Worksheet: Threat Modeling a Research Literature Assistant

Use this worksheet during Topic 07 Activity 1. Build a short threat-model entry for a research literature assistant using the threat-card shape, then map it to one OWASP LLM Top 10 category.

The Application

A reading helper for student researchers. It accepts uploaded PDFs, local notes, and a student question. It produces summaries, cross-paper comparisons, draft citations, and surfaced passages.

Pick One Scenario

- A paper with hidden instructions or malformed citations.
- A poisoned summary or fabricated comparison note.
- Private draft notes exposed through the assistant.
- A large upload that drives cost or latency.

Selected scenario: _____

Threat Card

- Component: _____
- Assumption: _____
- Attacker capability: _____
- Attack path: _____
- Impact: _____

OWASP Mapping

- Best-fit OWASP category (LLM01..LLM10): _____
- Why this category fits: _____

Boundary Walk

- One asset: _____
- One trust boundary crossed: _____
- One influence point (untrusted text becomes model-visible): _____
- One output reuse point (output drives a downstream action): _____

Reflection

- What part of the threat-card shape was hardest to fill in?
- Did the scenario cross more than one boundary? Which mattered most?
- What classroom prompt would make a student apply this same shape to a real system?