

Recommandations pour la transcription d'entretiens audio par IA

1. Les enjeux éthiques des solutions en ligne

Lorsque vous utilisez des services en ligne pour transcrire des interviews, vos fichiers audios sont envoyés vers des serveurs externes. Cela pose des problèmes importants en matière de confidentialité et de protection des données personnelles. Les interviews peuvent contenir des informations sensibles, des opinions ou des données privées. L'empreinte vocale en tant que telle constitue une donnée personnelle. En utilisant des plateformes tierces, vous perdez le contrôle sur ces données, ce qui peut entraîner des risques de fuite, une utilisation non consentie ou une non-conformité avec des réglementations comme la LIPAD ou le RGPD. Pour des projets universitaires, où la confidentialité est essentielle, il est nécessaire d'éviter ces solutions.

2. Alternative locale recommandée

Pour garantir la sécurité des données, il existe des outils qui fonctionnent directement sur votre ordinateur, sans transfert vers le cloud. Ces outils utilisent des modèles performants et, sur un ordinateur récent sans carte graphique dédiée, le temps de transcription est généralement équivalent à la durée de l'audio¹, ce qui reste tout à fait raisonnable pour un usage académique.

La qualité est parfois un peu en deçà de solution en ligne, mais quelle que soit la solution, un travail complémentaire manuel est toujours nécessaire afin de pouvoir capter toutes les subtilités de l'interview. Les solutions locales sont de toutes manières les seules à vous donner l'assurance du respect de la confidentialité.

Il existe plusieurs solutions mais nous vous recommandons le logiciel **aTrain** développé par des chercheurs de l'Université de Graz. Vous devriez les détails pour l'installer sur votre ordinateur sur cette page :

aTrain : <https://business-analytics.uni-graz.at/en/research/atrain/>

Le logiciel aTrain utilise Whisper² comme modèle d'IA, mais contrairement à d'autres services en ligne qui l'utilisent aussi, le modèle est ici exécuté par aTrain entièrement sur

¹ Par exemple, une heure d'enregistrement prendra environ une heure pour être transcrite avec le meilleur modèle.

² Modèle entraîné par la société OpenAI : <https://openai.com/fr-FR/index/whisper/>

votre ordinateur. La transcription se fait donc sans que des données ne soient partagées en ligne.

3. Recommandations pour une transcription optimale

Il existe plusieurs versions du modèle Whisper. Chaque version offre un compromis entre vitesse et précision de transcription. Par défaut, c'est la version Whisper turbo qui est installée, elle offre un bon rapport vitesse/précision. Pour obtenir la meilleure transcription possible, il vous faudra télécharger dans l'application aTrain le dernière modèle Whisper Large (pour l'instant en V3).

La qualité de l'audio est un facteur déterminant. Évitez les bruits de fond, utilisez si possible un micro externe et assurez-vous que les interlocuteurs-trices parlent clairement sans se couper la parole.

Avant la transcription, il est conseillé de nettoyer le son et de segmenter les interviews longues en plusieurs parties.

Attention, ici, encore une fois, toute solution en ligne est à proscrire. Pour toutes vos éditions et conversions audio vous pouvez utiliser le logiciel [Audacity](#) qui est un excellent outil gratuit et open source, idéal pour l'édition et l'amélioration de la qualité audio. Nous vous recommandons l'installation sans le Muse hub.

Le format du fichier audio va aussi avoir son importance. Les formats sans perte tel que WAV offrent la meilleure qualité mais si le fichier est trop volumineux, l'impact sur le temps de transcription peut être important. Les formats MP3 ou WMA permettent d'améliorer le temps de transcription mais il convient de ne pas trop pousser la compression pour ne pas perdre en précision de transcription. Suivant la qualité de votre audio, il est possible d'apporter certaines corrections : égalisation du niveau entre les locuteurs, diminution du bruit de fond, etc., mais le mieux est de réussir à faire une captation directement de bonne qualité. Toute correction postérieure aura des limites.