Ehi YouTube, in questo video ti mostrerò come puoi convertire rapidamente qualsiasi audio in testo usando il pacchetto open source gratuito in Python chiamato Whisper.Mostrerò che l'ho installato, mostrare un esempio di come l'ho eseguito e lo confronto con una libreria esistente.Quindi, iniziando, probabilmente vorrai andare al Whisper Ottieni un repository hub che stiamo guardando qui e danno istruzioni su come installarlo.Ora una cosa da tenere a mente quando pip installi solo il nome sussurro non installerà la versione giusta.Vogliamo installare da questo repository Git.Quindi prendi questo comando PIP Installa ed eseguilo nel tuo ambiente che stai eseguendo Python.E hanno anche detto qui che è necessario installare FFM PEG.Ci sono alcune istruzioni per farlo, ma l'ho già installato sul mio computer.Ora che ho l'installazione di sussurri, facciamo solo un po 'di audio su cui posso testarlo.Quindi dirò alcuni idiomi.I idiomi sono generalmente difficili da capire per i modelli.Anche se questo è solo un discorso al testo.Questo sarà un po 'divertente.Mi piacerebbe essere sul Cloud 9 come un pony di un trucco che non farebbe male a una mosca.Sarei come un pesce fuori dall'acqua e in forma come un violino per essere sotto il tempo.Salviamolo.Salviamolo come un'onda.Hanno istruzioni su come potremmo eseguirlo direttamente dalla riga di comando una volta installata.Ti mostrerò come usare l'API Python, che mostrano qui.Quindi è davvero semplice.Importiamo solo sussurri.Quindi creeremo il nostro modello, che è caricare.Modello che si chiama base.E poi solo usando questo oggetto modello, eseguiamo Transcrive sul nostro file audio.Quindi l'ho chiamato idiomi.Usiamo la versione wave.Vogliamo che questo restituisca il risultato.Ora, ho notato quando ho eseguito questo prima, ricevo questo errore a causa del mezzo tensore di Kuda e del tensore galleggiante.Sono stato in grado di risolverlo.Quindi è qualcosa da tenere a mente.Se non funziona per te, potrebbe essere necessario impostare il punto volante 16 per cadere.E puoi vedere dopo aver funzionato qui, ha rilevato la lingua già come inglese e quindi questo oggetto di risultato ha alcuni metodi diversi in essi, ma ciò che vogliamo entrare all'interno di questo è solo il testo e potremmo vedere che è un aspettocome se il risultato fosse buono, mi piacerebbe essere sul cloud nove come un pony di un trucco che non farebbe male a una mosca sarei come un pesce fuori

dall'acqua e questo ha rovinato un po 'questo pesce fuori dall'acquaFit come un violino e forse non l'ho detto abbastanza chiaramente un'altra cosa da sapere è quando lo esegui per la prima volta, dovrà scaricare il modello di base.Quindi potresti vedere una barra di avanzamento e dovrai scaricare quel modello.E dice che quando si esegue questo trascrizione, in realtà impiegano 30 secondi pezzi del tuo file audio ed esegue previsioni su di esso.Ora c'è anche un altro approccio che puoi adottare, che è un approccio di livello inferiore, in cui si crea effettivamente il modello e quindi si crea l'oggetto audio e il taglio del pattern.Quindi ti assicuri che questo pezzo audio sia di soli 30 secondi.Pochi secondi o lo darà una pacca con 30 secondi poiché questa è la lunghezza che il modello prevede di avere come input.Quindi sta realizzando uno spettrogramma del mouse log.Sta rilevando la lingua e possiamo decodificare qui e fornire molte più opzioni se volessimo.Se eseguo questa cella, ricevo di nuovo questo errore, che ora posso impostare nelle opzioni di decodifica, FP16 è uguale ai guasti.E in realtà questa volta sembra che abbia ottenuto tutto corretto.Sarei come un pesce fuori dall'acqua.ed è in forma come violino.Quindi è tutto per Whisper.Voglio solo confrontarlo con un tipo di modello esistente.E una biblioteca popolare per farlo è la biblioteca di riconoscimento vocale.Il modo in cui gestiamo la libreria di riconoscimento vocale è che la importa e quindi creiamo questo oggetto riconoscimento, con cui quindi possiamo caricare il nostro file audio.Successivamente, puoi prendere l'oggetto riconoscimento e ci sono alcuni diversi metodi di riconoscimento per questo.E useremo Google riconoscere e vediamo qual è il risultato.Quindi sembra che non abbia aggiunto alcuna punteggiatura e il cloud nove è diverso.Mi piacerebbe essere sul cloud nove come un pony di un trucco che non farebbe male a una mosca.Ma l'unica cosa da tenere a mente è che questo sta effettivamente usando l'API di riconoscimento vocale di Google.La biblioteca di Whisper, in realtà hai scaricato il modello ed è tuo da usare.Ti consiglio anche di dare un'occhiata al whisper Paper, che è stato rilasciato con questo codice.Entrano anche nei dettagli su come è stato addestrato il modello e sull'architettura che viene utilizzata.Whisper lavora su un sacco di lingue diverse.La performance che dicono varia in base alla lingua.Quindi puoi andare qui sul Repo di Github

dove hanno una trama che mostra quali lingue si comportano meglio per i bar qui.Più piccolo è migliore e più grande significa che funziona peggio.Quindi è comunque piuttosto impressionante il numero di lingue su cui funziona questo modello.