Audio2Art: Project Planning & Scheduling

# 1. Project Title

Audio2Art: Transforming Voice Prompts Into Visual Creations Using Transformers

# 2. Objective

To develop an AI-powered web application that converts spoken audio prompts into visually appealing images using transformer models like Wav2Vec2 for transcription and Stable Diffusion for image generation.

# 3. Scope

This project involves:

- Converting audio to text using Wav2Vec2

- Generating an image from transcribed text using Stable Diffusion

- Hosting the app with Streamlit and deploying via LocalTunnel

- Making the interface user-friendly and accessible

# 4. Timeline & Milestones

Week 1: Idea selection, requirement gathering, research

Week 2: Tool selection and environment setup (Google Colab)

Week 3: Implementation of promptgen (Wav2Vec2 logic)

Week 4: Implementation of text2image (Stable Diffusion logic)

Week 5: UI development using Streamlit

Week 6: Integration and testing

Week 7: Debugging, deployment with LocalTunnel

Week 8: Documentation (design, testing, planning, ideation) and presentation

# 5. Resources Required

- Google Colab (with GPU)

- Python packages: streamlit, diffusers, accelerate, transformers, torch, librosa

- LocalTunnel for temporary deployment

- Audio files (.wav format) for testing

- Internet connection

# 6. Team & Roles (if any)

- Developer: Implements models and frontend

- Tester (if applicable): Validates output and identifies bugs

- Reviewer: Compiles and prepares documentation

# 7. Risks and Contingencies

- Risk: Audio transcription inaccuracies → Mitigation: Use clear recordings, validate output

- Risk: GPU unavailability on Colab → Mitigation: Optimize inference or use smaller models

- Risk: Internet dependency for deployment → Mitigation: Document static alternatives

# 8. Deliverables

- Functional web app

- Image generation pipeline

- User interface

- Documentation: design, planning, testing, ideation