2024 AI Challenge – VoicedOver – Lazy Devs

*Talk your way through anything.*

## Problem

In today's fast-paced world, speed and efficiency are paramount, yet current technologies often fall short in delivering instant results in digital interactions. Users still encounter delays and inefficiencies across various digital platforms, whether accessing information, managing tasks, or communicating. This gap highlights a growing demand for quicker, more seamless processes. Our project addresses this need by introducing an innovative solution that enhances user interaction speeds with digital platforms, thereby advancing our global obsession with achieving ever-faster technological responses.

## Solution

VoicedOver adds another string to this endless pursuit, by allowing individuals to seamlessly communicate with their devices without even touching them. The VoicedOver gives any type of application the ability to talk to the user. The result is individuals can perform applications functions more easily and quickly.

## Target Audience

Currently, our target audience is the e-commerce stores, which is going to expand in the future.

## Engineering

### *High-Level Design*

## VoicedOver is an API tool, which is going to become a micro-service, that accepts user commands via audio input, utilizing Generative AI to execute tasks within integrated applications. These tasks are carried out automatically, triggering actions on the application's interface upon successful completion of VoicedOver API calls.

### *How It Works*

Following steps are listed down to understand its workflow:

1. Audio from the user is taken as input
2. Input is passed to LLM, OpenAi whisper in our case, which converts audio to text
3. This text is forwarded to OpenAi assistant. Function Calling is used to create response according to the instructions.
4. Response created by function calling not only prompt our machine to perform the said function, but the message part is also converted into Audio by means of another LLM, microsoft speecht5 in our case, which in turn, is delivered to the user.

### *Role of Generative AI*

Generative AI being used in three stages. The first stage of transforming audio to text is done via OpenAI whisper. The second stage of sending the returned text to the assistant for function calling. Once the function has been called and a message has been received, the message is then sent to another LLM for audio conversion. This audio is then finally delivered to the user.

### *Technologies*

* Python 3.12
* OpenAI whisper-1
* OpenAI assistants gpt-3.5-turbo-0125
* OpenAI tts-1/Microsoft speecht5\_tts
* Fast API 0.110.2

## Results and Analysis

We innitiated 100 tests and here are the results:

1. Audio in and conversion to text → 90% to the mark

2. Text sent to OpenAi assistant and returned a functional call successfully → 70%

3. Response from OpenAi conversion to speech → 90%