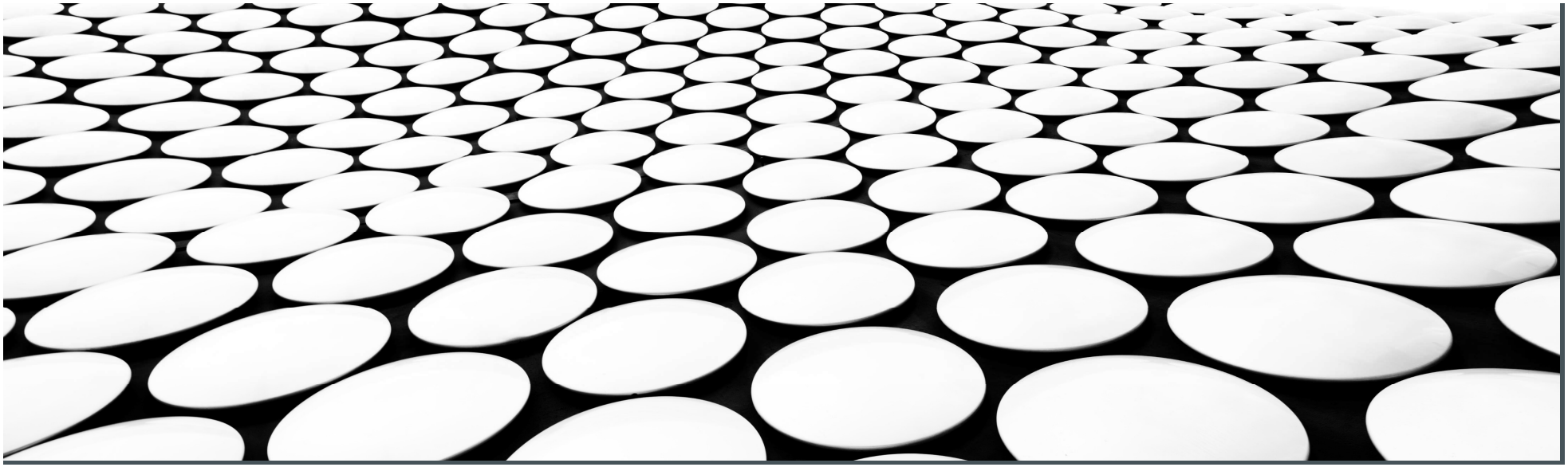
CERTIFIED AI SECURITY FUNDAMENTALS™ (CAISF™)

WORKSHOP 1 – HANDS-ON ACTIVITY



WORKSHOP 1: OBJECTIVE AND DETAILED EXPLANATION

- **Workshop Objective:** To apply the knowledge gained in the sessions by conducting threat modeling on hypothetical AI/ML systems. This involves identifying potential security threats and vulnerabilities within these systems.

  Duration: 1 hour

- Steps:

1. Introduction to the Hypothetical Business Scenario:
    1. An AI-enabled business where a platform connects consumers with repair contractors for home services, such as plumbing, electrical work, HVAC maintenance etc..
2. Importance of AI Threat Modeling:
    1. **Identifying AI Security Risks:** Explain that threat modeling is crucial for pinpointing security threats and vulnerabilities within AI/ML systems, particularly those that utilize large language models (LLMs) and generative AI (GENAI) technologies.
    2. **Mitigating AI Risks:** Emphasize how threat modeling helps in developing strategies to mitigate these risks, ensuring the platform remains secure and reliable.
    3. **Enhancing AI Security Posture:** Discuss how regular threat modeling can improve the overall security posture of the platform, making it resilient against potential cyberattacks.

DETAILED STEPS FOR THE SESSION:

1. Objective Clarification:
    1. The goal is to apply theoretical knowledge to a practical scenario, enhancing understanding through hands-on threat modeling.
2. Hypothetical Business Scenario:
    1. Introduce a fictitious platform that matches consumers with home repair contractors.
    2. The platform uses AI/ML to optimize matching, schedule appointments, and manage communications.
3. Threat Modeling Process:
    1. **System Overview:** Begin with an overview of the AI/ML system architecture, including data flow, key components, and user interactions.
    2. **Threat Identification:** Identify potential security threats, such as data breaches, unauthorized access, and adversarial attacks on AI models.
    3. **Vulnerability Analysis:** Analyze vulnerabilities within the system, focusing on areas where LLM and GENAI technologies are implemented.
    4. **Risk Assessment:** Assess the potential impact and likelihood of identified threats, prioritizing them based on severity.
    5. **Mitigation Strategies:** Develop strategies to mitigate identified risks, such as implementing stronger authentication, encrypting sensitive data, and conducting regular security audits.

CONCLUSION AND DISCUSSION:

1. Summarize the key findings from the risk assessment and threat modeling exercise.

2. Discuss the importance of ongoing threat modeling and risk management in maintaining a secure AI/ML platform.

3. Perform This Steps:
   A. **System Overview:** Begin with an overview of the AI/ML system architecture, including data flow, key components, and user interactions.
   B. **Threat Identification:** Identify potential security threats, such as data breaches, unauthorized access, and adversarial attacks on AI models.
   C. **Vulnerability Analysis:** Analyze vulnerabilities within the system, focusing on areas where LLM and GENAI technologies are implemented.
   D. **Risk Assessment:** Assess the potential impact and likelihood of identified threats, prioritizing them based on severity.
   E. **Mitigation Strategies:** Develop strategies to mitigate identified risks, such as implementing stronger authentication, encrypting sensitive data, and conducting regular security audits.
   F. **Other Optional Steps**

QUESTION AND ANSWER