Hackathon Project Phases for the **Audio2Art**.

Hackathon Project Phases

Project Title:

Audio2Art: Transforming Voice prompts into visual creations using transformers

Team Name:

VoxCanvas

Team Members:

- Boda Rishika
- Akula Archana
- Appam Chandana
- Annam Hasini

Phase-1: Brainstorming & Ideation

Objective:

Develop an **AI-powered creative tool** that utilizes **transformers** to convert **voice prompts into visual art**, enabling users to generate images through speech-based interaction.

Key Points:

1. **Problem Statement:**

- Many users lack the artistic skills to create digital artwork but have creative ideas they want to express.
- Traditional AI art generators require **text-based prompts**, limiting accessibility for those who prefer or require voice-based interactions.

2. **Proposed Solution:**

- A voice-to-art AI application that processes spoken descriptions, extracts key features, and translates them into AI-generated visuals.
- Uses transformers to analyze speech input, generate text-based prompts, and create artwork based on interpreted content.
- Provides different artistic styles and customization options for diverse creative outputs.

3. **Target Users:**

- Artists & Designers looking for quick inspiration or concept visualization.
- Casual users who want to generate artwork effortlessly.
- Accessibility-focused users, such as those with mobility impairments, who benefit from hands-free interaction.

4. **Expected Outcome:**

- A functional voice-to-art generation tool that enables users to create visual artwork from speech.
- A seamless user experience with intuitive voice input processing and real-time image generation.

Phase-2: Requirement Analysis

Objective:

Define the **technical and functional** requirements for Audio2Art.

Key Points:

Technical Requirements:

- **Programming Language:** Python
- **Backend:** Transformer-based Speech-to-Text & Text-to-Image models
- **Frontend:** Streamlit or a web-based UI framework
- **Database:** Not required initially (API-based queries)

1. **Functional Requirements:**

- Convert **spoken prompts** into **artistic images**.
- Support **multiple art styles** (e.g., realism, abstract, anime, sketch).
- Provide a **user-friendly voice interface** for prompt submission.
- Display **real-time AI-generated images** with customization options.

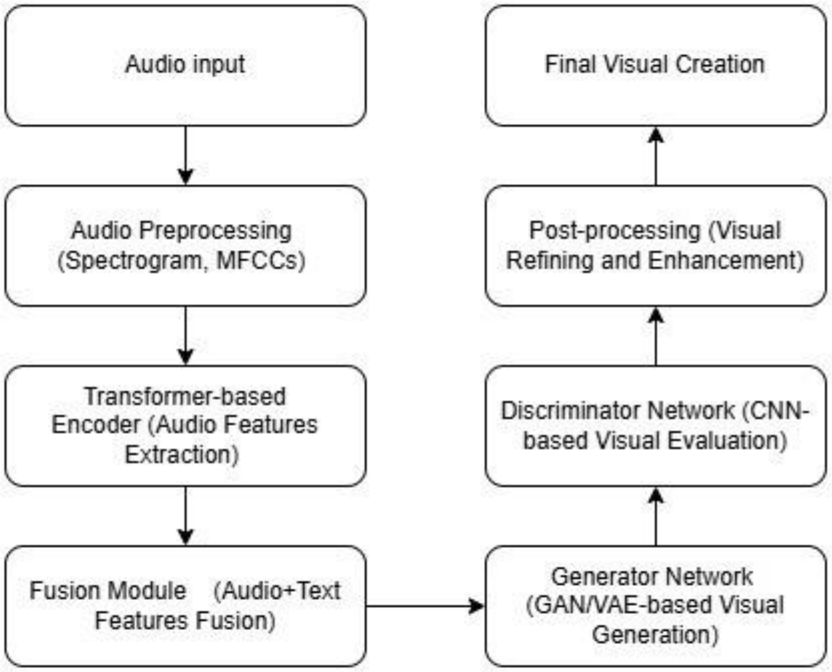
2. **Constraints & Challenges:**

- Ensuring **accurate speech-to-text conversion** to interpret prompts correctly.
- Handling **AI-generated image accuracy** based on voice descriptions.
- Managing **computation time & GPU requirements** for real-time image generation.

Phase-3: Project Design

Objective:

Develop the architecture and user flow of the application.


Audio input
Audio Preprocessing (Spectrogram, MFCCs)
Transformer-based Encoder (Audio Features Extraction)
Fusion Module (Audio+Text Features Fusion)
Generator Network (GAN/VAE-based Visual Generation)
Discriminator Network (CNN-based Visual Evaluation)
Post-processing (Visual Refining and Enhancement)
Final Visual Creation

Key Points:

System Architecture:

- **User speaks a creative prompt** into the app.
- The **Speech-to-Text transformer** converts the voice input into text.
- The **Text-to-Image transformer** processes the description and generates an image.
- The **frontend displays** the AI-generated artwork with **options for adjustments**.

User Flow:

- **Step 1:** User speaks a prompt (e.g., "A futuristic city at sunset in cyberpunk style").
- **Step 2:** The backend processes the voice input and converts it into a text description.
- **Step 3:** The AI model generates an image based on the processed description**.**
- **Step 4:** The user can fine-tune the output (e.g., change colours, adjust style).

UI/UX Considerations:

- A **minimalist, user-friendly interface** for smooth interaction.
- **Live voice input feature** with **real-time text conversion feedback**.
- Options to **adjust image settings** (brightness, style, colour scheme).

Phase-4: Project Planning (Agile Methodologies)

Objective:

Break down development tasks for efficient completion.

Sprint	Task	Priority	Duration	Deadline	Assigned To	Dependencies	Expected Outcome
Sprint 1	Environment Setup & API Integration	□ High	6 hours (Day 1)	End of Day 1	Boda Rishika	Transformer models	AI models successfully integrated
Sprint 1	Frontend UI Development	□ Medium	3 hours (Day 1)	End of Day 1	Akula Archana	UI framework setup	Basic UI with voice input field
Sprint 2	Voice Processing & Text Conversion	□ High	4 hours (Day 2)	Mid-Day 2	Appam Chandana	Speech-to-Text API	Accurate text extraction from voice input
Sprint 2	Image Generation & Fine-tuning	□ High	4 hours (Day 2)	Mid-Day 2	Annam Hasini	Text-to-Image Model	AI-generated images based on prompts
Sprint 3	UI Enhancements & Debugging	□ Medium	2 hours (Day 2)	Mid-Day 2	Akula Archana & Appam Chandana	UI layout complete	Smooth user interaction & experience
Sprint 3	Testing & Deployment	□ Low	2 hour (Day 2)	End of Day 2	Entire Team	Working prototype	Ready for final demo submission

Sprint Planning with Priorities

Sprint 1 – Setup & Integration (Day 1)

- ● **High Priority:** Set up the environment & install dependencies.
- ● **High Priority:** Integrate Speech-to-Text and Text-to-Image transformer models.
- □ **Medium Priority:** Build a basic UI with voice input functionality.

Sprint 2 – Core Features & Debugging (Day 2)

- ● **High Priority:** Implement voice processing and text conversion.
- ● **High Priority:** Generate images based on transformed text input.
- ● **High Priority:** Debug AI response accuracy and ensure seamless integration.

Sprint 3 – Testing, Enhancements & Submission (Day 2)

- □ **Medium Priority:** Test AI-generated images for different voice prompts.
- □ **Medium Priority:** Improve UI responsiveness and user experience.
- □ **Low Priority:** Prepare the final demo, optimize performance, and deploy.

Phase-5: Project Development

Objective:

Implement core features of **Audio2Art**.

Key Points:

Technology Stack Used:

- **Frontend:** Streamlit / Web UI
- **Backend:** Speech-to-Text & Text-to-Image transformers
- **Programming Language:** Python

Development Process:

- Implement Speech-to-Text conversion for prompt input.
- Integrate Text-to-Image models for generating artwork.
- Optimize user interaction for a smooth voice-based experience**.**

Challenges & Fixes:

- **Challenge:** Background noise affects speech recognition.
  - **Fix:** Use noise-reduction techniques and AI filters.
- **Challenge:** Long prompts might reduce image coherence.
  - **Fix:** Optimize text processing to summarize key elements**.**

Phase-6: Functional & Performance Testing

Objective:

Ensure that **Audio2Art** functions correctly.

Test Case ID	Category	Test Scenario	Expected Outcome	Status	Tester
TC-001	Functional Testing	Speak "A sunset over a calm ocean"	AI generates a relevant image.	✅Passed	Tester 1
TC-002	Functional Testing	Speak "An astronaut on Mars in watercolor style"	AI applies the correct art style	✅Passed	Tester 2
TC-003	Performance Testing	Voice processing response under 2 sec	Voice-to-text conversion is smooth	⚠ Needs Optimization	Tester 3
TC-004	Bug Fixes & Improvements	Fix unclear text interpretations	Text accuracy is improved	✅Fixed	Developer
TC-005	Final Validation	UI is responsive across devices	Works on desktop & mobile	❌Failed - UI broken on mobile	Tester 2
TC-006	Deployment Testing	Host the app on Streamlit	Accessible online	□ Deployed	DevOps

Final Submission

☑ **Project Report (Based on this template)**

☑ **Demo Video (3-5 Minutes)**

☑ **GitHub/Code Repository Link**

☑ **Presentation**