a holistic approach to risk management, aligning with the MLSecOps recommendations.
9. <https://www.helpnetsecurity.com/2024/03/07/ai-security-challenges/> - The article highlights the need for regular AI/ML security audits and robust security controls, which are essential steps in transitioning to an MLSecOps structure.
10. <https://perception-point.io/guides/ai-security/top-6-ai-security-risks-and-how-to-defend-your-organization/> - This source emphasizes the cultural and operational realignment needed within organizations to adopt MLSecOps, including fostering cooperative environments among security, IT, and ML teams.
11. <https://www.ey.com/en_us/insights/cybersecurity/ai-and-ml-are-cybersecurity-problems-and-solutions> - The article stresses the importance of ongoing training and awareness initiatives to sustain an effective MLSecOps culture as threats continue to evolve.
12. <https://news.google.com/rss/articles/CBMigwFBVV95cUxQYnNzX2FZUFJNYUV3X2hVSUF0VDFmSGk2dzd6WmhNVzNUQnhyVzM1MVV3ZG5RQy1NVTRFamdaQ1hWMHhRZ2w4QzVGdkQwZC1YbkVIbk5fNWxHczFsQjgtdXY1ZGxGZjZkTUI0ZUw2RjhIUWlXTnhKVkhrYnRyaG5IUlV3NA?oc=5&hl=en-US&gl=US&ceid=US:en> - Please view link - unable to able to access data