Phase 2 Design Document - Intermediate Projects

# 1. Introduction

This document describes the design and architecture of the Phase 2 Intermediate Projects developed as part of the LLM Awareness and Portfolio Development Program. Phase 2 consists of three main projects: Global News Topic Tracker, Multi-Modal Assistant, and Meeting Notes & Action Item Extractor. Each project leverages LLaMA3 and other AI models to demonstrate applied Large Language Model (LLM) capabilities.

# 2. Architecture Overview

The projects are built using a modular design with Streamlit as the user interface. Supporting models and APIs include LLaMA3 (via Ollama) for reasoning, Hugging Face BLIP for image captioning, and Whisper for audio transcription. The backend logic is written in Python, organized into separate modules.

Key Components:

* - LLaMA3 (via Ollama): Text reasoning and summarization
* - BLIP: Image captioning for visual understanding
* - Whisper: Speech-to-text transcription for meeting audio
* - Streamlit: Web-based user interface

# 3. Project File Structure

phase2\_Intermediate\_Projects/  
 ├── app.py # Main Streamlit UI with tabs  
 ├── news\_tracker.py # News Tracker logic  
 ├── multimodal\_assistant.py # Multi-Modal Assistant logic  
 ├── meeting\_notes.py # Meeting Notes logic  
 ├── utils/  
 │ ├── llama3\_utils.py # Query LLaMA3 functions  
 │ └── whisper\_utils.py # Audio transcription functions  
 └── requirements.txt # Dependencies

# 4. Global News Topic Tracker

The Global News Topic Tracker fetches trending headlines using RSS feeds (Google News) and summarizes the top topics using LLaMA3. Users can input a topic of interest, fetch the latest headlines, and get a concise summary.

Sample Input: 'Artificial Intelligence'  
Sample Output: 'AI is being widely adopted across industries. Recent news covers regulation, enterprise adoption, and generative AI developments.'

# 5. Multi-Modal Assistant

The Multi-Modal Assistant allows users to upload an image and ask questions about it. The BLIP model generates a descriptive caption of the image, which is then passed to LLaMA3 to answer user-specific queries. This demonstrates visual + text reasoning.

Sample Input: Image of a laptop, coffee cup, and books.  
Sample Question: 'What objects are in this image?'  
Sample Output: 'The image contains a laptop, a coffee cup, and books on a desk.'

# 6. Meeting Notes & Action Item Extractor

The Meeting Notes Extractor converts uploaded meeting audio into structured notes and action items. It uses Whisper to transcribe speech into text and LLaMA3 to summarize and extract action points. Supports .wav, .mp3, and .m4a audio formats.

Sample Input: Audio recording with instructions to complete tasks.  
Sample Output:  
- Meeting Notes: Sprint planning meeting. Backend API integration due Friday. Payment gateway assigned to John. UI design to Priya by Wednesday.  
- Action Items: John → payment gateway, Priya → UI design, Team → review on Thursday.

# 7. Conclusion

Phase 2 Intermediate Projects extend the portfolio with more advanced applications of LLMs. Together, they demonstrate multimodal capabilities (text, image, audio), integration of third-party models, and end-to-end user workflows. These projects provide practical value in summarization, productivity, and multi-modal assistance.