🏛️ AI Governance & Cybersecurity — Capstone Project Report

# 1. Project Overview

* Team Members: [List names]
* Chosen AI System: [e.g., Chatbot, MCP server, agentic browser]
* System Description: [Brief description of what it does, architecture, and context]

# 2. System Audit & Threat Model

* System Architecture Diagram (if available):
* Assets at Risk: [Data, secrets, model weights, workflows]
* Threat Actors: [Who might attack?]
* Threat Modeling Approach: [e.g., STRIDE, MITRE ATLAS]
* Prioritized Risks: [High/Med/Low with justification]

# 3. Exploit Demonstration

* Exploit Chosen: [Type: prompt injection, SQLi, RCE, agentic browser, etc.]
* Steps to Reproduce: [Commands, payloads, screenshots]
* Exploit Output: [Logs or screenshots]
* Root Cause Analysis: [Why did it work?]

# 4. Defense Implementation

* Defense Strategy: [e.g., input validation, sandboxing, denylist, monitoring]
* Steps Taken: [Code snippets, configuration changes]
* Defense Output: [Evidence attack now fails]
* Trade-offs / Limitations: [Impact on usability, performance, etc.]

# 5. Governance & Compliance Integration

* Relevant Frameworks: [EU AI Act, NIST AI RMF, ISO/IEC 42001, SOC-2, HIPAA, etc.]
* Governance Mapping: [How the exploit/defense ties into framework obligations]
* Accountability Discussion: [Who is responsible if this fails?]
* Lifecycle Monitoring: [How would this be continuously assured post-deployment?]

# 6. Reflection & Lessons Learned

* What was the biggest challenge?
* What did your team learn about AI security?
* How can this exercise influence real-world AI governance?

# 7. Appendix (Optional)

* Full Code Listings
* Screenshots
* References