

Projet Text-to-Speech (TTS)

Dans le cadre de mon projet, j'ai développé une application de machine learning dédiée à l'étude et à l'implémentation d'un système **Text-to-Speech (TTS)** en utilisant le modèle **Bark** de Suno.

Présentation du Projet :

- **Objectif** : Transformer du texte en audio naturel, multilingue et réaliste, capable de générer non seulement de la parole mais aussi des bruits de fond et des effets sonores simples.
- **Technologie utilisée** :
 - Le modèle Bark, basé sur des transformeurs, est au cœur de l'application. Il supporte plusieurs langues et dispose de préréglages vocaux personnalisables.
 - L'implémentation repose sur les bibliothèques **Hugging Face transformers** et **PyTorch**.

Mes Objectifs :

1. **Création de synthèse vocale avancée** :
 - Fournir une solution capable de générer un audio de haute qualité à partir de texte, dans plusieurs langues.
 - Répondre à des besoins divers tels que l'accessibilité pour les personnes malvoyantes, les assistants vocaux ou encore la création de contenu audio.
2. **Recherche et optimisation** :
 - Étudier les capacités et les limites des modèles transformeurs dans la génération audio.
 - Améliorer les temps d'inférence et développer des techniques de post-traitement pour garantir une qualité optimale.
3. **Développement pratique** :
 - Proposer un outil pratique et accessible pour créer du contenu audio multilingue.
 - Permettre une large gamme d'applications, allant des ressources éducatives aux loisirs.

Caractéristiques de mon Projet :

1. Traitement phrase par phrase :

- Chaque phrase est traitée indépendamment afin d'assurer une meilleure clarté et de réduire les erreurs ainsi que le bruit.
- Une attention particulière est portée aux segments bruyants situés au début ou à la fin de l'audio, grâce à une analyse basée sur des seuils d'amplitude.

2. Post-traitement audio :

- J'ai implémenté un algorithme de découpage qui nettoie les segments audio bruyants, notamment ceux générés à la fin des fichiers.
- Cette étape garantit une amélioration significative de la qualité audio finale.

3. Utilisation de Bark :

- Le modèle Bark, avec ses **1 milliard de paramètres**, est utilisé pour la conversion texte-à-audio.
- Des paramètres ajustables, tels que les seuils d'amplitude et les étapes temporelles, permettent d'optimiser les résultats.

4. Visualisation et analyse :

- Des outils de visualisation des ondes audio sont intégrés pour faciliter l'identification des bruits et affiner les réglages de nettoyage.