1. Attack Summary

Describe the most successful attack vectors used (prompt injection, jailbreak, etc.)

2. Defense Measures

List concrete mitigations applied (isolation, regex filter, moderation API…)

3. Evaluation

Did the attack still succeed post-patch? ☐ Yes ☐ No
Evidence / logs:

4. Lessons Learned

Key takeaways & potential next-step hardening.

Reviewer Sign-off

Name: ________________________ Signature: ___________________