

Hé YouTube, dans cette vidéo, je vais vous montrer comment vous pouvez rapidement convertir n'importe quel audio en texte en utilisant le package open source gratuit dans Python appelé Whisper. Je vais montrer que je l'ai installé, montrer un exemple de la façon dont je l'ai dirigé et le compare à une bibliothèque existante. Donc, en commençant, vous voudrez probablement aller au référentiel Whisper Get Hub que nous regardons ici et ils donnent des instructions sur la façon dont vous pouvez l'installer. Maintenant, une chose à garder à l'esprit lorsque vous installez simplement le nom de nom, il ne va pas installer la bonne version. Nous voulons installer à partir de ce référentiel GIT. Alors prenez simplement cette commande PIP Installation et exécutez-la dans votre environnement que vous exécutez Python. Et ils ont également mentionné ici que vous avez besoin de FFMPEG installé. Il y a des instructions pour le faire, mais je l'avais déjà installée sur mon ordinateur. Maintenant que j'ai Whisper Installer, faisons simplement un audio sur lequel je peux tester. Alors je vais dire quelques idiomes. Les idiomes sont généralement difficiles à comprendre pour les modèles. Même si ce n'est que de la parole au texte. Ce sera plutôt amusant. J'adorerais être sur Cloud 9 en tant que poney à un tour qui ne ferait pas de mal à une mouche. Je serais comme un poisson hors de l'eau et aussi en forme qu'un violon pour être sous le temps. Économisons cela. Sauvons-le comme une vague. Ils ont des instructions sur la façon dont nous pourrions exécuter cela juste à partir de la ligne de commande une fois qu'il est installé. Je vais vous montrer comment utiliser l'API Python, qu'ils montrent ici. C'est donc vraiment simple. Nous importons simplement chuchoter. Ensuite, nous allons créer notre modèle, c'est-à-dire que nous allons charger. modèle qui s'appelle Base. Et puis en utilisant simplement cet objet modèle, nous exécutons Transcribe sur notre fichier audio. Je l'ai donc nommé idioms. Utilisons la version Wave. Nous voulons que cela renvoie le résultat. Maintenant, j'ai remarqué que lorsque j'ai couru ceci avant, je reçois cette erreur à cause du demi-tenseur et du tenseur de flotteur de Kuda. J'ai pu résoudre ce problème. C'est donc quelque chose à garder à l'esprit. Si cela ne fonctionne pas pour vous, vous devrez peut-être régler le point flottant 16 pour tomber. Et vous pouvez voir après qu'il soit exécuté ici, il a détecté la langue déjà en tant qu'anglais, puis cet objet de résultat a quelques méthodes différentes, mais ce que nous voulons entrer à l'intérieur, c'est juste le texte et nous pourrions voir qu'il est apparenté. Comme le résultat est bon, j'aimerais être sur le nuage neuf en tant que poney à un tour qui ne ferait pas de mal à une mouche, je serais comme un poisson hors de l'eau et cela a gâché un peu ce poisson hors de l'eau dans. Aussi apte à un violon et peut-être que je ne l'ai pas dit assez clairement, une autre chose à savoir, c'est quand vous exécutez ceci pour la première fois, il devra télécharger le modèle de base. Vous pourriez donc voir une barre de progression passer et vous devrez télécharger ce modèle. Et il

est dit que lorsque vous exécutez ce transcript, il faut en fait 30 secondes de votre fichier audio et exécuter des prédictions dessus. Maintenant, il y a aussi une autre approche que vous pouvez adopter, ce qui est une approche de niveau inférieur, où vous créez réellement le modèle, puis vous créez l'objet audio et le modèle de la garniture. Vous vous assurez donc que ce morceau audio ne mesure que 30 secondes. secondes ou cela le tapera avec 30 secondes car c'est la longueur que le modèle prévoit d'avoir comme entrée. Ensuite, il fait un spectrogramme de la souris log. Il détecte la langue et nous pouvons décoder ici et offrir beaucoup plus d'options si nous le voulions. Si j'exécute cette cellule, obtenez encore cette erreur, que je peux maintenant définir dans les options de décodage, FP16 est égal aux défauts. Et en fait, cette fois, il semble que tout soit correct. Je serais comme un poisson hors de l'eau. et est en forme comme un violon. C'est tout pour chuchoter. Je veux juste le comparer à un type de modèle existant. Et une bibliothèque populaire pour ce faire est la bibliothèque de reconnaissance vocale. La façon dont nous exécutons la bibliothèque de reconnaissance vocale est que nous l'importons, puis créons cet objet de reconnaissance, avec lequel nous pouvons alors charger notre fichier audio. Après cela, vous pouvez prendre l'objet de reconnaissance et il existe quelques méthodes de reconnaissance différentes pour cela. Et nous allons utiliser le Google Reconnus et voyons quel est le résultat. Il semble donc que cela n'ajoute aucune ponctuation, et le nuage neuf est différent. J'adorerais être sur le nuage neuf en tant que poney à un tour qui ne ferait pas de mal à une mouche. Mais la seule chose à garder à l'esprit est que cela utilise réellement l'API de reconnaissance vocale Google. La bibliothèque Whisper, vous avez en fait le modèle téléchargé et c'est à vous. Je vous recommande également de jeter un œil au papier Whisper, qui a été publié avec ce code. Ils entrent également dans les détails de la façon dont le modèle a été formé et de l'architecture qu'elle est utilisée. Whisper fonctionne sur un tas de langues différentes. Les performances qu'ils disent varient en fonction de la langue. Vous pouvez donc aller ici sur le dépôt GitHub où ils ont une intrigue montrant quelles langues fonctionnent le mieux pour les bars ici. Plus petit est meilleur et plus grand signifie qu'il fonctionne pire. Donc toujours assez impressionnant le nombre de langues sur lesquelles ce modèle fonctionne.