APISecurityEngine

Owasp Top 10 LLM 02 : Insecure Output Handling

What Is Insecure Output Handling ?

Output handling refers to program output, including user interface, file storage, network communication, and processing.

Insecure handling in software indicates inadequate risk consideration for generated content.

Insecure output handling can manifest in various ways :

Injection attacks can occur when programs fail to validate user inputs, causing malicious code to be injected, resulting in data breaches or unauthorized actions.

Sensitive data exposure occurs when sensitive information is included without encryption or obfuscation.

Improper encoding can cause issues like broken web pages, data parsing, and security vulnerabilities.

Programs without validation or sanitization can manipulate output, causing unintended consequences.

Inconsistent output confuses users and attackers.

Attackers exploit error messages to gain insights into system functionality and weaknesses.

Solutions to Insecure Output Handling

1. Treating Model Output as Untrusted User Content and Validating Inputs :

Validate input data before using model to prevent vulnerabilities from malicious inputs or injection attacks.

Validate model output for format, content, and potential harmful content to ensure accuracy.

2. Encoding Model Output to Mitigate Undesired Code Interpretations :

Properly encode HTML tags and special characters for web interface output to prevent cross-site scripting attacks and malicious code execution.

Use appropriate methods to escape special characters to prevent code injection attacks.

Set Content-Type headers for model output to prevent browsers from treating content as executable code.

3. Penetration Testing (Pentesting) for Uncovering Insecure Outputs :

Pentesting detects model output vulnerabilities and mishandling issues.

Regular security audits identify model improvement opportunities.

Input fuzzing tests system response to malicious inputs, revealing model vulnerabilities.

Contact@apisecurityengine.com