Georgian Scientists
ქართველი მეცნიერები
Vol. 7 Issue 3, 2025
https://doi.org/10.52340/gs.2025.07.03.01

The Use of Artificial Intelligence in CI/CD Systems: Enhancing Security and Managing Risks

Iveri Jajanidze[1], Ioseb Kartvelishvili[2]

[1]PhD Student, Georgian Technical University, jajanidze.iveri24@gtu.ge, Phone: 599072599;
[2]Professor, Georgian Technical University, s.kartvelishvili@gtu.ge, Phone: 595220922. ORCID: 0000-0001-9890-4099

Abstract

Organizations in the public and private sectors increasingly implement Continuous Integration (CI) and Continuous Deployment (CD) processes to streamline software delivery. These processes, supported by automation and DevOps practices, enable frequent code releases. However, the rise of adversarial threats—such as AI-enhanced social engineering, adversarial machine learning, and large-scale data leaks—introduces complex security challenges.

This article explores the integration of Artificial Intelligence (AI) and Machine Learning (ML) into CI/CD pipelines, highlighting how these technologies enhance security posture through real-time threat detection, automated incident response, and predictive analytics. Key areas include Explainable AI (XAI), Federated Learning (FL), adversarial attack prevention, and AI-powered malware detection. The study further discusses visualization techniques for security monitoring (ROC curves, histograms, pie charts), supported by 2024 statistical insights into AI adoption in CI/CD environments.

Keywords: AI in DevSecOps pipelines, Explainable AI (XAI) in cybersecurity, Federated Learning for secure model training, Adversarial attacks on machine learning models,AI-driven social engineering threats, Security visualization in CI/CD systems

- **CI/CD (Continuous Integration/Continuous Deployment)**: Software engineering practices enabling regular and reliable delivery of code changes into production.
- **Artificial Intelligence (AI)**: Computer systems simulating human intelligence to perform tasks like reasoning, learning, and problem-solving.

- **Machine Learning (ML)**: A subset of AI where systems learn patterns from data and improve over time without being explicitly programmed.
- **Explainable AI (XAI)**: A branch of AI focused on making model decisions interpretable and transparent to humans.
- **Federated Learning**: A decentralized ML approach where training data remains on local devices while only model updates are shared centrally.
- **Adversarial Attacks**: Malicious techniques targeting AI/ML models to produce incorrect outputs or degrade performance.
- **Social Engineering Attacks:** Psychological manipulation techniques used by attackers to trick individuals into revealing confidential information, granting system access, or performing unsafe actions—often through phishing, pretexting, baiting, or impersonation. These attacks exploit human trust rather than technical flaws.

Introduction

CI/CD methodologies have become foundational pillars in modern software development. CI refers to the continuous merging of code changes into a central repository, followed by automated builds and tests. CD ensures that these changes are automatically deployed to production. This paradigm, driven by DevOps (Development and Operations) culture, drastically reduces the time between development and delivery.

However, the speed and automation associated with CI/CD pipelines reduce the effectiveness of traditional security controls, which were designed for slower, manual release cycles. This creates a need for security mechanisms that are equally fast and adaptive.

Integrating **AI and ML** into CI/CD security enables:

- **Real-time threat detection** through anomaly detection in logs, metrics, and APIs;
- **Automated incident response**, such as triggering alerts, opening SIEM tickets, or applying patches;
- **Risk prediction** via behavioral analysis of source code metadata and developer interactions.

CI/CD tools like Jenkins, GitLab CI, and Azure DevOps increasingly integrate with AI frameworks such as TensorFlow Extended and Kubeflow, forming the foundation of intelligent DevSecOps pipelines.

Explainable AI (XAI) and Regulatory Compliance

In highly regulated sectors, such as finance and healthcare, AI-based decisions must be auditable and justifiable. This is where **Explainable AI (XAI)** comes into play—offering insight into the logic and feature contributions behind model predictions.

- **LIME (Local Interpretable Model-agnostic Explanations)** explains individual predictions by training a surrogate model locally around the instance being evaluated, highlighting which features were most influential.
- **SHAP (SHapley Additive exPlanations)** uses concepts from cooperative game theory to assign importance scores to each input feature, ensuring consistency and fairness in explanations.

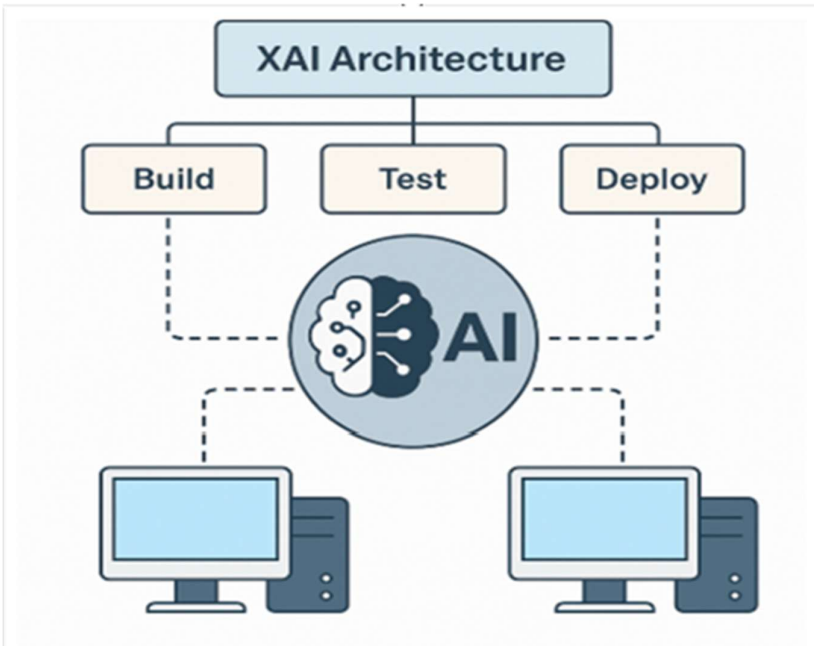
These tools facilitate:

- Compliance with regulations like **GDPR**, **DORA**, and **NIST SP 800-53**, which demand accountability in automated decision-making;
- Improved communication between developers, auditors, and security teams;
- Greater trust in AI-powered CI/CD workflows.

Federated Learning and Data Confidentiality

Federated Learning (FL) allows AI models to be trained across multiple decentralized devices or servers without transferring raw data. This privacy-preserving technique is particularly important in sensitive environments such as banking or healthcare.

Fig. 1: Federated Learning in CI/CD Architecture



Key benefits:

- **Data locality**: Private data never leaves the host environment (e.g., endpoints or internal infrastructure).
- **Collaborative learning**: Different organizations or departments contribute to a shared model, enhancing its generalization capabilities.
- **Secure aggregation**: Only model updates—such as gradients or parameters—are shared, not the data itself.

FL reduces the risk of data leakage while preserving model performance and learning efficiency.

Adversarial Attacks and AI-Driven Malware

As AI becomes more embedded in CI/CD, so too do the threats targeting it. **Adversarial attacks** seek to exploit ML models by manipulating their inputs to cause incorrect predictions.

Common techniques include:

- **Adversarial Examples**: Slight perturbations imperceptible to humans are introduced to input data, fooling the AI into making wrong predictions.
- **Poisoning Attacks**: Malicious actors insert corrupted samples into training datasets to bias model learning.
- **Evasion Attacks**: Attackers craft inputs that evade detection at runtime, particularly during the inference phase.

AI-powered malware now employs **Reinforcement Learning (RL)** to dynamically adapt and optimize attack paths. These threats can:

- Automatically craft payloads using deep learning;
- Adjust behavior to avoid detection by IDS/IPS systems;
- Obfuscate code using AI-guided encryption techniques.

AI-Augmented Social Engineering

Social engineering attacks, such as phishing and pretexting, become more convincing and scalable with AI. By analyzing linguistic and behavioral cues, AI can automate and personalize malicious messages.

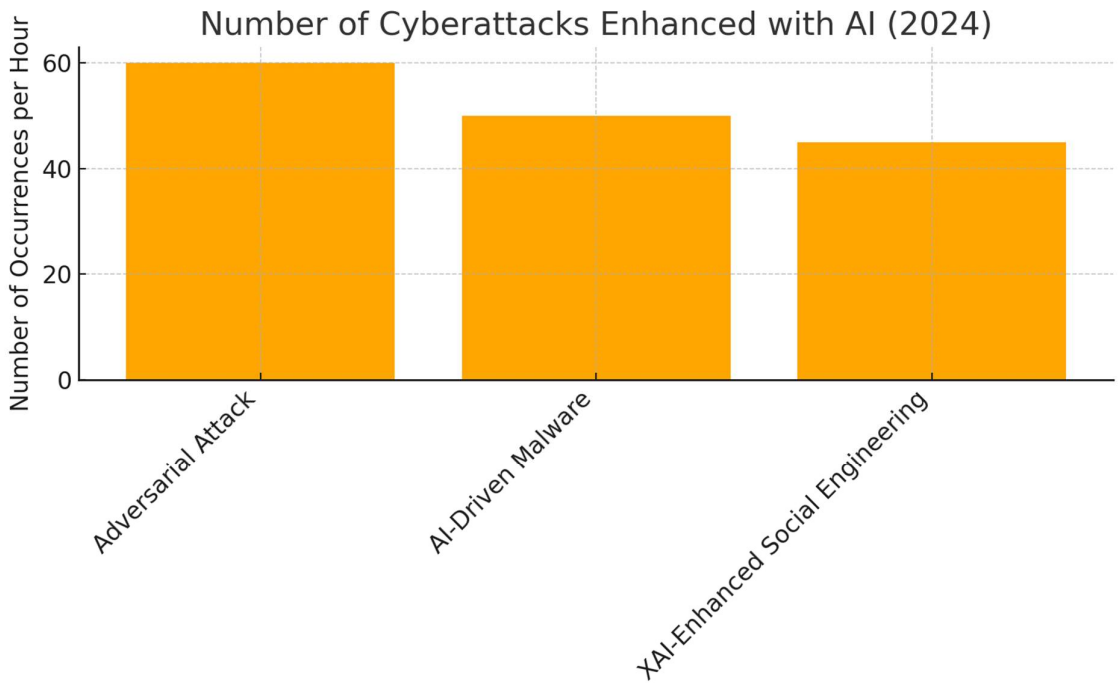
Key features include:

- **Sentiment analysis**: Models detect user tone and generate targeted phishing content;

- **Contextual adaptation**: Messages are translated and localized for industry, language, and demographic;
- **Automated persona creation**: Bot networks create fake social media profiles to spread misinformation or engage victims.

Such capabilities make detection more difficult, necessitating dynamic security controls, such as AI-driven behavioral baselining and MFA enforcement.

Fig. 2: AI-Driven Attacks



Security Monitoring and Visualization

Security visualization tools transform complex datasets into intuitive graphics that facilitate quicker response and insight.

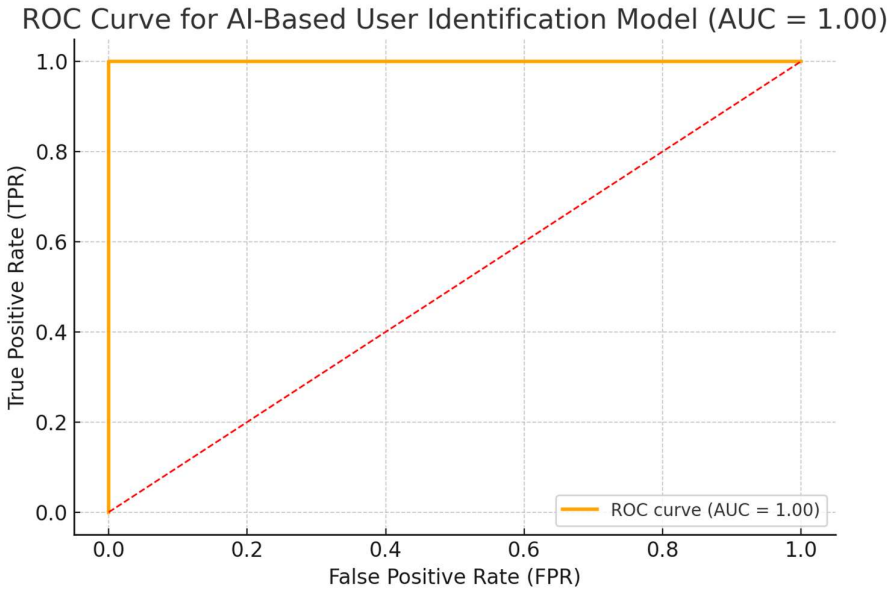
- **ROC Curves (Receiver Operating Characteristic)**: Plot the trade-off between true positive and false positive rates, with **AUC (Area Under the Curve)** as a metric of detection quality.
- **Histograms & Density Plots**: Highlight deviations and anomalies in system behavior.
- **Pie Charts & Category Summaries**: Visualize proportions of threat types or affected assets.

Such dashboards empower DevSecOps teams to:

- Prioritize alerts based on severity;
- Track security metrics over time;

- Justify decisions to stakeholders or compliance officers.

Fig. 3: ROC Curve for Security Monitoring in CI/CD Pipelines



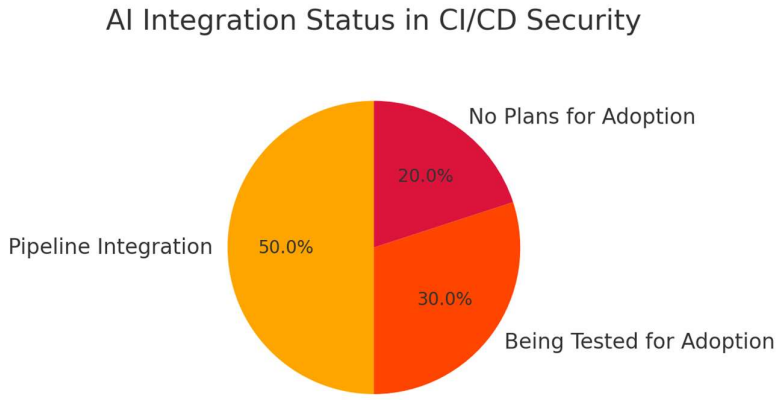
CI/CD AI Security Adoption Statistics

A 2024 industry study surveying over 200 organizations revealed:

- **50%** actively use AI in their CI/CD security processes;
- **30%** are in the testing and pilot stages;
- **20%** plan to adopt AI solutions within the next 12 months.

These figures reflect a growing recognition of AI's role in ensuring agile yet secure software delivery.

Fig. 4: AI Adoption Rates in CI/CD Security (2024)



Conclusion

Integrating AI and ML into CI/CD pipelines represents a paradigm shift in software delivery security. Techniques such as XAI, Federated Learning, and adversarial defense mechanisms enhance visibility, trust, and resilience. Moreover, the ability to detect, visualize, and respond to threats in real-time enables DevSecOps teams to meet modern demands without sacrificing velocity or compliance.

By embracing these technologies holistically, organizations position themselves to address both current and emerging cybersecurity challenges.

References

Cath, C. და სხვ. (2018). *Artificial intelligence and the 'good society': the US, EU, and UK approach.* Science and Engineering Ethics, 24(2), 505–528. Gunning, D. და სხვ. (2019). *XAI—Explainable artificial intelligence.* Science Robotics, 4(37).

Ucci, D. და სხვ. (2019). *Survey of machine learning techniques for malware analysis.* Computers & Security, 81, 123–147.

Yang, H. და სხვ. (2019). *Federated machine learning: Concept and applications.* ACM Transactions on Intelligent Systems and Technology, 10(2), 1–19.

ქართველიშვილი ი., კუჭავა გ. CI/CD-ის დახმარებით DevOps-ში პროგრამული უზურნველყოფის მიწოდების ხარისხისა და სიჩქარის ოპტიმიზაცია. საქართველოს ტექნიკური უნივერსიტეტის საერთაშორისო სამეცნიერო-პრაქტიკული კონფერენცია "თანამედროვე გამოწვევები და მიღწევები ინფორმაციულ და საკომუნიკაციო ტექნოლოგიებში - 2024“, საქართველო, თბილისი, 1-2 ნოემბერი, 2024 წელი.

Avtandil Bichnigauri, Ioseb Kartvelishvili, Otar Shonia, Daviti Bichnigauri, Otar Gudadze. Strengthening Cyber Defenses - The Crucial Role of Phishing Simulation in Modern Security Strategies. International Scientific Journal "Defence and Science". No. 3 (2024). https://doi.org/10.61446/ds.3.2024.8467

ქართველიშვილი ი., ოხანაშვილი მ., ჩორხაული ნ. ქსელური შეტევების აღმოჩენის არსებული მეთოდების მიმოხილვა და ანალიზი. საქართველოს ტექნიკური უნივერსიტეტის საერთაშორისო სამეცნიერო-პრაქტიკული კონფერენცია "თანამედროვე გამოწვევები და მიღწევები ინფორმაციულ ტექნოლოგიებში - 2023“, საქართველო, თბილისი, 12-13 ოქტომბერი, 2023 წელი.