

Les Clonage de voix des Intelligences artificielles



1) Qu'est-ce qu'une IA ?

L'intelligence artificielle (IA) est un processus d'imitation de l'intelligence humaine qui repose sur la création et l'application d'algorithmes exécutés dans un environnement informatique dynamique. Son but est de permettre à des ordinateurs de penser et d'agir comme des êtres humains. Pour Cela on utilise trois composants nécessaires : des systèmes informatiques, des données avec des systèmes de gestion et des algorithmes D'IA avancé. Pour se rapprocher le plus du comportement humain l'IA a besoin d'une multitude de données et d'une capacité à les traiter. Nous allons nous intéresser au de voix de l'intelligence artificielle.

2) Qu'est-ce qu'un clonage de voix de L'IA ?

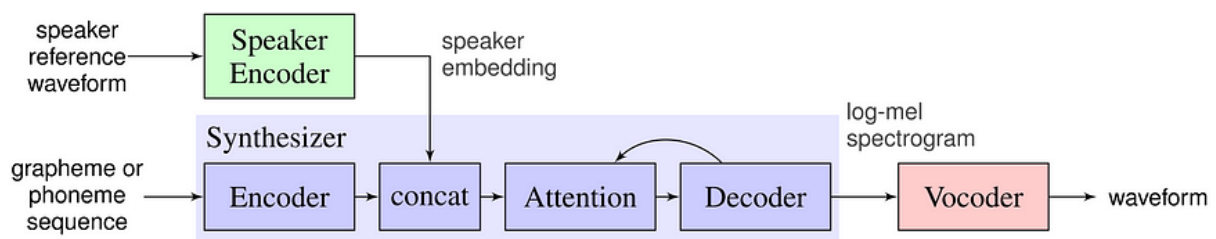
Le clonage de voix par l'IA est le fait d'utiliser une voix existante et de lui faire dire ce que bon nous semble. Il suffit de prendre un extrait de la voix de la personne que l'on veut et de la mettre sur un site (Deep Voice) dédié à cela puis de mettre le texte que l'on veut. Cela permettra à l'IA d'identifier et d'utiliser la voix puis de traiter le texte afin de lier les deux et de nous sortir la phrase demandée avec la voix demandée.

3) Comment fonctionne l'algorithme ?

Les IA et les algorithmes de clonage de voix sont composés de petite IA qui communiquent entre elles. Pour cela et pour convertir le texte en sons les IA ont besoin d'un synthétiseur celui-ci va découper le texte en une séquence phonétique. Il faut donc entraîner l'IA avec pleins de données, des textes qui seront d'un côté écrit et de l'autre lu pour qu'il apprenne la phonétique. Avec cela l'IA sera capable de parler avec un texte donné et une voix donnée mais sans la tonalité, les émotions, les accents.

4) Comment les IA de clonage de voix s'adaptent-elles aux accents, à la tonalité et aux émotions ?

Les synthétiseurs vont créer des diagrammes de Mel pour chaque son émis, les diagrammes de Mel représentent toutes les caractéristiques d'un son. On utilise un speaker encoder dans le but que le diagramme de Mel produit par le synthétiseur se rapproche le plus de celui utilisé par notre voix. Après avoir récupéré l'intonation l'IA doit se saisir de notre voix pour cela il suffit d'entraîner un vocoder qui permettra d'utiliser ou de modifier notre voix. Le vocoder est une technologie qui était déjà utilisée depuis longtemps dans le domaine de la musique. Entraîner un vocoder ne prend pas énormément de temps.



5) Quelles sont les bienfaits ?

Avec ces nouvelles technologies on peut gagner beaucoup de temps surtout dans les métiers audiovisuels faire plusieurs projets demandant nos voix en même temps ou utiliser une voix de doublage d'un défunt et la réutiliser. Par exemple si un acteur américain ayant un doubleur français depuis des années et que celui-ci décède on peut utiliser sa voix pour continuer le doublage de son film. On peut aussi utiliser la voix d'un ancien chanteur pour refaire des musiques. En ce moment on peut voir Michael Jackson chanter sur des musiques de The Weeknd.

6) Les risques ?

Alliés au Deepfake qui est le fait d'utiliser l'image de quelqu'un, la modifier etc... On peut facilement usurper l'identité d'une personne, il y a donc un danger d'usurpation d'identité et cela peut donc augmenter le cyber attaque. De plus les métiers comme la chanson ou le doublage peuvent être menacés.

