🛡️DEF CON 33 Workshop Exercise: Writing Agile Abuse Stories for GenAI Systems

Agile Abuse Stories are user stories written from an attacker's perspective. They help identify potential abuse early in development and align with agile workflows.  
  
Each participant/team will write abuse stories for different components of a GenAI system. Use the structure below and complete at least 2 examples during this exercise.

# 📌 Template Format:

- Abuser Persona: [Define the attacker — e.g., malicious user, insider, compromised plugin]  
- Abuse Scenario: [Describe what the attacker does]  
- Impact: [What harm this causes — e.g., data leakage, privilege escalation]  
- Mitigation: [How to prevent or minimize the impact]

🔁 Repeat the following structure for each abuse story:

Example:  
"As a malicious user, I want to exploit insecure direct object references to access unauthorized data."  
"I will tamper with parameters to view another user's data."  
"This will result in unauthorized data disclosure, violating confidentiality."  
"Mitigate: Implement access controls and validate all user input."

### 1. ****Three-Tier Web Architecture**** (App + API + Database)

**Scenario:**

An attacker notices that object IDs in the URL of a SaaS web application can be guessed or incremented to access resources belonging to other users.

**Prompt to Participants:**

As an attacker, how would you exploit Insecure Direct Object References (IDOR) in the web layer?  
Write the abuse persona, scenario, impact, and potential mitigation.

# ✍️ Write Your Abuse Stories Here:

--- Abuse Story #1 ---

Abuser Persona: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Abuse Scenario: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Impact: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Mitigation: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### ✅ 2. ****GenAI-Powered Chatbot System****

**Scenario:**

A malicious user uploads a PDF document to the GenAI chatbot with hidden instructions embedded in white text. The model processes the PDF and unknowingly executes the hidden prompt, leaking internal user credentials via error messages.

**Prompt to Participants:**

As a prompt injection attacker, describe how this hidden input could exploit the LLM.  
Write the abuse story including who the attacker is, what the model does, and how to mitigate it.

--- Abuse Story #2 ---

Abuser Persona: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Abuse Scenario: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Impact: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Mitigation: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

### ✅ 3. ****Cloud Infrastructure (IAM & Serverless)****

**Scenario:**

A serverless function in AWS processes incoming alerts and escalates them to PagerDuty. The function uses an IAM role with full ses:\*, s3:\*, and ec2:\* permissions due to misconfigured trust boundaries.

**Prompt to Participants:**

As an insider or compromised automation script, how could this excessive IAM permission be abused?  
Write an abuse story focusing on excessive privilege, cloud misconfig, and defense controls.

--- Abuse Story #3 ---

Abuser Persona: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Abuse Scenario: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Impact: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Mitigation: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_