AI Governance & Cybersecurity — Lab Rubric

Each lab is evaluated across three dimensions: Exploit Demonstration, Defense Implementation, Governance & Reflection.

Scoring per lab: 10 points max (40 points total across 4 labs).

|  |  |  |  |
| --- | --- | --- | --- |
| Lab | Exploit (3 pts) | Defense (4 pts) | Governance & Reflection (3 pts) |
| Lab 1 — Prompt Injection | Show secret exfiltration | Denylist, guardrails, redaction | Explain confidentiality risk, map to frameworks |
| Lab 2 — AI-Generated Code | Exploit eval/SQLi | Use safe AST eval & parameterized queries | Discuss insecure AI code + compliance |
| Lab 3 — MCP RCE | Demonstrate RCE exploit | Allowlist + shell=False | Map to config mgmt & governance |
| Lab 4 — Agentic Browser | Exfiltrate via meta instructions | Domain allowlist + consent | Accountability & post-market monitoring |