Hackathon Theme:
"MULTIMODAL RAG SYSTEM CHALLENGE"

This problem statement defines a single, unified hackathon challenge focused on the design and implementation of robust Multimodal Retrieval-Augmented Generation (RAG) systems. Participants are expected to demonstrate deep system design, multimodal reasoning, and reliability under uncertainty. This is not a prompt-engineering or chatbot challenge.

Problem Statement

CAUSE / AIM

In real-world environments, knowledge is fragmented across multiple modalities such as text documents, images, scanned files, audio recordings, and other unstructured sources. These modalities are often incomplete, noisy, misaligned, contradictory, or missing contextual metadata.

Most existing AI systems fail under such conditions, producing confident but ungrounded responses. This challenge aims to push participants to build a Multimodal Retrieval-Augmented Generation (RAG) system that can ingest, store, retrieve, reason, and generate responses grounded in evidence while explicitly acknowledging uncertainty, conflicts, and data gaps.

The goal is to engineer a system that behaves responsibly under ambiguity, adapts its retrieval strategy based on evidence quality, and knows when it should refuse to answer.

CORE OBJECTIVE

Design and implement a Multimodal RAG system capable of:
- Ingesting heterogeneous multimodal data
- Performing unified cross-modal retrieval
- Generating evidence-grounded responses
- Detecting conflicts and uncertainty
- Adapting retrieval and reasoning strategies dynamically

MANDATORY SYSTEM REQUIREMENTS

1. **Multimodal Data Ingestion:**
   The system must ingest and process multiple data formats including text (PDF, DOC, TXT),

images (scans, diagrams, photographs), and audio recordings. Each modality must be treated as a primary knowledge source.

2. **Unified Multimodal Storage:**
   All ingested data must be stored in a unified knowledge representation that enables semantic retrieval across modalities without isolating formats into silos.
3. **Cross-Modal Retrieval & Alignment**:
   The system must retrieve relevant information across different modalities for a single query and infer relationships even when explicit metadata is missing or unreliable.
4. **Intent-Aware Retrieval Strategy:**
   The system must analyze the user query to determine which modalities are relevant and adjust retrieval depth and scope accordingly.
5. **Evidence-Based Generation (RAG):**
   All generated responses must be grounded in retrieved evidence. The system must clearly reference the sources used for generation.
6. **Uncertainty & Confidence Awareness:**
   The system must assess the strength of available evidence, distinguish between verified facts and inferred assumptions, and explicitly acknowledge uncertainty where applicable.
7. **Conflict Detection & Presentation:**
   When retrieved sources contradict each other, the system must detect the conflict and present multiple interpretations instead of collapsing them into a single answer.
8. **Adaptive Retrieval Loop:**
   If retrieved evidence is insufficient or low-confidence, the system must trigger additional retrieval cycles and justify why further retrieval was required.
9. **Hallucination Suppression:**
   When adequate evidence is unavailable, the system must refuse to answer and clearly state what information is missing.
10. **Failure-Tolerant Operation:**
    The system must continue to operate when some modalities are unavailable, embeddings are partially corrupted, or external services fail.

CONSTRAINTS

- The solution must not rely solely on a **single cloud-based LLM and must not use heavy resource.**
- Blind vector search without justification will be penalized.
- Systems that generate answers without citing retrieved evidence will be disqualified.
- Monolithic, non-explainable pipelines are discouraged.

EXPECTED DELIVERABLES

- Working prototype demonstrating multimodal ingestion, retrieval, and generation
- System architecture diagram
- Explanation of retrieval and reasoning strategy
- Demonstration of failure and uncertainty handling

EVALUATION CRITERIA

- Cross-modal retrieval accuracy and relevance
- Quality of evidence grounding and citations
- Effectiveness of conflict and uncertainty handling
- System robustness and adaptability
- Overall engineering depth and clarity

NOTE TO PARTICIPANTS:

Judges may introduce adversarial inputs such as missing data, contradictory sources, or partial modality failures during evaluation. Solutions that prioritize correctness, transparency, and system integrity over superficial output quality will be favoured.