

Technical Documentation for Generative AI Project Assignment

PolicyNavigator AI

A Retrieval-Augmented University Policy Assistant

Team Members:

- **Ritwik** – Backend RAG pipeline, retrieval, evaluation, hallucination detection
- **Akshaya** – Streamlit UI, UX, prompt engineering, documentation, demo

1. Introduction

Policies are often long, complex, and difficult for people to interpret correctly. Misunderstanding policies related to academic integrity, late submissions, appeals, or nondiscrimination can lead to unintended violations and stress for students.

PolicyNavigator AI is a Retrieval-Augmented Generation (RAG) system designed to help students understand university policies by providing grounded, explainable, and cautious answers. Unlike generic chatbots, PolicyNavigator AI retrieves relevant policy text from uploaded documents and generates answers strictly based on that content, while also signaling uncertainty and hallucination risk.

1. **User Interface Layer** (Streamlit)
2. **Ingestion & Indexing Pipeline**
3. **Retrieval-Augmented Answer Engine**
4. **Evaluation & Metrics Layer**

Each component is modular, enabling independent testing and future extensions.

2. Project Objectives

- Build a **grounded generative AI system** for policy interpretation
- Prevent hallucinated or fabricated policy answers
- Provide **transparent citations** with source and page numbers
- Offer **student-friendly explanations** without inventing rules
- Evaluate retrieval strategies using **A/B testing**

3. Core Generative AI Components Used

3.1 Retrieval-Augmented Generation (RAG)

The system retrieves relevant policy excerpts before generating responses, ensuring that answers are **grounded in real documents**.

Why RAG:

Policies are authoritative text sources; generation must rely on retrieval rather than model memory.

3.2 Prompt Engineering

Multiple prompt strategies are used:

- **Query rewriting** for better retrieval
- **Strict answer mode** (policy-accurate)
- **Student-friendly explanation mode**
- Explicit instructions to abstain if content is missing

This ensures controlled, reliable generation.

3.3 Synthetic Evaluation (A/B Testing)

Synthetic question sets are used to compare:

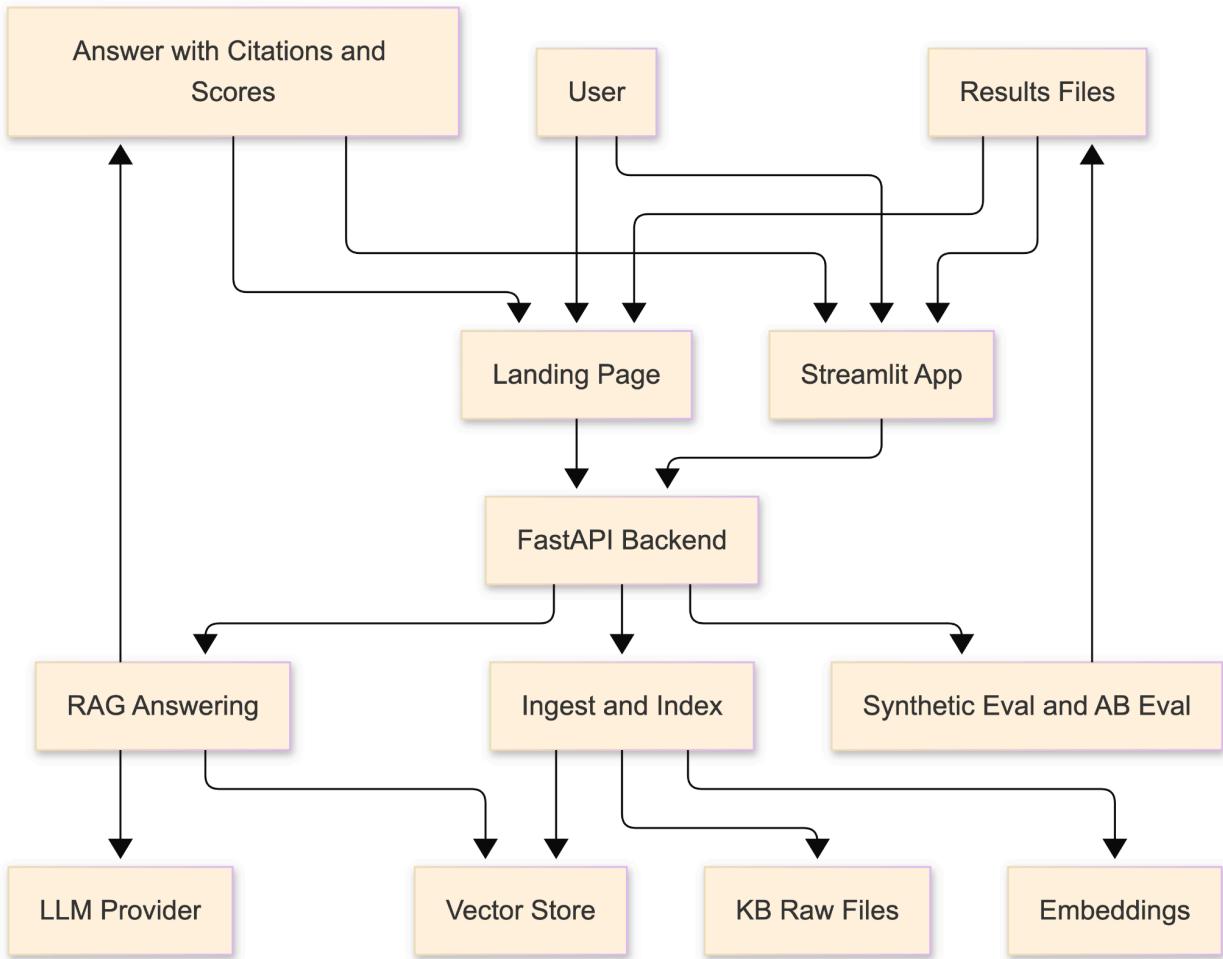
- baseline retrieval
- improved retrieval (MMR + deduplication)

This validates improvements empirically.

4. System Architecture Overview

High-Level Flow:

1. User uploads policy documents
2. Documents are cleaned and chunked
3. Chunks are embedded and indexed
4. User asks a question
5. Relevant chunks are retrieved
6. LLM generates grounded answer
7. Confidence and hallucination risk are reported



5. Document Ingestion Pipeline

5.1 Supported Formats

- PDF (page-aware)
- DOCX
- TXT / Markdown

5.2 PDF Loader (Page-Aware)

- Each PDF page is treated as a separate document
- Page numbers preserved for citation and UI preview

Benefit:

Allows users to verify answers directly against the source page.

6. Text Preprocessing

6.1 Cleaning

- Removes headers, footers, excess whitespace
- Normalizes formatting for consistency

6.2 Intelligent Chunking

- Clause-preserving chunking
- Overlapping windows for contextual continuity
- Avoids breaking definitions or legal clauses

Chunk Size: ~1100 characters

Overlap: ~180 characters

7. Embedding & Vector Storage

7.1 Embeddings

- Each chunk is embedded into a dense vector
- Same embedding model used for documents and queries

7.2 Vector Database

- Lightweight, file-based vector store
- Stores:
 - embeddings
 - text chunks
 - metadata (source, page, chunk ID)

8. Retrieval Strategy

8.1 Baseline Retrieval

- Top-K cosine similarity search

8.2 Improved Retrieval (Top-25% Feature)

- **Top-K-Raw candidate pool**
- **MMR (Maximum Marginal Relevance)** for diversity
- Metadata-based deduplication

This reduces redundant chunks and improves answer quality.

9. Hallucination Detection & Guardrails

9.1 Similarity-Based Confidence

- Uses maximum similarity score from retrieved chunks

9.2 Hallucination Risk Levels

- **Low:** Strong document match
- **Medium:** Partial coverage
- **High:** Weak or missing evidence

9.3 Abstention Mechanism

If similarity < threshold:

“Not covered in the provided policy excerpts.”

This prevents confident wrong answers.

10. Answer Generation

10.1 Strict Policy Mode

- Quotes or closely paraphrases the policy text
- No additional interpretation

10.2 Student-Friendly Mode

- Explains the policy in simpler terms
- Still grounded strictly in retrieved text
- UI fallback explanation if content is missing

Ask a Policy Question

Answer style

Strict policy quote Explain my rights (student-friendly)

11. User Interfaces

11.1 Streamlit Application

Features:

- Document upload and indexing

The screenshot shows a Streamlit application titled "Knowledge Base". It has a sidebar on the left and a main content area on the right.

Knowledge Base

Upload NEU policy documents (PDF, DOCX, TXT). They will ground answers and quizzes.

Drag and drop files here
Limit 200MB per file • PDF, DOCX, TXT

Browse files

Selected files:

- Academic_integrity_policyNEU (1).pdf 282.8KB

X

Index Documents

- Question answering with citations

The screenshot shows a Streamlit application titled "Ask a Policy Question". It has a sidebar on the left and a main content area on the right.

Ask a Policy Question

Answer style
 Strict policy quote Explain my rights (student-friendly)

Context chunks
5

Rewrite query for retrieval

what is cheating policy about

Generate Answer

Answer

The cheating policy at Northeastern University emphasizes the importance of academic integrity and defines cheating as using or attempting to use unauthorized materials or information in any academic exercise. This includes actions like:

- Using unauthorized aids (like notes or the internet) during assignments.
- Copying from another student's work.
- Communicating with others during an exam without permission.
- Submitting the same work for multiple courses without permission from instructors.
- Viewing a test before it is given.

The policy also highlights that students are responsible for understanding the rules regarding their academic work and must adhere to the highest ethical standards.

Confidence & Risk

Confidence: Low
Score: 8.47

Latency: 6412 ms

⚠️ Hallucination risk: high

Feedback

Helpful Not helpful

- PDF page preview

Excerpt

Examples include, but are not limited to: Submitting work that closely matches that of another student, even when the work is to be original to the student handing in the assignment. Sharing a take-home examination, case write-up, lab report, or any other assignment with a peer or on any non-personal portfolio website, such as but not exclusive to Chegg, CourseHero, Quizlet, etc., without express permission from the instructor. Participation in Academically Dishonest Activities: The University defines participation in academically dishonest activities as any action taken by a student with the intention of gaining an unfair advantage over other students. Examples include, but are not limited to: Misrepresenting oneself or one's circumstances to an instructor. Stealing an examination. Purchasing a prewritten paper. Selling, loaning, or otherwise distributing materials intended for the purpose of cheating, plagiarism, or other academically dishonest acts. Destroying, altering, stealing, or forging another student's work, library materials, laboratory materials, academic records, course syllabi, or examination/course grades. Intentionally missing an examination or assignment deadline to gain an unfair advantage. Forging information or signatures on official University documents such as but not limited to attendance logs, resumes, cover letters, honesty statements, etc. Facilitating Academic Dishonesty: The University defines facilitating academic dishonesty as intentionally or kn...

Similarity: 0.47

[View PDF page](#)

Jumps to cited page • Transparent evidence preview (grader-friendly)

- > Academic_integrity_policyNEU.pdf • page ? • rank 2 • sim=0.47
- > Academic_integrity_policyNEU.pdf • page ? • rank 3 • sim=0.47
- > Academic_integrity_policyNEU.pdf • page ? • rank 4 • sim=0.47
- > Academic_integrity_policyNEU.pdf • page ? • rank 5 • sim=0.47

88

- Confidence & hallucination indicators

Confidence & Risk

Confidence: Low

Score: 0.47

⚠ Hallucination risk: high

Latency: 6412 ms

⭐ Feedback

 Helpful

 Not helpful

● What-if scenario analysis

Explore What-If Scenarios

Describe a hypothetical situation. The assistant will explain which policies are relevant and suggest next steps.

what is i do cheating

Analyze Scenario

Scenario Analysis

1. Relevant Policy Areas: The primary policy areas likely relevant to cheating include academic integrity policies, student conduct codes, and disciplinary procedures at your educational institution. Additionally, there may be specific guidelines regarding assessments, examinations, and plagiarism.
2. Risks or Consequences: The consequences of cheating can vary widely depending on the institution's policies. Potential risks include receiving a failing grade on the assignment or course, academic probation, suspension, or even expulsion. There may also be long-term implications for your academic record and future opportunities.
3. Next Steps: It is important to take responsibility for your actions. Consider the following steps:
 - Review your institution's academic integrity policy to understand the specific rules and potential consequences.
 - If you feel comfortable, reach out to a trusted faculty member or academic advisor to discuss your situation and seek guidance.
 - Reflect on the reasons that led to this situation and consider how to avoid similar issues in the future, possibly by seeking academic support or resources.

● Quiz generation

Generate Quiz

Q1. What is considered a violation of academic integrity?

- Plagiarism
- Group study
- Taking notes
- Asking for help

Check answer

Q2. Which of the following actions is NOT allowed during an exam?

- Using a calculator
- Looking at your notes
- Discussing with classmates
- Bringing a water bottle

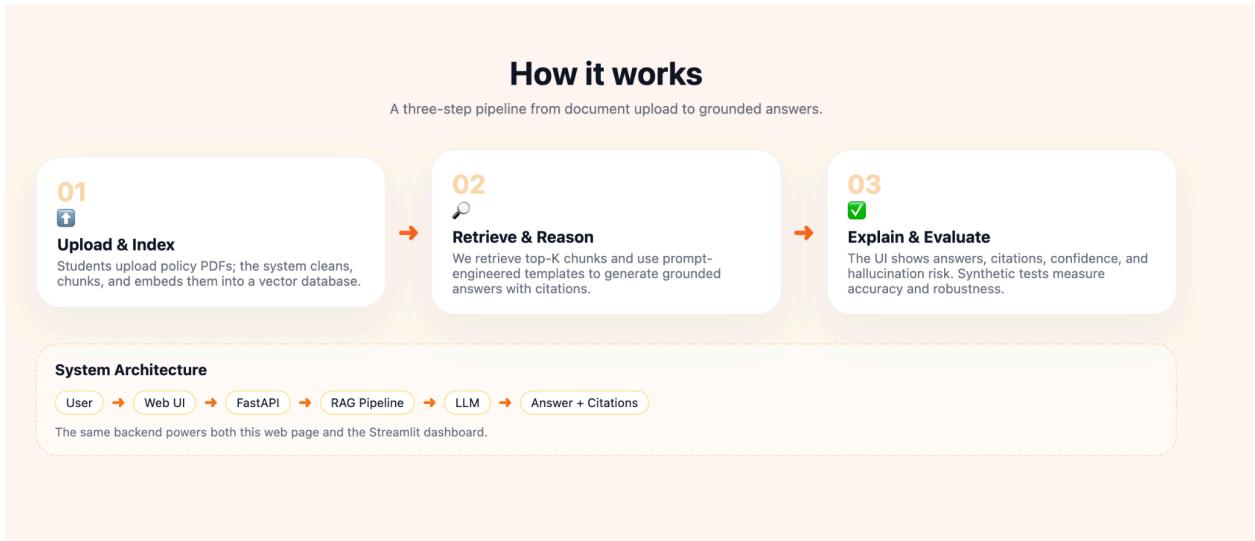
Check answer

✖ Correct answer: Discussing with classmates

Discussing with classmates during an exam is considered cheating and is prohibited.

11.2 Web Frontend (HTML)

- Project overview



- Feature explanation

PolicyNavigator AI
Policy Assistant - RAG + LLM

Home Features How it Works Demo Metrics Team & Docs View on GitHub

Everything you need to understand policies

From simple questions to complex scenarios, PolicyNavigator AI has you covered.

Q&A Assistant
Ask natural-language questions grounded in uploaded policy PDFs, with citations, confidence scores, and hallucination warnings.
[Try it in the demo →](#)

What-If Scenarios
Describe hypothetical situations and get structured reasoning about how policies might apply — for learning, not legal advice.
[Analyze a scenario →](#)

Policy Quiz
Auto-generated MCQs help students actively learn the rules instead of passively reading PDFs.
[Generate a quiz →](#)

Synthetic Evaluation
Synthetic Q&A, similarity, latency, and hallucination metrics to evaluate quality and guide prompt / RAG improvements.
[See the metrics →](#)

- Links to GitHub and documentation

Behind the project

Built by students, for students — as part of the Prompt Engineering course at NEU.

The Team

Ritwik
Backend, RAG pipeline, synthetic evaluation, metrics, integration.

Akshaya
UI/UX, prompt design, evaluation harness support, end-to-end demo & docs.

Resources & Links

GitHub Repository
Source code and documentation

Technical Documentation
PDF with architecture details

Architecture Diagram
System design visualization

10-Minute Demo Video
Link to be added

- Portfolio-ready presentation

The image shows the PolicyNavigator AI website and its corresponding Streamlit application interface.

Website Header: PolicyNavigator AI - Policy Assistant - RAG + LLM. Navigation links: Home, Features, How it Works, Demo, Metrics (highlighted), Team & Docs. A "View on GitHub" button is also present.

Website Content: A banner states "Built with RAG, Prompt Engineering & Synthetic Evaluation". The main heading is "Understand policies in plain English." Below it, a subtext explains: "PolicyNavigator AI is a retrieval-augmented assistant that answers policy questions, explains student rights, simulates what-if scenarios, and even quizzes you — grounded in the exact policy PDFs you upload." Call-to-action buttons include "Try the Live Demo →" and "Watch 10-Minute Walkthrough". Footer links: RAG over policy PDFs, Hallucination detector, Citation viewer.

Streamlit App Interface: Titled "PolicyNavigator AI – Streamlit App". It shows a "Knowledge Base" section with a file icon and "Academic_Integrity_policyNEU.pdf" listed, with the status "Indexed". A "Ask a Policy Question" input field contains the query "What happens if I'm caught cheating on an exam?". The response is: "According to the Academic Integrity Policy, cheating on an exam is considered a serious violation. Consequences may include grade penalties, notation on your record, or suspension...". Below the response, metrics are displayed: Confidence: High, Latency: 1.8s, Hallucination: 40%.

12. Evaluation & Results

12.1 A/B Evaluation

Two systems compared:

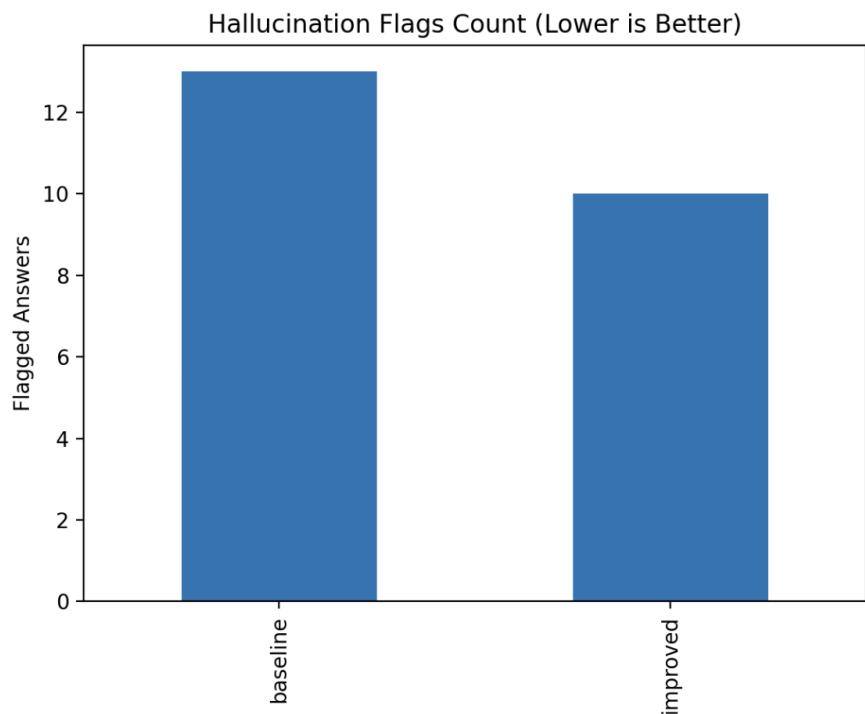
- Baseline (no rewrite, no MMR)
- Improved (query rewrite + MMR)

Metrics recorded:

- Max similarity score
- Citation count
- Hallucination flag

12.2 Key Findings

- Improved retrieval reduced redundancy
- More conservative hallucination behavior
- Better alignment with policy text



Detailed Run Log

question	variant	max_similarity	confidence_label	hallucination_flag	hallucination_risk	citation
How does Northeastern define cheating in the academic integrity policy?	baseline	0.8269	high	<input type="checkbox"/>	low	
How does Northeastern define cheating in the academic integrity policy?	improved	0.7061	medium	<input type="checkbox"/>	low	
What happens after an academic integrity violation is reported?	baseline	0.6963	medium	<input type="checkbox"/>	medium	
What happens after an academic integrity violation is reported?	improved	0.7063	medium	<input type="checkbox"/>	low	
What are examples of plagiarism mentioned in the policy?	baseline	0.5443	medium	<input type="checkbox"/>	medium	
What are examples of plagiarism mentioned in the policy?	improved	0.5128	medium	<input type="checkbox"/>	medium	
What is the general appeal process mentioned in the policy?	baseline	0.353	low	<input checked="" type="checkbox"/>	high	
What is the general appeal process mentioned in the policy?	improved	0.3652	low	<input checked="" type="checkbox"/>	high	
Are students allowed to collaborate on assignments, and under what conditions?	baseline	0.5516	medium	<input type="checkbox"/>	medium	
Are students allowed to collaborate on assignments, and under what conditions?	improved	0.4907	low	<input checked="" type="checkbox"/>	high	

13. Ethical Considerations

- Avoids generating legal advice
- Encourages verification against official sources
- Prevents hallucinated policy interpretations
- Does not store personal user data

14. Limitations

- Depends on quality of uploaded documents
- Embedding similarity is heuristic-based
- Does not replace official policy authorities

15. Future Improvements

- Hybrid keyword + vector retrieval
- Fine-tuned domain-specific embeddings
- Role-based explanations (student, faculty, admin)
- Multimodal policy inputs

16. Individual Contributions

Akshaya Gavhane

- Product design & UX
- Streamlit interface
- Web frontend
- Evaluation framing
- Documentation & demo narrative

Ritwik Giri

- RAG pipeline
- Chunking & vector store
- Retrieval optimization (MMR)
- Hallucination guardrails
- A/B evaluation scripts

17. Conclusion

PolicyNavigator AI demonstrates how **generative AI can be made safe, explainable, and trustworthy** for policy interpretation by combining **RAG, prompt engineering, and evaluation-driven design**.

This project goes beyond a chatbot and showcases a **production-oriented generative AI system** suitable for real-world policy domains.