FACULTY OF TECHNOLOGY
Information & Communication Technology
Subject: Capstone Project

Name: Fenil Vadher

En Roll no: 92200133023

Subject: Capstone Project

Stakeholder Identification and Needs Analysis

For this capstone project, the primary domain selected is **Artificial Intelligence (AI) in Multimodal Information Retrieval**. The stakeholders include:

1. **Content Consumers (Movie Enthusiasts, Researchers, and Students):**
    - Need an efficient system to **search and retrieve movie dialogues and scenes** across large databases.
    - Current platforms (e.g., IMDb, streaming platforms) lack **fine-grained multimodal search** (dialogue + scene context + visual cues).
    - Consumers often face **time-consuming manual browsing** to find specific moments or quotes.
2. **Media and Entertainment Companies:**
    - Require tools for **metadata enrichment** and **content recommendation**.
    - Challenges include **manual annotation costs** and **scalability** of indexing multimodal content.
3. **Academic and Research Communities:**
    - Scholars analyzing **storytelling patterns, dialogue sentiment, or character interaction** need structured retrieval.
    - Current methods lack **context-aware multimodal indexing** across text, audio, and visuals.

4. **Technology Developers (AI/ML Engineers, Cloud Service Providers):**
    - Need frameworks that are **scalable, interoperable, and cost-effective** for deployment.
    - Challenges include **high compute requirements**, **lack of standardized evaluation**, and **data diversity**.

Insights from Market and Industry Reports:

- The global video analytics market is expected to grow at a CAGR of 21.5% by 2030, driven by AI-based search and recommendation (MarketsandMarkets, 2024).
- A Deloitte study (2023) highlights that **71% of media companies struggle with data silos and inefficient content retrieval systems**.
- A PwC survey (2023) indicates **61% of consumers demand smarter search and personalization** in entertainment platforms.

These findings confirm that **stakeholders demand a robust, AI-driven solution for multimodal script and scene retrieval**.

Problem Statement

Current multimedia search engines are limited to **keyword-based retrieval** or **basic metadata filtering**, which fails to capture the **contextual and multimodal nature of movie scripts and scenes**. As a result, stakeholders (audiences, researchers, and companies) face:

- **Inefficient search experiences** (dialogue-only or scene-only search).
- **High costs of manual annotation** for companies.
- **Lack of standardization** in evaluating multimodal retrieval performance.

Problem Statement:
There is a critical need for an intelligent, context-aware multimodal search system that integrates textual, visual, and audio modalities to enable precise dialogue and scene retrieval in large-scale movie databases.

Ideation of Solutions

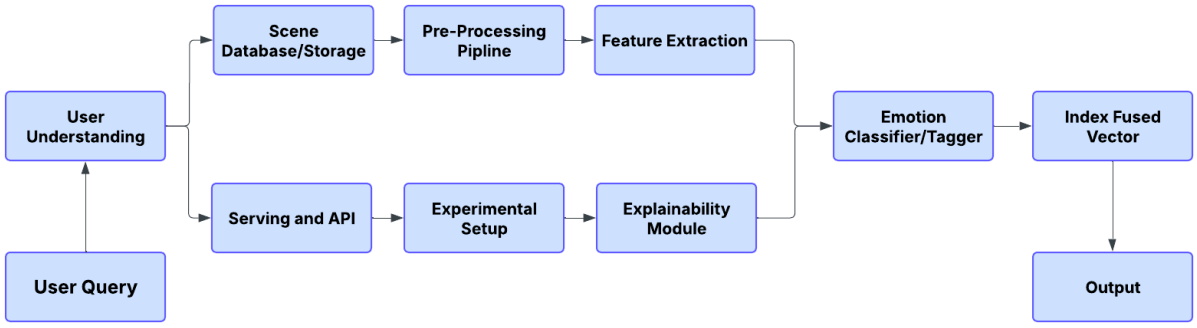
To address the above problem, the following innovative solution ideas are proposed:

Idea 1: Multimodal Embedding Search Engine (MMESE)

- **Description:** Develop a **deep learning-based retrieval engine** that encodes scripts, dialogues, and visual frames into a shared embedding space.
- **Features:**
    - Text embeddings (dialogues, descriptions).
    - Visual embeddings (scene frames).
    - Audio embeddings (tone, background sounds).
    - Unified retrieval mechanism.
- **Stakeholder Benefit:** Enables **context-aware search** where users can query by text, image, or even example dialogue.

Idea 2: AI-Powered Contextual Recommendation System

- **Description:** Extend retrieval to **personalized content recommendations** by analyzing search patterns and preferences.
- **Features:**
    - Hybrid recommendation (collaborative + content-based filtering).
    - Semantic understanding of queries.
    - Adaptive learning from user feedback.
- **Stakeholder Benefit:** Enhances **user engagement** for consumers and **business revenue** for media companies.



Relevance to ICT Domain

This project is highly relevant to the ICT domain:

- **Artificial Intelligence (AI/ML):** Utilizes **transformer-based models** (e.g., BLIP-2, Vid2Seq, GIT2) for multimodal embeddings and retrieval.
- **Information Retrieval (IR):** Advances beyond traditional IR by incorporating **multimodal context awareness**.
- **Current Trends Alignment:**
    - **Generative AI and Large Multimodal Models (LMMs)** (OpenAI, Google, Meta) are being deployed for advanced content understanding.
    - **Edge AI and 5G streaming** highlight the need for efficient, real-time retrieval.

Impact on Stakeholders:

- **Consumers:** Faster and more engaging search experience.
- **Companies:** Reduced costs in metadata management and improved monetization.
- **Researchers:** Richer data for academic analysis of storytelling and multimedia content.
- **Developers:** A standardized, AI-powered platform for innovation.