Agent Threat Model Worksheet

1. Map the workflow and the authority behind each action.

System / workflow

Review date

Accountable owner / team

Intended task and allowed outcome

Inputs and trust boundaries

Include external documents, tool results, memory, and other agents.

Identities, delegation, and credentials

Who acts, whose authority applies, and when does access expire?

Tools, destinations, and actions that need approval

Agent Threat Model Worksheet

2. Test abuse paths, record evidence, and plan recovery.

Security property that must hold

Example: an agent cannot export data beyond the approved recipient.

Abuse case: entry point, action, and impact

Control, enforcement boundary, and owner

Test result and evidence reference

Containment trigger and recovery conditions

Residual risk and follow-up decision

Reviewer / decision owner

Next review date