Alert

4/25/2024

COZEN O'CONNOR

Joint Guidelines for Secure AI Deployment

New cybersecurity guidance for artificial intelligence (AI) systems, available here, was recently issued jointly by the U.S. Cybersecurity and Infrastructure Security Agency (CISA), the FBI, the National Security Agency's Artificial Intelligence Security Center, and cybersecurity agencies of Australia, New Zealand, the U.K., and Canada. The Guidelines, *Deploying AI Systems Securely: Best Practices for Deploying Secure and Resilient AI Systems*, are particularly notable because they focus on best practices for organizations that deploy AI developed by a third party rather than targeting developers of AI systems.

The Guidelines include, among other recommendations:

1. Manage and Secure the Deployment Environment

Prior to deployment, deployers of AI systems should verify that the deployment environment adheres to sound security principles, which include robust governance, well-designed architecture, and secure configurations. Deployers should also require the developer of the AI system to provide information about likely security threats to the system and account for deployment environment security requirements when developing contracts for AI system products or services.

2. Validate the AI System Before and During Use

The Guidelines include specific recommendations intended to ensure the integrity of the AI system. Recommended methods include the use of cryptographic methods, such as digital signatures and checksums, to confirm the authenticity of all artifacts. Additionally, hashed and encrypted copies of each release of the AI model and system should be created and stored in a tamper-proof location. All forms of code (including source code and executable code) and artifacts (*e.g.*, models, parameters, configurations, and data) should be stored in a version control system with proper access controls. The supply chain should be evaluated for any external AI models and data, ensuring vendors adhere to organizational standards and risk management policies.

3. Secure API

Application programming interfaces (APIs) should be secured by implementing authentication and authorization mechanisms. All input data should be validated and sanitized.

4. Actively Monitor Model Behavior

Deployers should collect logs that include inputs, outputs, intermediate states, and errors and should ensure that automated alerts are triggered for suspicious conditions. The model's architecture and configuration settings should be monitored for any unauthorized changes or unexpected modifications that might compromise the model's performance or security.

5. Protect Model Weights

Interfaces for accessing model weights, the parameters that are adjusted during the training process to create the designed output of the AI system, should be hardened to increase the effort it would take for an adversary to exfiltrate the weights. Weight storage should be aggressively isolated. For example, model weights should be stored in a protected storage vault, in a highly restricted zone (such as a separate dedicated enclave), or using a hardware security module (HSM).

Christopher Dodson

Member

cdodson@cozen.com
Phone: (215) 665-2174
Fax: (215) 372-2408

Andrew Baer

Chair, Technology, Privacy & Data Security

abaer@cozen.com
Phone: (215) 665-2185
Fax: (215) 372-2400

Related Practice Areas

- Artificial Intelligence
- Business
- Corporate
- Technology, Privacy & Data Security

6. Enforce Strict Access Controls

Deployers should apply role-based access controls (RBAC) or attribute-based access controls (ABAC), where feasible, to limit access to authorized personnel. Additionally, they should use multi-factor authentication (MFA) and privileged access workstations (PAWs) for administrative access.

7. Conduct Audits and Penetration Testing

Third-party security experts should be engaged in conducting audits and penetration testing on AI systems.

8. Update and Patch Regularly

The AI system should be kept up-to-date, and when updating to a new/different version, a full evaluation should be run to ensure that accuracy, performance, and security tests are within acceptable limits before redeploying the AI system.

Overall, the Guidelines provide specific, actionable best practices. Organizations deploying AI systems would be well served by following the recommendations.