Cover Gen

Pushkar | Karan | Ayush | Sreehari

INTRODUCTION

This project aims to generate images from audio or text provided by the user as an application for the latest OpenAI Tools in the order of Whisper, ChatGPT, and DALL-E. This caters to creating cover photos and pictures for podcast albums, story books, and documents.

METHODOLOGY

- The pipeline developed consists of converting audio into transcripts using OpenAI Whisper. Further, that transcript is transformed into a prompt using GPT-2 and then the image using DALL-E.
- The pipeline involves multi models for each process and has been fine-tuned for each task. These models are mainly developed by OpenAI, which has been further applied to this use.
- The text input is also addressed by summarising it to create a prompt that further makes images using stable diffusion.

RESULTS

The multi-model setup has been tested for various domains. It has been fine-tuned at the prompt and summary step for each task. Namely story-books covers, podcast covers, meeting covers, and document/book covers. Negative prompting and advanced models were also used for stable diffusion.

PROCEDURE

- The image gets generated by following the pipeline when uploading an audio file or text from a book, storybook, or document.
- The user has to upload a document or podcast on which the album cover image has to be obtained.
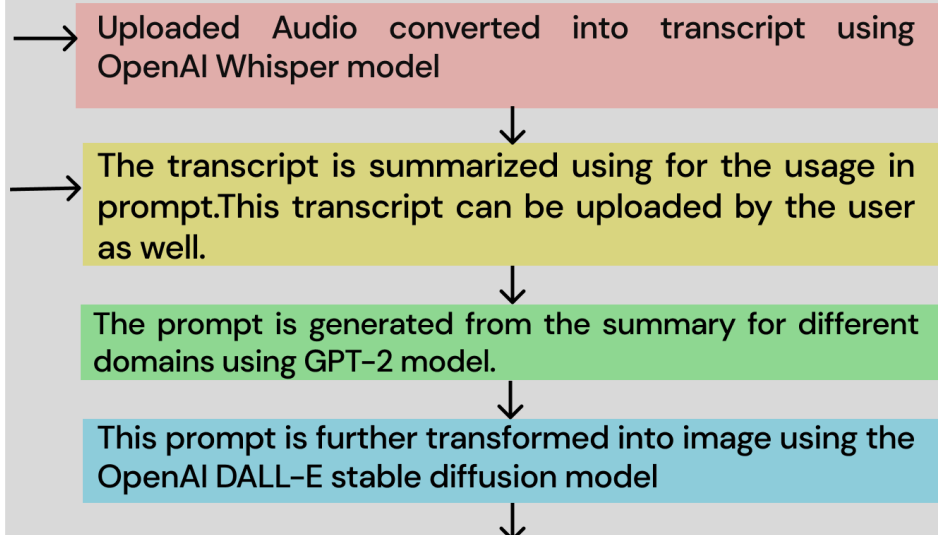
FUTURE GOALS

We propose this pipeline to work on making more application based products such as story-book, podcasts and Books covers. The Music industry could be targeted by converting song into lyrics.

PIPELINE

Uploaded Audio converted into transcript using OpenAI Whisper model

↓

The transcript is summarized using for the usage in prompt.This transcript can be uploaded by the user as well.

↓

The prompt is generated from the summary for different domains using GPT-2 model.

↓

This prompt is further transformed into image using the OpenAI DALL-E stable diffusion model

↓

The image obtained for the audio input of a meeting