

PoliticAI / ELECTOR Feed Project Report

Prepared by: Raghuvaran D

Position: Freelancer

Project Type: AI Safety Hackathon MVP Report

Track Fit: Global South AI Safety Challenge, Asia Track

Submission Focus: RAG-based political news and civic intelligence system

Index

1. Executive Summary
2. Project Overview
3. Problem Statement
4. Objectives
5. Why This Matters in Asia
6. Proposed Solution
7. System Architecture
8. Data and Source Strategy
9. RAG Pipeline Design
10. Safety and Hallucination Controls
11. MVP Features
12. Evaluation Plan
13. Expected Outcomes
14. Limitations
15. Future Scope
16. Conclusion

1. Executive Summary

PoliticAI, also referred to in the product direction as ELECTOR Feed, is a retrieval-augmented generation (RAG) political news and civic intelligence MVP built to organize recent, source-linked political content into a safe and queryable knowledge system. The core design goal is to help users explore political developments, election-related reporting, misinformation trends, and region-specific public-interest issues through grounded retrieval rather than unsupported AI generation.[cite:493][cite:495]

The project is framed as a safety-first AI system for the Global South AI Safety Challenge Asia Track. Its relevance is strongest in multilingual and politically sensitive contexts, where factual

grounding, source attribution, and misinformation resilience matter more than fluent but unverified responses.[cite:493][cite:503]

2. Project Overview

PoliticAI is designed as a structured political intelligence interface rather than a generic chatbot. Instead of answering from model memory alone, it retrieves trusted and recent source material, assembles relevant evidence, and then generates a bounded answer linked to those sources.[cite:495][cite:497]

The platform is intended to support safer civic exploration by allowing users to query political news, election developments, fake-versus-real news claims, and governance issues while preserving traceability. In this design, the assistant should refuse unsupported answers when the retrieved evidence is insufficient, rather than speculate.[cite:495]

3. Problem Statement

Political information environments are noisy, fast-changing, and vulnerable to misinformation, especially during elections and other high-attention public events. In such settings, standard LLM chat systems may produce outdated, fabricated, or weakly supported claims if they are not constrained by retrieval and evidence-based response policies.[cite:495][cite:497]

This problem becomes more serious in multilingual and Global South contexts, where institutional coverage may be fragmented, local language content may be underrepresented, and region-specific harms from information pollution can be intensified. Recent Global South AI safety discussions have emphasized stronger accountability, literacy, fact-checking, and region-aware governance responses to information harms.[cite:493][cite:503]

4. Objectives

The project has five main objectives:

- Build a queryable knowledge layer over political and civic reporting using structured, source-linked records.
- Reduce hallucination by requiring retrieval evidence before answer generation.[cite:495][cite:497]
- Preserve citation and source traceability for every high-stakes response.
- Support multilingual and region-sensitive use cases relevant to Asia and the broader Global South.[cite:493][cite:503]
- Demonstrate a practical AI safety MVP that is feasible within a hackathon timeline.

5. Why This Matters in Asia

The Asia track of the challenge highlights AI safety issues such as multilingual model performance, misinformation, governance, and region-specific deployment risks. A system like PoliticAI directly fits that framing because it focuses on political information safety, evidence retrieval, and multilingual extensibility rather than open-ended political generation.[cite:503][cite:493]

This is especially relevant in India and neighboring contexts where users often need clear civic information, trusted election references, and safer ways to examine political claims. Election literacy and voter education are already major public-interest priorities, with the Election Commission of India promoting structured voter education and electoral literacy initiatives through SVEEP and related programs.[cite:494][cite:498][cite:504]

6. Proposed Solution

The proposed solution is a safety-first RAG application that ingests curated political and civic sources, stores them in structured form, indexes them for retrieval, and uses an LLM only after relevant evidence is retrieved. The model response is therefore not the primary knowledge source; the retrieved documents are.[cite:495][cite:497]

The system should support two broad user flows:

- **Exploration flow:** users ask about political developments, election procedures, governance events, or misinformation narratives and receive grounded, cited summaries.
- **Verification flow:** users test claims or ask for clarification on recent developments, and the system either returns evidence-backed answers or explicitly refuses if evidence is missing.[cite:495]

7. System Architecture

The MVP can be organized into four layers:

7.1 Ingestion Layer

This layer collects trusted inputs such as official civic information, election-related reference material, and curated political reporting. For India-facing civic modules, official Election Commission and SVEEP resources provide a strong foundation for voter education and electoral process information.[cite:494][cite:498][cite:504]

7.2 Knowledge Layer

Documents are normalized, chunked, tagged by region, topic, date, language, and source type, and stored in a searchable knowledge base. Structured metadata enables safer retrieval than raw article dumps because the system can filter by jurisdiction, recency, and source class before generation.[cite:495]

7.3 Retrieval and Reasoning Layer

When a user submits a query, the system retrieves the most relevant chunks, optionally reranks them, and then passes only that evidence into the model prompt. This retrieval-first pattern is central because RAG improves reliability only when evidence quality is high; weak or partial evidence can still produce unsupported outputs if not properly gated.[cite:495]

7.4 Interface Layer

The interface presents answers with linked sources, timestamps, and refusal messages when grounding is weak. This improves user trust because the answer becomes inspectable rather than opaque.[cite:495]

8. Data and Source Strategy

The project should avoid uncontrolled scraping for the MVP. A safer initial source policy is:

- Official civic and election sources, including ECI and SVEEP material.[cite:494][cite:498][cite:504]
- Curated and attributable recent political reporting.
- Region-tagged public-interest issue summaries.
- Structured internal records for issue tracking, misinformation examples, and source provenance.

This strategy matters because evidence quality determines answer quality. Research on RAG reliability shows that retrieval can improve accuracy with near-complete evidence, but partial or unsupported evidence can still lead to hallucination and citation fabrication.[cite:495]

9. RAG Pipeline Design

A practical MVP pipeline would follow these steps:

1. **Ingest** trusted political and civic content.
2. **Clean and chunk** documents into manageable evidence units.
3. **Attach metadata** such as date, geography, topic, source, and language.
4. **Index** for retrieval using lexical or hybrid search.
5. **Retrieve** the most relevant chunks for each query.
6. **Apply confidence rules** to decide whether evidence is sufficient.
7. **Generate** a response only from the retrieved context.
8. **Return citations** and refusal messages where needed.[cite:495][cite:497]

For a fast MVP, hybrid retrieval is preferable to purely generative answering because it gives more control over relevance and traceability. The answer template should explicitly instruct the model to use only provided evidence, cite the supporting items, and refuse if the context does not fully support the query.[cite:495]

10. Safety and Hallucination Controls

The main safety principle of PoliticAI is bounded generation. In practice, this means the model should not answer political questions unless supporting evidence has been retrieved and passed into context.[cite:495]

Recommended controls include:

- Strict source-bounded system prompts.
- Refusal when no strong context match is found.
- Citation of retrieved sources in every factual answer.
- Separation of trusted civic knowledge from volatile unsupported claims.
- Query logs for later audit and failure review.
- Optional multilingual expansion with the same retrieval-and-refusal logic.[cite:493]
[cite:495]

This is important because research shows RAG is not automatically safe; if the system retrieves incomplete or weak evidence, hallucinations can still occur. Therefore, retrieval quality checks and refusal thresholds are as important as the model itself.[cite:495]

11. MVP Features

The strongest hackathon MVP would include:

- A political query interface for recent issue lookup.
- Source-linked answer cards.
- Region- and topic-aware filtering.
- Election and civic explainer mode using official references.[cite:494][cite:504]
- Fake-versus-real news verification flow using curated evidence sets.
- Refusal engine for unsupported questions.
- Session history and basic analytics.

These features make the product useful while keeping the build focused on safety, retrieval, and explainability instead of broad social or prediction features.

12. Evaluation Plan

The system should be evaluated on groundedness rather than eloquence. A simple benchmark can include:

- In-scope factual queries with known evidence.
- Paraphrased versions of the same questions.
- Out-of-scope questions that should trigger refusal.
- Ambiguous questions with insufficient source support.
- Region-sensitive political prompts and misinformation-style claims.

Useful metrics include grounded answer rate, refusal precision, unsupported-claim rate, citation correctness, and retrieval hit rate. This evaluation style aligns with recent work emphasizing that evidence completeness is a key determinant of RAG reliability.[cite:495]

13. Expected Outcomes

If implemented well, the project should demonstrate that a political AI assistant can be made substantially safer by prioritizing retrieval, attribution, and refusal over free-form generation. It should also show that civic and political information tools for Asia can be built with stronger local relevance and safer operating boundaries.[cite:493][cite:503]

For hackathon judging, the strongest outcome is not “the model knows everything,” but “the system reliably distinguishes what it knows from what it cannot support.” That is a stronger AI safety story than a conventional political chatbot.[cite:495]

14. Limitations

The MVP will still have important limitations:

- Political news changes quickly, so source freshness is a constant challenge.
- Retrieval quality depends on how well the dataset is curated and tagged.[cite:495]
- Multilingual performance may vary if local-language sources are sparse.
- Bias can still enter through source selection, even with grounded retrieval.
- The system is safer for summarization and lookup than for interpretation of intent, motive, or future outcomes.

These limitations should be stated openly in the submission because transparency is itself part of responsible AI system design.

15. Future Scope

The next version of PoliticAI / ELECTOR Feed could expand in several directions:

- Indic-language and code-mixed retrieval support.
- Better source ranking and contradiction handling.
- Dedicated misinformation case-tracking dashboards.
- Region-by-region civic intelligence modules.
- Stronger evaluation suites for political safety and retrieval robustness.
- Human reviewer workflows for disputed or high-risk claims.

Future development should preserve the same core principle: retrieval-first, source-linked, refusal-capable political intelligence rather than unconstrained generative politics.

16. Conclusion

PoliticAI / ELECTOR Feed is a strong AI safety hackathon concept because it focuses on a real and difficult deployment problem: how to help users explore political information without encouraging hallucination, bias amplification, or unsupported claims. By combining structured source ingestion, retrieval-first answer generation, and refusal when evidence is weak, the project offers a practical and regionally relevant AI safety MVP for the Asia track. [cite:493][cite:495][cite:503]

Its value lies not only in political information access, but in showing how high-risk public-interest AI systems can be designed to be inspectable, evidence-aware, and safer by default. That design choice makes it both a usable product concept and a credible AI safety submission. [cite:495]