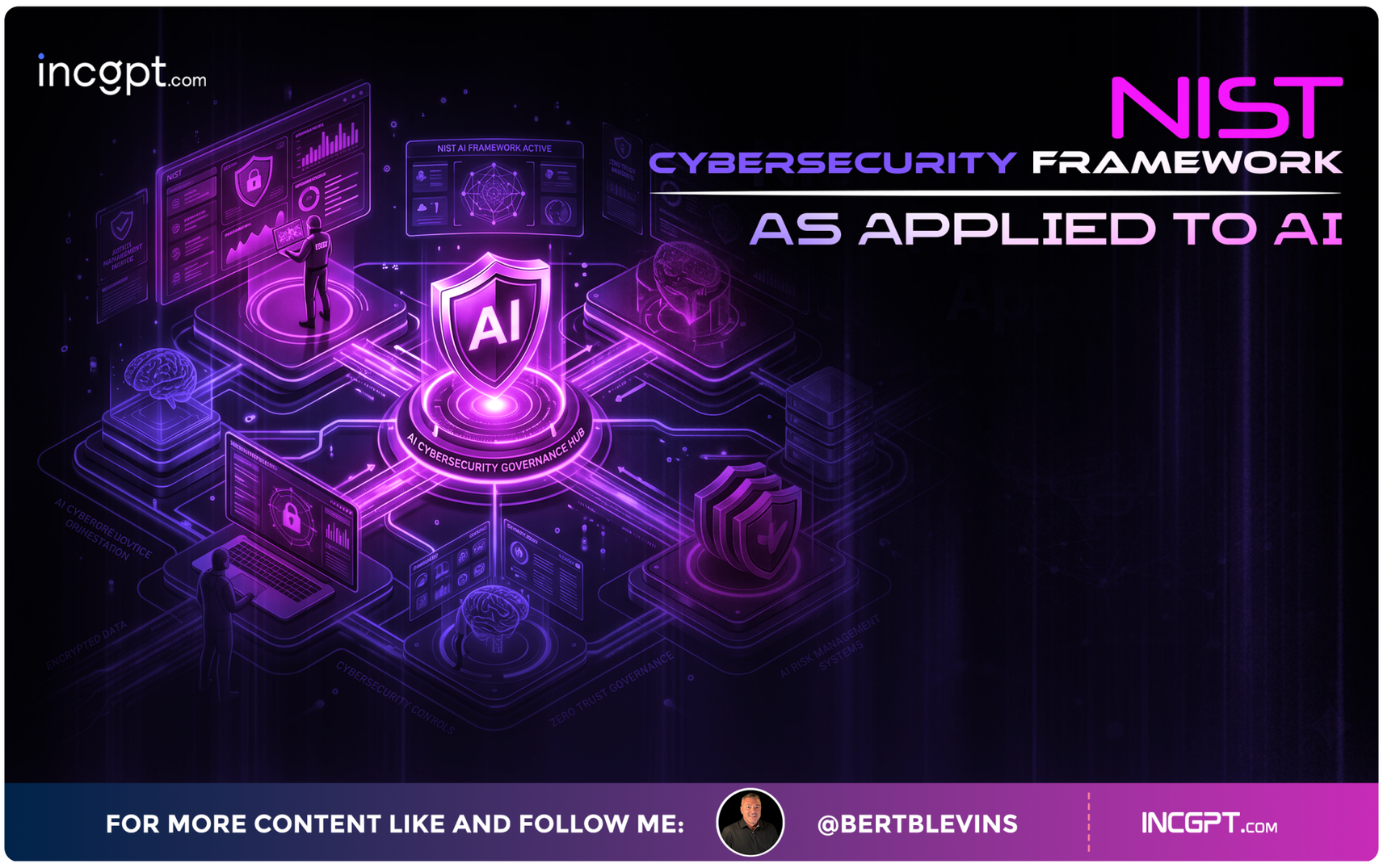
NIST Cybersecurity Framework as Applied to AI:
An In-Depth Guide to the Cyber AI Profile



Introduction

The integration of Artificial Intelligence (AI) into cybersecurity presents both vast opportunities and unique risks. As AI technologies evolve, they significantly enhance an organization's ability to detect, respond to, and mitigate cybersecurity threats. However, AI also introduces new vulnerabilities and challenges, such as adversarial machine learning attacks and data poisoning. To help organizations effectively manage these emerging risks, the NIST Cybersecurity Framework (CSF) 2.0 is being adapted to address AI-specific concerns through the development of the Cyber AI Profile.

This document serves as a comprehensive guide to understanding how the CSF 2.0 can be applied to AI in the context of cybersecurity, and it highlights the foundational principles outlined in the Preliminary Draft of NIST IR 8596 (December 2025). The Cyber AI Profile aligns with CSF 2.0's six core functions—Govern, Identify, Protect, Detect, Respond, and Recover—offering organizations a structured approach to integrate AI into their cybersecurity operations while managing risks. This guide expands on the Cyber AI Profile to provide practical, actionable insights for organizations navigating the intersection of AI and cybersecurity.

Acknowledgements

The creation of the Preliminary Draft of the Cyber AI Profile involved significant collaboration between cybersecurity and AI experts, industry stakeholders, and organizations like MITRE Corporation. NIST's dedication to stakeholder engagement is evident in the development of this profile, ensuring it reflects a wide range of perspectives and expertise. Public workshops, sessions, and consultations have helped identify critical AI-specific cybersecurity challenges and provided feedback that shaped the profile's development. This collaborative process ensures that the Cyber AI Profile is practical, comprehensive, and adaptable to the rapidly evolving cybersecurity landscape.

Emerging AI Opportunities in Cybersecurity

The rapid growth of AI brings transformative possibilities to the cybersecurity domain. AI can automate repetitive security tasks, analyze vast amounts of data in real-time, and identify complex patterns that may otherwise go undetected. For example, AI systems are already being deployed to:

- Automate incident detection and response, allowing for quicker identification and mitigation of threats.
- Enhance threat intelligence gathering, identifying anomalies and predicting potential vulnerabilities before they are exploited.
- Improve anomaly detection in network traffic and systems by analyzing large datasets and identifying deviations from typical behavior.

However, alongside these benefits, AI introduces challenges, especially when adversaries seek to exploit its vulnerabilities. Adversarial machine learning, in which attackers manipulate the input data used to train AI models, poses significant risks. Additionally, model drift and data poisoning are persistent threats to AI-powered systems. These vulnerabilities underscore the need for a structured approach to manage AI's integration into cybersecurity operations.

The Cyber AI Profile addresses these challenges by aligning AI opportunities with CSF 2.0's Identify Function (ID.RA: Risk Assessment), where organizations can use AI for enhanced threat detection and predictive vulnerability assessments, while also preparing for the potential risks introduced by AI.

Cyber AI Profile: A Framework for AI-Cybersecurity Integration

The Cyber AI Profile is designed to help organizations integrate AI into their cybersecurity strategies in a way that enhances defenses while managing risks effectively. It provides a structured framework that maps AI-related cybersecurity considerations to CSF 2.0's six core functions. This approach helps organizations understand how to protect AI systems, leverage AI to improve security, and defend against AI-enabled cyberattacks.

01

For example, under the Protect Function (PR.DS: Data Security), the profile focuses on safeguarding AI training data against tampering, such as data poisoning. The integration of controls from NIST SP 800-53 SC-12 is recommended to maintain data integrity and ensure that the AI models perform as expected without being compromised. The Cyber AI Profile also emphasizes the need for securing AI systems at every level, from model development to deployment, addressing potential vulnerabilities in AI's design and operational phases.

02

Incorporating principles from the AI Risk Management Framework (AI RMF), the profile provides guidance on establishing AI governance, maintaining robust configuration management, and ensuring that AI systems can be audited to prevent manipulation. This holistic approach to AI risk management ensures that organizations can confidently adopt AI technologies while maintaining a high level of security.

Key Focus Areas of the Cyber AI Profile

The Cyber AI Profile identifies three key focus areas critical for integrating AI into cybersecurity practices. These focus areas provide organizations with actionable insights on how to secure AI systems, leverage AI for cybersecurity defense, and protect against AI-driven cyberattacks:

01

Securing AI System Components

This focus area emphasizes the security of the AI systems themselves, including the algorithms, models, and underlying infrastructure. Securing AI system components is crucial to prevent attacks that target the foundational elements of AI systems, such as model manipulation and data poisoning.

Under CSF 2.0's Protect Function (PR.PS: Platform Security), the Cyber AI Profile recommends tracking AI hyperparameters, versions, and maintaining a secure development lifecycle to safeguard the integrity of AI models. Organizations are advised to implement strong controls over the AI supply chain (GV.SC-01) to prevent vulnerabilities from entering the system through third-party components.

By securing AI models and their components, organizations reduce the risk of attackers exploiting weaknesses in the AI infrastructure to bypass security measures.

02

Conducting AI-Enabled Cyber Defense

AI can enhance traditional cybersecurity defenses by automating incident detection and response. AI systems can process large volumes of data in real-time, detect anomalies, and identify new or emerging threats faster than traditional methods.

In the Detect Function (DE.AE: Adverse Event Analysis), the profile encourages the use of AI to correlate events across different security systems, enabling faster identification of potential threats. For example, AI can automate the triage process during incidents, allowing for rapid prioritization of threats (RS.MA-02). However, the profile also cautions that AI systems must be continuously monitored to avoid false positives or misclassifications.

Additionally, the profile stresses the importance of a hybrid approach to AI-driven cybersecurity, where human expertise complements AI's capabilities to ensure accuracy and resilience in detecting and mitigating threats.

03

Thwarting AI-Enabled Cyber Attacks

As AI becomes more advanced, it can also be weaponized by attackers. AI-driven cyberattacks, such as adaptive malware and advanced spear-phishing, are becoming increasingly sophisticated. The Cyber AI Profile provides guidelines for defending against AI-driven attacks by leveraging AI's capabilities to detect and respond to evolving threats.

Under the Respond Function (RS.MI: Incident Mitigation), the profile highlights the need to quickly contain AI-enabled attacks, such as by disabling compromised AI agents or countering adaptive malware that can change its behavior in response to detection.

The profile also encourages organizations to integrate threat intelligence into their defenses, enabling them to stay ahead of emerging attack vectors and adapt to new AI-driven threats.

Preliminary Draft of the Cyber AI Profile

The Preliminary Draft of the Cyber AI Profile serves as a foundation for refining the document based on input from cybersecurity professionals, industry stakeholders, and AI experts. The draft outlines the essential steps organizations need to take to secure their AI systems and effectively manage the risks associated with AI adoption. It maps the risks and opportunities of AI to the specific subcategories of CSF 2.0, assigning priority levels to each action.

01

For example, in the Identify Function (ID.AM-07), the profile stresses the importance of tracking metadata for AI systems to detect tampering or data poisoning. The draft also includes informative references to frameworks like OWASP AI Exchange and ENISA Threat Landscape 2025 to help organizations stay updated on emerging risks.

02

Organizations are encouraged to provide feedback on the Preliminary Draft through public workshops and consultations, ensuring that the final version of the profile remains practical and effective in addressing the evolving cybersecurity needs associated with AI technologies.

NIST Cybersecurity Framework (CSF) 2.0

The NIST Cybersecurity Framework (CSF) 2.0 is an established, widely recognized standard that provides organizations with a flexible and adaptable approach to managing cybersecurity risks. It is designed to be a comprehensive resource for developing, managing, and improving an organization's cybersecurity posture. The framework is built around five core functions: Govern, Identify, Protect, Detect, Respond, and Recover, each of which contains specific categories and subcategories that guide organizations in managing risks and vulnerabilities. These functions are designed to be flexible and scalable, making the CSF suitable for organizations of all sizes, sectors, and maturity levels.

01

While the CSF 2.0 covers a broad spectrum of cybersecurity needs, it does not specifically address the unique challenges presented by Artificial Intelligence (AI). Recognizing the growing role of AI in cybersecurity, NIST introduced the AI Risk Management Framework (AI RMF) to complement CSF 2.0. The AI RMF is designed to provide a more targeted, AI-specific approach to risk management, covering challenges like model drift, adversarial machine learning attacks, and data integrity issues.

02

Together, CSF 2.0 and the AI RMF offer a comprehensive methodology for managing the evolving risks associated with AI adoption, while also enabling organizations to capitalize on the powerful benefits AI brings to the cybersecurity landscape. This combined approach provides organizations with the tools and strategies to effectively incorporate AI technologies into their existing cybersecurity frameworks, ensuring that the opportunities of AI do not outweigh the risks.

Integrating AI into the CSF 2.0 Framework

The Cyber AI Profile is a crucial tool for organizations seeking to integrate AI into their cybersecurity strategies while addressing AI-specific risks. It effectively utilizes the flexible structure of CSF 2.0 to help organizations implement risk management strategies that are customized to their specific environments. This means that the profile allows for adjustments based on the organization's size, industry, or cybersecurity maturity. For example, a small business in the healthcare sector might use the profile to develop AI-specific risk management strategies for handling sensitive health data, while a financial institution might focus on securing AI models that assess financial transactions for fraud detection.

The Cyber AI Profile ensures that organizations can systematically assess and manage risks related to AI technologies within their existing cybersecurity risk management processes. By integrating AI considerations into the CSF 2.0 functions, organizations gain a more holistic view of their overall cybersecurity posture. The Identify function, for example, would not only focus on traditional asset identification but would also encompass identifying AI systems, understanding their vulnerabilities, and assessing associated risks, such as the potential for model poisoning or data manipulation.

Addressing AI-Specific Challenges in the CSF 2.0 Context

Incorporating the AI RMF into the CSF 2.0 framework introduces AI-specific challenges into the broader cybersecurity landscape. For instance:

01

Model Drift:

As AI models learn and evolve, their performance may degrade over time due to changes in the data they encounter. This can lead to inaccurate predictions and decisions. By addressing model drift within the Detect and Respond functions of the CSF 2.0, organizations can establish continuous monitoring processes to detect shifts in model behavior and implement corrective actions.

02

Adversarial Attacks:

These attacks manipulate input data to deceive AI models into making incorrect predictions or classifications. The Protect function in the CSF 2.0 framework, which is responsible for securing systems and data, can be tailored to address the specific vulnerabilities that make AI systems susceptible to adversarial manipulation. By applying protective measures such as data integrity checks, encryption, and anomaly detection, organizations can mitigate the risks of adversarial machine learning.

03

Data Integrity:

Since AI models rely on large datasets to make decisions, the integrity of the data used to train and operate these models is critical. The Govern function within the CSF 2.0 provides guidance on es..tablishing governance structures to ensure the integrity of AI data. Organizations can implement policies and procedures that enforce data validation and quality controls to reduce the risk of data corruption or manipulation, ensuring that AI models are based on reliable information.

Actionable Steps for Managing AI Risks

The Cyber AI Profile provides clear and actionable steps for managing AI-specific risks within the CSF 2.0 framework, offering organizations a structured pathway to secure AI systems and minimize potential cybersecurity threats. Some of the key actionable steps include:

01

Governance and Risk Assessment:

The profile encourages organizations to establish AI governance frameworks that integrate AI risk management directly into their existing cybersecurity governance structures. By linking AI-specific risks to the Govern function of CSF 2.0, organizations can ensure that AI adoption is aligned with their overall risk management strategies and regulatory requirements. This could include creating roles and responsibilities for AI risk management and ensuring that AI governance is continuously reviewed and updated.

02

Securing AI Systems and Data:

Under the Protect function of CSF 2.0, the Cyber AI Profile emphasizes securing AI systems, including training data, models, and outputs. This involves implementing strict data security measures, such as encryption and secure access controls, to protect AI systems from manipulation or unauthorized access. Furthermore, organizations are encouraged to protect AI models from adversarial attacks by using techniques like adversarial training and anomaly detection to detect attempts at model tampering.

03

Continuous Monitoring and Response:

The Detect and Respond functions are particularly critical when it comes to monitoring AI systems and responding to threats. The profile suggests using continuous monitoring to identify any signs of model drift or adversarial attacks. This could include establishing anomaly detection systems that can quickly identify when an AI model behaves unexpectedly or when an input has been manipulated. Furthermore, it provides guidelines for rapid response protocols, ensuring that AI models can be swiftly shut down or retrained in the event of a security breach.

04

AI in Recovery:

The Recover function of CSF 2.0 emphasizes the importance of having a plan in place for recovering from a cybersecurity incident. The profile highlights the role of AI in supporting recovery efforts, such as using AI-powered systems to identify affected areas and optimize recovery strategies. This can help reduce downtime and ensure a faster, more efficient recovery process, ensuring that organizations can return to normal operations as quickly as possible.

Maximizing AI Benefits While Minimizing Risks

By leveraging the complementary nature of CSF 2.0 and AI RMF, organizations can maximize the benefits of AI in cybersecurity while minimizing the associated risks. For example, AI's ability to enhance threat detection, incident response, and risk assessments can significantly improve an organization's overall cybersecurity posture. However, to fully realize these benefits, organizations must also address the unique challenges AI introduces, such as the risk of model vulnerabilities or the potential for data manipulation.

01

The Cyber AI Profile provides a practical roadmap for balancing the innovative advantages of AI with the necessary risk management. By following the actionable steps outlined in the profile, organizations can adopt AI technologies securely and responsibly, ensuring that their AI-driven cybersecurity strategies are both effective and resilient.

02

In summary, the NIST Cybersecurity Framework (CSF) 2.0, when integrated with the AI Risk Management Framework (AI RMF), offers a holistic and flexible approach to cybersecurity that addresses the unique risks posed by AI. The Cyber AI Profile brings these frameworks together, providing clear, actionable guidance for organizations to effectively manage AI risks, secure AI systems, and maximize the benefits AI offers in enhancing cybersecurity defenses.

Conclusion

The Cyber AI Profile provides a clear, structured framework for organizations aiming to integrate Artificial Intelligence (AI) into their cybersecurity operations, while also addressing the unique challenges and risks associated with AI technologies. As organizations continue to adopt AI, they must ensure that these systems are not only efficient but also secure. The Cyber AI Profile is designed to help organizations achieve this balance by aligning with the NIST Cybersecurity Framework (CSF) 2.0, which offers a proven, flexible approach to managing cybersecurity risks. Additionally, the AI Risk Management Framework (AI RMF) is incorporated to address the specific threats AI technologies can introduce.
By following the Cyber AI Profile, organizations can:

01

Secure AI systems:

The profile emphasizes the need for robust security measures, including securing the AI development lifecycle, protecting data, and mitigating risks such as adversarial attacks and data poisoning.

02

Enhance defenses:

AI can significantly improve threat detection, anomaly detection, and incident response when properly integrated into cybersecurity strategies. The profile highlights how AI can be leveraged to automate certain tasks while ensuring human oversight to avoid misclassifications or false positives.

03

Mitigate AI-related risks

The integration of AI presents new risks, such as model drift, adversarial machine learning, and vulnerability in AI models. The Cyber AI Profile offers detailed guidance on how to address these risks using CSF 2.0's core functions like Identify, Protect, Detect, and Respond.

This guide serves as an essential resource for organizations looking to navigate the complexities of AI adoption in cybersecurity, offering practical insights that enable the creation of more resilient and secure AI systems. With clear recommendations for AI governance, data protection, and incident response, the profile provides actionable steps that can be customized to an organization's unique needs.

For those seeking the latest version of the Cyber AI Profile, and to stay updated on its development, it is recommended to refer to NIST resources, such as NIST IR 8596.iprd, where the most current guidelines, feedback from stakeholders, and iterative updates are made available. This ongoing work ensures the Cyber AI Profile evolves in response to emerging AI threats and the evolving cybersecurity landscape.