

Traducteur Vocal en Temps Réel à Base d'IA Générative

Contexte

Dans le cadre du Hackathon Google Cloud, l'objectif est de concevoir un traducteur vocal en temps réel utilisant des technologies avancées d'intelligence artificielle.

Ce projet s'appuiera sur Google Cloud et ses outils (comme Vertex AI et les APIs Cloud AI) pour implémenter un système innovant et fonctionnel qui traduit la voix en différentes langues avec précision.

Objectifs du Projet

Objectifs Principaux :

- Développer une application capable de traduire la voix en temps réel en plusieurs langues.
- Exploiter les modèles d'IA existants sur Vertex AI pour minimiser le temps de développement.
- Mettre en œuvre les technologies suivantes :
 - **Speech-to-Text** (reconnaissance vocale).
 - **Text-to-Text** (traduction textuelle).
 - **Text-to-Speech** (synthèse vocale).

Objectifs Secondaires :

- Intégrer un serveur de communication pour transmettre les traductions vers des périphériques d'écoute externes.
- Fournir un retour éducatif pour les utilisateurs souhaitant améliorer leur diction ou apprendre une nouvelle langue.

Innovations :

- Intégration avec des modèles génératifs comme Gemini pour enrichir la traduction en ajoutant du contexte.
- Déploiement d'une interface utilisateur (UI) intuitive et attrayante.

Attendus Fonctionnels

Reconnaissance et traduction en temps réel avec une latence minimale.

Flexibilité linguistique : permet à l'utilisateur de choisir les langues d'entrée et de sortie.

Interface utilisateur (UI) : simple, offrant un retour visuel des traductions en temps réel.

Capacités éducatives :

- Suggestions pour améliorer la diction.
- Indications grammaticales dans la langue traduite.

Interopérabilité : le système doit être compatible avec des périphériques d'écoute variés.

Exigences Techniques

Composants Logiciels

Google Cloud Services :

- Speech-to-Text API.
- Translate API.
- Text-to-Speech API.

Langages :

- Python (intégration des APIs Google).
- HTML/CSS/JavaScript (pour l'interface utilisateur).

Bibliothèques :

- sounddevice pour la gestion des flux audio.
- queue pour la synchronisation des étapes de traitement.

Exigences en Performance :

- Temps de traitement total (voix → texte → traduction → voix) inférieur à 2 secondes.
- Gestion de flux audio stable, même en présence d'interruptions ou de silences.

Architecture :

- Un système basé sur un pipeline parallèle pour gérer simultanément les différentes étapes (capturer → transcrire → traduire → synthétiser).
- Possibilité de redémarrer automatiquement en cas de détection de silence.

Paramètres Modifiables :

- Langue d'entrée et de sortie.
- Temps limite de détection de silence.

Scalabilité :

- Optimisation pour fonctionner sur des environnements contraints (ordinateurs portables standards ou serveurs cloud).
- Capacité à étendre les fonctionnalités pour inclure d'autres langues ou dialectes.

Exigences Non-Fonctionnelles

Sécurité :

- Les flux audio ne doivent pas être enregistrés ou partagés sans consentement explicite.
- Les données doivent être traitées dans un environnement sécurisé Google Cloud.

Accessibilité :

- Interface utilisateur accessible pour les personnes ayant des déficiences visuelles ou auditives.

Robustesse :

- Résistance aux erreurs de réseau ou aux coupures d'entrée audio.