White Paper

Introduction to the NetApp AI Security Framework

Justin Spears, NetApp

May 2024 | WP-7365

Abstract

AI systems are attractive targets for malicious actors. This white paper introduces the NetApp® AI Security Framework and describes its role in providing proper security across the AI pipeline.

TABLE OF CONTENTS

NAISF overview 3

Secure your data 3

Enhance your ecosystem security 4

Secure your model 4

Validate your data 4

Monitor and review your environment 5

AI governance 5

Conclusion 6

About the author 6

LIST OF FIGURES

Figure 1) Six key tenets of the NetApp AI Security Framework. 3

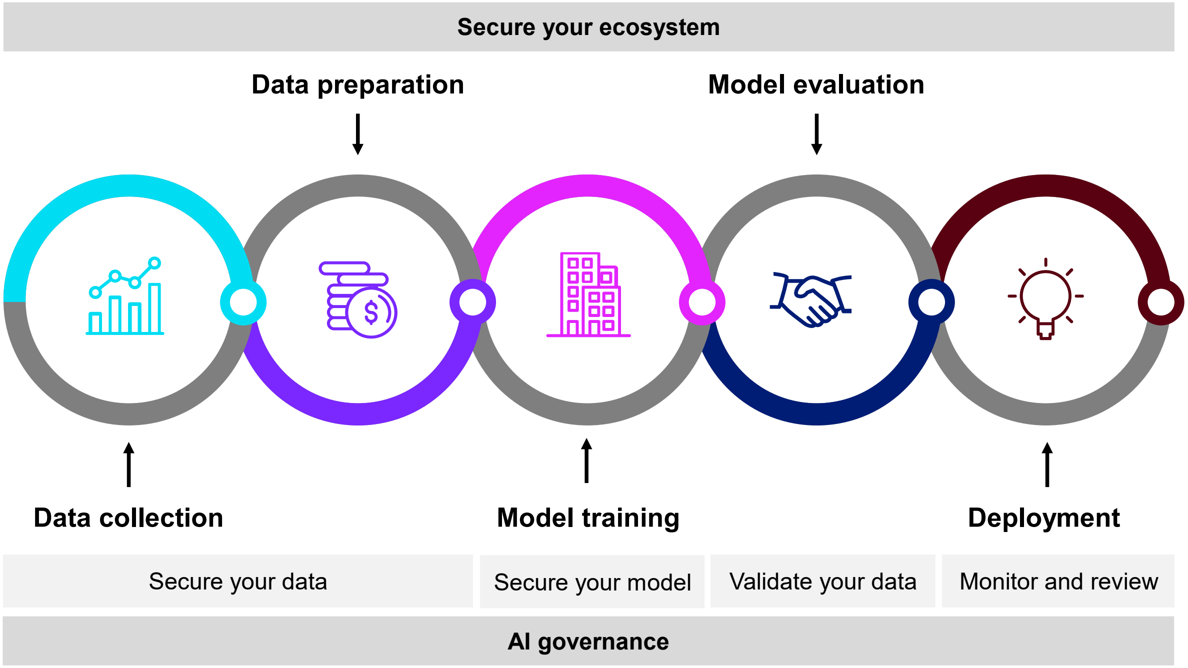
The potential of AI, especially generative AI, to change the fundamental landscape of our world is massive. But as AI systems become more prevalent and sophisticated, they also become more attractive targets for malicious actors. Adversarial attacks, data breaches, and AI-specific attacks like model poisoning or evasion can have severe consequences. Furthermore, AI models and algorithms can have vulnerabilities that attackers can exploit. These vulnerabilities can arise from biased or incomplete training data, algorithmic weaknesses, or design flaws.

A security framework helps organizations stay ahead of emerging threats, implement strong security controls, and continuously monitor and protect AI systems. Such a framework can ultimately help businesses navigate the complexities of AI security and safeguard their AI systems and the data they handle. This white paper introduces the NetApp® AI Security Framework (NAISF).

NAISF overview

The NAISF is built on six key tenets that constitute a holistic approach to providing proper security across the AI pipeline. These tenets include securing your data, ensuring proper ecosystem security, securing your training models, validating your data, monitoring your environment, and implementing AI governance (Figure 1).

Figure 1) Six key tenets of the NetApp AI Security Framework.



Secure your data

Data security plays a critical role in any AI framework, because it ensures the confidentiality, integrity, and availability of the data throughout the AI lifecycle. By incorporating comprehensive data security practices, organizations can safeguard their AI assets and maintain the integrity and ethical standards of their AI systems. The two key elements of data security include:

- **Access control.** Implement robust access controls to regulate who can access, modify, or delete data. This includes user authentication, authorization mechanisms, role-based access controls (RBAC), attribute-based access controls (ABAC), and Zero Trust models to ensure that only authorized personnel can access sensitive data.

- **Encryption.** Use encryption techniques to protect data at rest (stored data), data in transit (data being transmitted over networks), and data in use. A variety of encryption algorithms can be employed to secure data and keep it confidential even if it falls into unauthorized hands.

Enhance your ecosystem security

Ecosystem security plays a crucial role in a comprehensive AI security framework. Securing the entire AI ecosystem involves not only the AI models and data but also the supporting infrastructure, devices, third-party integrations, and the overall environment in which the AI system operates. A holistic approach to ecosystem security enhances the overall resilience, reliability, and trustworthiness of AI systems in real-world environments. Some of the key elements of ecosystem security include:

- **Secure development lifecycle.** Integrate security into every phase of the AI development lifecycle, from design and development to testing and deployment, by adopting secure coding practices and conducting regular security assessments.
- **Consistency in controls.** Support a consistent set of controls across the IT ecosystem for secure-by-default protection, so that no individual product becomes the weakest link.
- **Vendor and third-party security.** Evaluate the security practices of vendors and third-party providers who have access to your data. Ensure that they have appropriate security controls and contractual obligations in place to protect the data they handle on your behalf.
- **Employee training and awareness.** Educate employees about data security best practices, the importance of handling data securely, and the risks associated with data breaches or mishandling of data. Foster a culture of data security to help employees be vigilant and proactive in protecting data.

Secure your model

Securing AI training models is vital to protecting the integrity, confidentiality, and availability of the models during the training process. By adopting several mechanisms, organizations can enhance the security posture of their AI training models and reduce the risk of unauthorized access, data breaches, code theft, or tampering throughout the model development lifecycle. Some of the key elements of securing your model include:

- **Model security.** Apply techniques like model encryption, differential privacy, and federated learning to protect sensitive models and data used for training.
- **Model versioning and tracking.** Use a version control system or repository to track changes to the AI models. This helps maintain a historical record and enables easy rollback or analysis if security issues are detected later. It is also important to implement rigorous model verification and validation practices to maintain the integrity and correctness of the trained models.
- **Secure training environment.** Implement secure configurations, regular security updates, and isolation measures for the AI training infrastructure to minimize the risk of unauthorized access or compromise.

Validate your data

Data validation plays a crucial role in the AI pipeline, providing several valuable benefits throughout the data preprocessing and model training stages. Integrating robust data validation practices contributes to the development of reliable, fair, and trustworthy AI systems. A few key elements of data validation include:

- **Regulatory compliance.** Organizations must stay abreast of relevant regulations such as the General Data Protection Regulation (GDPR), California Consumer Privacy Act (CCPA), and Health

Insurance Portability and Accountability Act (HIPAA), and ensure that AI systems comply with legal and regulatory requirements regarding data privacy and security.

- **Ensuring model safety and reliability.** Validating the data used during model development helps models operate safely and reliably. By validating inputs, monitoring outputs, and verifying that the models are performing within expected bounds, organizations can mitigate risks associated with erroneous or unsafe model behavior.
- **Mitigating data leakage risks.** Data validation helps identify unintended sensitive information or personally identifiable information (PII) and remove it from the training data. The data used in the pipeline must be properly anonymized or de-identified so that organizations can reduce the risk of data leakage and can maintain privacy and comply with data protection regulations.

Monitor and review your environment

Monitoring is a crucial component of a security framework for AI, because it provides continuous visibility into the behavior, performance, and security of AI systems. Monitoring empowers organizations to maintain a strong, adaptive security posture for their AI systems, maintaining the integrity, availability, and trustworthiness of the AI assets and the underlying infrastructure. There are a few key areas of monitoring, such as:

- **Threat detection and response.** Use AI-powered threat detection tools to identify and mitigate security risks in real time and establish incident response protocols to address security incidents promptly.
- **Monitoring and logging.** Implement mechanisms to track model performance, detect anomalies, and monitor for suspicious activities. This allows timely detection and response to potential security incidents.
- **Regular auditing.** Conduct regular security audits of the training process to identify and address vulnerabilities or potential attack vectors. This includes reviewing data handling practices, model architecture, and training methodologies.

AI governance

In the context of AI security, governance plays a crucial role in ensuring responsible and ethical use of AI systems. It involves establishing policies, guidelines, and frameworks that govern the development, deployment, and operation of AI models and applications. Overall, governance in an AI security framework makes sure that AI systems are developed and used responsibly, with appropriate measures in place to address risks, adhere to ethical principles, and comply with laws and regulations. It provides a structured approach to managing the lifecycle of AI systems and minimizes potential harm while maximizing benefits. Here's a detailed description of the core elements of governance in an AI security framework:

- **Definition of ethical principles.** Governance establishes a set of ethical principles and guidelines for AI development and deployment. These principles should encompass fairness, transparency, accountability, privacy, and inclusivity. They ensure that AI systems align with societal values and do not hurt individuals or communities.
- **Risk assessment.** Governance involves conducting thorough risk assessments to identify potential risks and threats associated with AI systems. This includes assessing the potential biases, vulnerabilities, and security risks of AI models, datasets, and deployment environments. Risk assessment helps in identifying potential issues and mitigating them before they cause harm.
- **Accountability and transparency.** Governance emphasizes accountability and transparency for AI systems. It defines processes for documenting and tracking AI system behavior, decisions, and outcomes. This includes keeping records of model versions, training data, and evaluation metrics.

Transparency measures such as robust documentation and disclosure of system capabilities and limitations promote trust and enable users to make informed decisions.

- **Continuous improvement.** Governance in AI security frameworks promotes a culture of continuous improvement. It encourages organizations to learn from incidents, feedback, and emerging best practices. Regular audits, reviews, and updates to policies and procedures help in adapting to evolving threats and technologies.

Conclusion

An AI security framework delivers business value by protecting intellectual property, enhancing customer trust, maintaining regulatory compliance, minimizing financial losses, preserving brand reputation, improving operational efficiency, and enabling future innovation. Organizations can create a more secure, trustworthy, and responsible AI ecosystem by leveraging the six key tenets described in the NetApp AI Security Framework. This framework includes principles such as data security, ecosystem security, model and data validation, monitoring, and AI governance. By aligning organizations' AI initiatives with such a framework, companies ensure the integrity, availability, and confidentiality of their AI systems while generating value and mitigating risks throughout their AI journey.

About the author

Justin Spears is a consummate technical executive with 20+ years of experience in cybersecurity, product development, data science, and engineering. Currently residing at the intersection of security, AI, and data, Justin's primary focus is driving the cybersecurity roadmap for all core NetApp products. In previous roles, he was responsible for product and engineering work at Dell and EMC and spent several years in academic research at his alma mater, the University of Pittsburgh. When not trying to circumvent virtual evildoers, you can find Justin powerlifting and renovating houses with his wife and four daughters.

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

Copyright information

Trademark information