OWASP

GenAI SECURITY PROJECT

genai.owasp.org

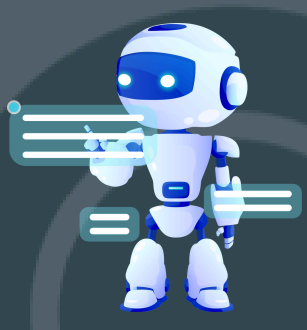
AGENTIC AI THREAT NAVIGATOR

A cheat sheet for identifying and assessing agentic AI threats!

Agency & Reasoning

Does the AI agent independently determine the steps needed to achieve its goals? Related threats are:

- **Misaligned and Deceptive Behaviors**
- **Intent Breaking and Goal Manipulation**
- **Repudiation and Untraceability**



Memory & Context

Does the AI agent rely on stored memory for decision-making? Related threats are:

- **Memory Poisoning**
- **Cascading Hallucinations**

Tools and Execution

Does the AI agent execute actions using tools, system commands, or external integrations? Related threats are:

- **Tool Misuse**
- **Resource Overload**
- **Unexpected RCE and Code Attacks**
- **Privilege Compromise**

Identity and Authentication

Does the AI system rely on authentication to verify users, tools, or services? Related threat is:

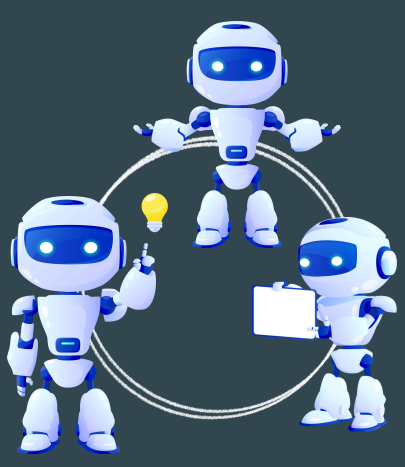
- **Identity Spoofing and Impersonation**

Human Engagement

Does AI require human engagement to achieve its goals or function effectively? Related threats are:

- **Overwhelming Human-in-the-Loop (HITL)**
- **Human Manipulation**

Multi-Agency



Does the AI system rely on multiple interacting agents? Related threats are:

- **Agent Communication Poisoning**
- **Rogue Agents**
- **Human Attacks**

genai.owasp.org/initiatives/#agenticinitiative