Science and Technology 2024, 14(2): 17-21
DOI: 10.5923/j.scit.20241402.01

Demystifying Evaluation Framework for Modern AI Security Tools

Anupam Mehta[1,*], Lionel Dsouza[2], Sharon Augustus[3]

[1]Product Security, Salesforce, Ashburn, USA
[2]Software Integrity Group, BlackDuck, Reston, USA
[3]Product Security, Salesforce, Boston, USA

Abstract Software development in the modern era follows a practice of rapidly developing, testing, and deploying software products or features. This approach is commonly seen in agile and DevOps environments, where the emphasis is on quick iterations, continuous integration, and continuous delivery (CI/CD). The use of generative Artificial Intelligence (AI) in high-velocity software development has become increasingly significant as organizations strive to accelerate their development processes while maintaining high quality and security standards. As such, Artificial Intelligence (AI) plays a transformative role in the field of security, enhancing the ability to protect software, systems, data, and networks from a wide range of threats.

Keywords Cybersecurity, Artificial Intelligence, Security Tool

1. Introduction

This article describes a framework for evaluating modern security tools equipped with capabilities in the Artificial Intelligence (AI) space. Security products/tools have been significantly on the rise of development, at various stages of the Software Development Lifecycle (SDLC) such as AI-driven threatmodeling and secure design, AI-driven threat detection, AI for Incident Response, AI-Powered Security Information and Event Management (SIEM), AI in Vulnerability Management and so on.

Security tools are not one-size-fits-all solutions; they must be chosen based on the specific needs, architecture, and threat environment of the organization. By applying well-defined evaluation criteria, organizations can ensure that their security investments are not only effective but also aligned with their broader strategic goals. This approach helps in selecting tools that provide robust protection, adapt to emerging threats, integrate seamlessly with existing infrastructure, and ultimately contribute to a stronger overall security posture.

On a high level, the key factors for evaluating a security product or service empowered with AI are as follows:

- Threat Identification and False Positive
- Scalability and Deployment
- Ease of Integration
- Customization, Usability, and User-Friendly Interface
- Real-time alerts and Notifications
- Compliance and Reporting
- Performance and Availability
- Proactive Defense and Incident Response
- Adaptability and Learning
- Support and Documentation
- Cost-effectiveness and Licensing
- Security and Privacy by Design

The following sections provide a detailed description of key evaluating indicators for any security product or service equipped with AI capabilities.

2. Evolution of Modern Security Tools

In the modern era, security tools have evolved to address increasingly sophisticated cyber threats, leveraging advancements in technology like artificial intelligence (AI), machine learning (ML), and cloud computing. These tools are designed to provide comprehensive protection across all layers of an organization's IT infrastructure, including endpoints, networks, applications, data, and users.

Modern security tools are typically integrated into broader security architectures, enabling seamless interaction between various components to provide real-time threat detection, prevention, response, and remediation. They are often cloud-based or hybrid, offering scalability, flexibility, and the ability to leverage global threat intelligence. These tools are crucial in protecting against a wide range of threats, including malware, ransomware, phishing, insider threats, and advanced persistent threats (APTs). These modular,

* Corresponding author:
anupammeht@gmail.com (Anupam Mehta)
Received: Aug. 22, 2024; Accepted: Sep. 16, 2024; Published: Sep. 21, 2024
Published online at http://journal.sapub.org/scit

layered architectures ensure that modern security tools are flexible, scalable, and capable of providing comprehensive protection against the wide range of threats that organizations face today. By integrating AI and machine learning, these tools are also increasingly proactive, adaptive, and efficient in identifying and mitigating security risks.

3. Common Applications of Security Tools and Artificial Intelligence

As the AI advancements in security have exponentially increased, numerous security tools have adopted approaches that contextualize data in relation to associated risks or threats. This shift has led to several use cases that significantly enhance the efficiency of risk discovery. Some of the common use cases include:

AI-driven Threat Modeling - This involves processing architecture diagrams, data flow diagrams (DFDs), and application-specific documentation to generate business-driven and application specific threats. This includes a comprehensive list of potential attacks, attack goals, threat actors, and the overall attack surface. The output can be manually reviewed by a penetration tester or fed into AI-driven Dynamic Application Security Testing (DAST) tools. Additionally, this tool can be fine-tuned to perform a gap analysis of security controls, identifying both existing controls and those that are missing. This allows organizations to invest in missing controls more effectively.

AI-Enhanced Source Code Review - This involves leveraging machine learning to analyze code and identify security vulnerabilities. These tools detect issues and offer multiple remediation options with near-accurate syntax suggestions that developers can directly implement, avoiding the reliance on generic remediation guidance. These tools are capable of detecting code changes, detecting issues with minimal to no false positives, and highlighting issues that might need a manual review. AI-powered tools should be able to prioritize vulnerabilities based on the context of the application.

AI-based Threat Detection - AI-powered detection tools use advanced machine learning models to continuously enhance threat detection capabilities. These tools are designed to process and analyze vast amounts of data at scale, tailored to the specific needs of an organization. They excel at identifying attack patterns based on both behavioral and signature-based analysis, achieving detection speeds that far surpass human capabilities. Moreover, AI-powered threat detection tools are particularly effective in categorizing threats while significantly minimizing false positive rates.

4. Criteria for Evaluation of Product

In today's highly competitive market, the success of a product hinges not only on its ability to meet consumer needs but also on how well it stands up to critical evaluation criteria. The process of evaluating a product involves assessing various attributes that contribute to its overall value, performance, and suitability for its intended purpose. Security tools often go under strict reviews and evaluations with the nature of the data they handle, process, or generate for the organizations.

This evaluation serves multiple stakeholders engineering org, security engineering teams, vulnerability management teams, and more, each with their own set of expectations and requirements. For security engineering, the evaluation criteria might focus on usability, reliability, cost-effectiveness, true positive detection, integration, etc. Engineering, on the other hand, may prioritize factors such as integration efficiency, scalability, customization, performance, etc.

This paper aims to define and explore the key criteria to consider when evaluating a security product [1] [3] equipped with Artificial Intelligence, offering a structured approach that can be applied across various organizations and industries. By understanding these criteria, organizations can better assess their products, make informed decisions about product development, and ultimately integrate the right security tools through their development lifecycle.

4.1. Threat Identification and False Positive

- ***Identification of Threats:*** A security tool should be able to identify threats in a given stage during the SDLC. If required it must be able to apply advanced methods like machine learning, behavioral analysis, and anomaly detection to identify sophisticated and emerging threats.
- ***Business Context for Risk Identification:*** The AI system understands the specific environment it is operating in, including the typical behavior of users, systems, applications, and data flows, it is also aware of the system architecture, including how different components interact, what data is most sensitive, and which assets are most critical. This allows the AI to prioritize risks and threats that could have the most significant impact. For example, what is considered normal activity in a financial institution may differ from that in a healthcare setting. This may require training the model with the organization's context and data.
- ***Fine Tuning Policies or Rulesets:*** The tool must allow capabilities to fine-tune and configure rules/policies to allow lowering the overall false positive rate.
- ***Low False Positives:*** The tool should be accurate, with a low rate of false positives, to ensure that security teams are not overwhelmed with unnecessary alerts.

4.2. Scalability and Deployment

- ***Adaptability to Growth:*** The security tool should be scalable, meaning it can handle an increasing amount of contextual data to increase predictions, without compromising performance or protection.
- ***Support for Diverse Environments:*** It should work effectively in different environments, whether

on-premises, in the cloud, or hybrid settings.

- ***Deployment Solutions:*** The tool should have multiple deployment models and connectivity patterns to LLMs and datasets. Organizations have various security policies on data leaving the trust boundary to a third-party environment with hosted LLMs.

4.3. Ease of Integration

- ***Compatibility with Existing Systems:*** The tool should integrate seamlessly with other security solutions and the existing infrastructure. This allows an organization to easily adapt and plug the tools in the pipeline.
- ***APIs and Plugins:*** It should offer APIs or plugins that facilitate easy integration with other tools and systems. Allowing APIs exposed from the deployed environment makes it easy to centralize the notifications /threats identified by the tool.

4.4. Customization, Usability, and User-Friendly Interface

- ***Intuitive Interface:*** The tool should have an easy-to-navigate interface that allows security practitioners to quickly understand the identified threats, and their mitigating actions and allow developers to provide feedback on the threat.
- ***Custom Dashboards:*** Security personas such as Security leadership, security engineers and security product managers should be able to customize dashboards and reports to focus on the metrics and alerts most relevant to them.
- ***Automation:*** The tool should offer automation for routine tasks, reducing the manual effort required by security teams. Some of the common tasks are automated reports aggregated over a period of time, notification to teams, configurable risk prioritization, etc.
- ***Custom Risk Prioritization and Exception:*** The tool should allow the team to file/seek exceptions through the dashboard. It should also allow organizations to be able to configure risk prioritization based on their policy and standards.

4.5. Real-Time Alerts and Notifications

- ***Timely Alerts:*** The tool should provide real-time alerts and notifications when a threat is identified, allowing for quick responses from security teams and engineering teams. AI can enhance this process by using machine learning algorithms to analyze normal behavior patterns and trigger alerts when deviations occur. AI-driven tools can also prioritize alerts based on contextual risk factors, automatically filtering out low-risk activities while highlighting high-risk events that require immediate action.
- ***Configurable Alerting:*** Users should be able to configure alert thresholds and types to reduce noise and focus on critical threats. AI can dynamically adjust alert thresholds based on historical patterns, reducing the need for constant manual tuning. AI tools can also learn from user feedback to improve accuracy, gradually filtering out false positives and enhancing the relevance of alerts over time.
- ***Advanced Blocking Features:*** Security tools should offer the capabilities to allow products from being blocked to release/deployment for customers if critical /high-risk threats are identified during the product evaluations.

4.6. Compliance and Reporting

- ***Regulatory Compliance:*** The tool should help ensure compliance with relevant regulations and standards, such as GDPR, HIPAA, PCI-DSS, and others. In an AI-driven security tool, models should be trained to recognize compliance violations and suggest remediation steps. For example, the tool could automatically flag non-compliant data storage or detect vulnerabilities that might lead to data breaches and non-compliance.
- ***Audit Trails:*** It should maintain detailed logs and audit trails that can be used for compliance audits and forensic analysis. For example, Generative AI can assist in making sense of large volumes of audit logs by providing insights, flagging anomalies, and recommending corrective actions.
- ***Reporting Capabilities:*** The tool should generate detailed reports that can be customized for different audiences, such as executives, auditors, or technical teams.

4.7. Performance and Availability

- ***Minimal Impact on System Performance:*** The tool should be designed to operate efficiently without significantly slowing down the systems it protects.
- ***High Availability:*** It should offer high availability and redundancy features to ensure continuous protection, even during system failures or maintenance.

4.8. Adaptability and Learning

- ***Interoperability:*** Ensure that the tool can seamlessly work with existing security infrastructure, enabling centralized monitoring and management.
- ***Machine Learning and AI:*** The tool should leverage AI and machine learning to adapt to new threats and continuously improve its detection and prevention capabilities.
- ***Behavioral Analysis:*** It should be able to learn and recognize normal behavior patterns within the network and identify deviations that may indicate a security breach.

4.9. Support and Documentation

- ***Vendor Support:*** The tool should come with strong vendor support, including regular updates, patches, and a responsive support team.
- ***Comprehensive Documentation:*** It should have

thorough documentation that guides installation, configuration, usage, and troubleshooting.

4.10. Cost-Effectiveness and Licensing

- ***Reasonable Cost:*** The tool should offer a good balance between cost and functionality, ensuring it delivers value for the investment.
- ***Flexible Licensing:*** It should provide flexible licensing options that allow organizations to scale usage based on their needs without excessive costs.
- ***Model Training and Cost:*** This process requires significant computational resources, including GPUs or TPUs, to handle the intensive operations. The training process can be time-consuming and expensive. This should be factored into the overall evaluation.

4.11. Security and Privacy by Design

- ***Data Encryption:*** Ensure that the tool encrypts data both in transit and at rest, using strong encryption algorithms to protect sensitive information.
- ***Data Minimization:*** The tool should only collect and process the minimum amount of data necessary to perform its functions, adhering to the principle of data minimization.
- ***Anonymization/Pseudonymization:*** Look for features that anonymize or pseudonymize sensitive data to protect user privacy and comply with data protection regulations like GDPR.
- ***Model Integrity:*** The tool should have mechanisms to protect the integrity of the AI model, ensuring it cannot be tampered with or corrupted by unauthorized users.
- ***Model Retraining:*** The tool should offer the ability to retrain its AI models with new data to adapt to evolving threats and changing environments.
- ***Model Network Connectivity:*** The tool should secure the networks that host and interact with generative AI models. This ensures the safe and reliable operation to interact with confidential data (source code, designs, vulnerability).
- ***Multi-Tenant Isolation***: The tool's architecture and design must implement strict data segregation and access controls to ensure that each tenant's data and interactions are isolated from others. Utilize encryption, tenant-specific authentication, and real-time monitoring to prevent unauthorized access and cross-tenant data leakage.

4.12. Proactive Defense and Incident Response

- ***Threat Intelligence Integration:*** The tool should incorporate threat intelligence feeds to stay updated on the latest threats and vulnerabilities.
- ***Automated Incident Response:*** It should have the capability to automatically respond to certain threats, such as isolating infected systems or blocking malicious IP addresses.
- ***Forensic Capabilities:*** The tool should offer features for conducting forensic analysis after a security incident, helping to understand the scope and impact of the attack.

4.13. Key Performance Indicators (KPIs) [2] for a Tool

- ***Coverage KPIs:*** The percentage of an organization's assets (e.g., servers, endpoints, applications, networks) that are protected or monitored by the security tool. Target: Ideally, this should be close to 100%, indicating that all critical assets are protected.
- ***Effectiveness KPIs:*** Detection Rate (True Positive Rate). The percentage of actual security threats that the tool successfully detects. A high detection rate (e.g., 95% or above) is desirable, indicating the tool is effectively identifying threats.
- ***Continuous Improvement KPIs (False Negative Rate):*** The percentage of actual threats that the tool fails to detect. Target: This should be as close to 0% as possible, indicating that the tool is not missing threats.

5. Challenges for Security Tools Integrating AI

While AI brings powerful capabilities to security tools, the integration of AI introduces a range of challenges, including privacy concerns, vulnerability to adversarial attacks, ethical issues, and more. Overcoming these challenges requires careful consideration of data governance, transparency, and system design, as well as constant monitoring and adaptation to evolving threats. Organizations must also weigh the costs and scalability of AI tools, ensuring they can operate effectively in diverse and complex environments.

Integrating AI into security tools offers significant advantages, such as enhanced threat detection, automated responses, and predictive analytics. However, these benefits come with challenges that must be addressed to ensure the tools are effective, secure, and ethical. Below are key challenges associated with integrating AI into security tools:

5.1. Privacy Concerns

AI security tools often require large amounts of data, including sensitive personal and organizational information, to function effectively. This creates several privacy-related challenges:

- ***Data Collection and Use:*** AI systems must collect, process, and analyze vast amounts of data, which may include personal identifiers, proprietary information, and confidential data. Ensuring that data is collected and used in compliance with privacy regulations (e.g., GDPR, HIPAA) is critical.
- ***Data Anonymization:*** It is challenging to anonymize data effectively while retaining its utility for AI models. Insufficient anonymization may lead to inadvertent exposure of sensitive information.
- ***User Consent and Transparency:*** Users and

organizations need to be informed about how their data is used by AI-driven security tools. Gaining consent, ensuring transparency, and maintaining trust can be difficult, particularly when AI models rely on continuous data collection.

5.2. Adversarial Attacks

AI models, including those in security tools, are vulnerable to adversarial attacks, where attackers manipulate inputs to deceive or exploit the AI system:

- ***Model Poisoning:*** Attackers may inject malicious data into the training dataset to influence the AI model's behavior, causing it to make errors when deployed. Poisoned data can cause the model to overlook certain types of threats or vulnerabilities.
- ***Evasion Attacks:*** In real-time systems, attackers may alter their behaviors to bypass AI-driven detection mechanisms, effectively learning how to evade the system's protections.

5.3. Ethical Issues

AI-driven security tools raise several ethical concerns, especially in how decisions are made and how data is used:

- ***Decision Bias:*** AI models are only as good as the data they are trained on. If biased or unrepresentative data is used, the model may disproportionately impact certain groups or make unfair decisions, such as flagging specific types of behaviors or users more frequently.
- ***Automated Decision-Making:*** AI security tools often make decisions autonomously, such as blocking access or flagging behavior as suspicious. This raises concerns about accountability, fairness, and the right to challenge or appeal decisions, especially if users are incorrectly classified as threats.
- ***Ethical Data Usage:*** Security tools must balance their need for large amounts of data with ethical considerations about surveillance and privacy. Using AI for security without overreaching into unnecessary or invasive data collection is a critical challenge.

5.4. Miscellaneous

AI models include some additional challenges of their own. A non-exhaustive list of challenges is defined below:

- ***Resource Intensive:*** Training and deploying AI models require significant computational resources, particularly for deep learning models. Ensuring the tool can operate efficiently at scale without impacting system performance can be difficult.
- ***Deployment in Multi-Cloud Environments:*** Many organizations operate in complex, multi-cloud environments, making it challenging to deploy and scale AI-driven security tools consistently across different platforms and infrastructures.
- ***Incomplete Data***: If key data is missing or incomplete, the AI model may fail to recognize certain threats, reducing its effectiveness.
- ***Integration with Legacy Systems:*** Many organizations still rely on legacy systems that may not easily integrate with modern AI-driven security tools:

6. Conclusions

In summary, AI enhances the ability of organizations to detect, prevent, and respond to security threats with greater speed and accuracy. It enables the automation of routine security tasks, provides real-time insights, and helps in anticipating and mitigating emerging threats, making it a critical component of modern cybersecurity strategies. Evaluating security tools is a critical process that determines their effectiveness in safeguarding an organization's assets, data, and operations against an ever-evolving landscape of threats. As we have explored in this paper, the criteria for evaluating security tools must be comprehensive, covering aspects such as functionality, performance, scalability, integration, user experience, security by design, critical KPIs, and compliance. This rigorous evaluation of security tools is an indispensable part of modern cybersecurity strategy [4]. As threats continue to grow in sophistication, the ability to select and implement the right security tools, based on sound evaluation criteria, will be paramount in maintaining resilience and trust in the digital era. The framework in this paper aims to guide organizations in making informed decisions that enhance their security capabilities and protect their most valuable assets.

REFERENCES

[1] Viega, J., & McGraw, G. (2001). *Selecting and Evaluating Security Tools*. IEEE Security & Privacy, 1(2), 65-69.

[2] Swende, M. H. (2012). *Key Performance Indicators for Security Operations*. In *Journal of Information Security*.

[3] Cranor, L. F. (2000). *Evaluation Criteria for Security Tools*. In *Proceedings of the Computer Security Applications Conference*.

[4] Dafoe, A., & Hadfield, G. (2018). *Securing the Future of Artificial Intelligence and Machine Learning in Cybersecurity*. Journal of Cybersecurity, 4(1), 29-48.