assessment brief

Assessment Brief

Assessment Brief: Secure RAG Implementation (Assignment 2)

Unit: LLM Security and Governance (Postgraduate Unit)

Weighting: 30% of total unit marks

Submission Requirements:

* A written report detailing the implementation of a secure RAG pipeline, including:
    + Code snippets for key components
    + Explanation of design decisions and trade-offs
    + Discussion of potential vulnerabilities and mitigation strategies
* A separate appendix containing:
    + Prompt-injection test suite code (including test cases)
    + Results of testing (attack success rate, false block rate, JSON validity rate, citation c
overage)

Assessment Criteria:

1. **Design and Implementation (40%)**
    * Clarity and effectiveness of RAG pipeline design
    * Correctness and efficiency of implementation
    * Adherence to secure coding practices
2. **Prompt-Injection Testing (30%)**
    * Thoroughness and comprehensiveness of test suite
    * Accuracy and reliability of results (attack success rate, false block rate, JSON validity
 rate, citation coverage)
3. **Compliance and Ethics (30%)**
    * Clarity and thoroughness of compliance and ethics documentation
    * Adherence to relevant cyber law and regulatory requirements

Submission Guidelines:

* Submit a single PDF document containing the written report and appendix.
* Use a clear and consistent formatting style throughout.
* Include a cover page with student name, ID number, and unit code.

Word Limit: 2,500 words (excluding references)

Due Date: [Insert due date]

Late Submission Penalty: [Insert late submission penalty policy]

Marking Rubric: Secure RAG Implementation

| Criteria | Excellent (90-100%) | Good (80-89%) | Satisfactory (70-79%) | Needs Improvement (B
elow 70%) |
| --- | --- | --- | --- | --- |

Design and Implementation	Clear, effective design; efficient implementation; secure coding practices.	Good design; some minor issues with implementation or security.	Adequate design; significant issues with implementation or security.	Poor design; major issues with implementation or security.
Prompt-Injection Testing	Thorough and comprehensive test suite; accurate and reliable results.	Good test suite; some minor issues with accuracy or reliability.	Adequate test suite; significant issues with accuracy or reliability.	Poor test suite; major issues with accuracy or reliability.
Compliance and Ethics	Clear, thorough compliance and ethics documentation; adherence to relevant cyber law and regulatory requirements.	Good compliance and ethics documentation; some minor issues with adherence.	Adequate compliance and ethics documentation; significant issues with adherence.	Poor compliance and ethics documentation; major issues with adherence.

Student Submission Checklist: Secure RAG Implementation

* Have I implemented a secure RAG pipeline?
* Is my prompt-injection test suite thorough and comprehensive?
* Have I reported accurate results for attack success rate, false block rate, JSON validity rate, and citation coverage?
* Is my compliance and ethics documentation clear and thorough?
* Have I adhered to relevant cyber law and regulatory requirements?

Note: This checklist is for student reference only.