

Fiche descriptive de projet Hub

MY_ELIZA

Contexte et but du projet

Ayant obtenu un stage à l'Université de recherche à Strasbourg dans la team Icube spécialisé dans l'IA, mon futur tuteur m'a fait part de certaines technologies que j'utiliserais durant mon stage (SV2TTS, DeepFaceLab,...). C'est pour cela que j'ai décidé de me remettre au python, et d'expérimenter avec le repo SV2TTS en projet HUB.

L'idée serait de créer un assistant vocal de type Home Assistant en utilisant différentes API comme Whisper pour la partie Speech Recognition, pour un prototype, utilisé une API GPT 3.5 et expérimenté avec SV2TTS pour la partie voice-cloning / TTS. Dans l'objectif de faire autant de travail que possible dessus, pour un potentiel usage personnel et en guise de préparation pour mon stage, je pense remplir 70 heures de travail officiellement, et plus en pratique, peut-être même en continuant le projet par la suite. (Potentiel pour faire un mode offline en utilisant des models comme LLAMA-2 en local et une solution de Speech recognition qui tournerais en local dans le but de faire un mode en local. Potentiellement, par la suite je pourrais encore séparer le projet en un client et un server-side, où le serveur ferait tout le processing, et le client ne ferait que microphone, haut-parleur, et API/custom server call.

Le nom du projet viens de la toute première machine à effectuer des opérations sur langage (NLP) dans les 1960 à l'institut MIT.

Résumé du projet

Mettre sous forme de liste à puces les étapes du projet

- Implémenter le Speech-To-Text avec l'API Whisper.
- Implémenter les call d'API pour un LLM (potentiellement GPT 3 / 3.5 ou davinci)
- Implémenter le Text-To-Speech en utilisant le repo SV2TTS.
- Expérimenter avec du voice-cloning
- Implémenter un système de contexte pour que l'assistant garde un sens de la conversation actuel car les call API sont toujours uniques contrairement à une conversation.
- Implémenter un model custom dans SV2TTS
- Fine-Tune SV2TTS pour avoir un résultat plus poli.

Porteur(s) du projet

- Responsable du projet : Cyrian Pittaluga

Environnement technique / technologique

Exposer le contexte technique et technologique (matériel, langage, environnement d'exécution, ressources, ...) dans lequel le projet s'inscrit

Matériel : Desktop / Laptop / Potentiellement un server dédié

Langages : Python

Environnement d'exécution : GNU/Linux Fedora / Ubuntu server

Ressources : Documentation variée et potentiel assistance d'un chercheur dans le domaine

Description du livrable

- Speech-To-Text avec Whisper : Déployé avec documentation.
- Call d'API pour LLM : Déploiement avec documentation.
- Text-To-Speech avec SV2TTS : Déployé avec documentation.
- Expérimentation avec du voice-cloning : Documentée.
- Système de contexte : Déployé avec documentation.
- Model custom dans SV2TTS : Déployé avec documentation.
- Fine-Tuning de SV2TTS : Déployé avec documentation.

Organisation et temporalité

1 – Tâche 1

Explications de la tâche

Prototype d'assistant basic,
Speech-To-Text API > LLM API > Text_To_Speech API

Temps estimé pour cette tâche:

20 H

2 – Tâche 2

Explications de la tâche

Objectif ; Obtenir des réponse rapide et décente vocalement.
Speech-To-Text API > LLM API > SV2TTS

Temps estimé pour cette tâche:

20 H

2 – Tâche 3

Explication de cette tâche

Objectif ; Créer et implémenter un model custom SV2TTS dans le projet de façon

Temps estimé pour cette tâche:

30 H

Estimation du projet

Temps de travail, plus de 70 heures.

Nous estimons à 20 **XP** le projet, ce qui fait 20 **XP** par personne.

- **Pourquoi ?**

J'ai déjà eu des expériences avec certains outils similaires, et plus simples d'utilisation, et d'expérience, produire un tel projet prendras nettement plus de temps car ce projet a pour but d'accumuler autant d'expérience que possible avec l'utilisation du repo SV2TTS. Donc, j'utiliserais tout mon temps libre pour ce projet jusqu'au début de mon stage dans le but de gagner en efficacité le jour venu. De plus, je vais devoir apprendre à créer de la documentation et le faire extensivement.