

Hola YouTube, en este video te mostraré cómo puedes convertir rápidamente cualquier audio en texto usando el paquete de código abierto gratuito en Python llamado Whisper. Voy a mostrar que lo instalé, mostrar un ejemplo de cómo lo ejecuté y compararlo con una biblioteca existente. Entonces, comenzando, probablemente querrá ir al repositorio Whisper Get Hub que estamos viendo aquí y dan instrucciones sobre cómo puede instalarlo. Ahora, una cosa a tener en cuenta, cuando PIP instale solo el nombre, no va a instalar la versión correcta. Queremos instalar desde este repositorio Git. Así que solo tome este comando de instalación de PIP y ejecute en su entorno que esté ejecutando Python. Y también mencionaron aquí que necesita instalado FFMPEG. Hay algunas instrucciones para hacerlo, pero ya lo tuve instalado en mi computadora. Ahora que tengo una instalación de susurro, hagamos un poco de audio en el que pueda probar esto. Así que voy a decir algunos modismos. Los modelos suelen ser difíciles de entender para los modelos. A pesar de que esto es solo discurso a texto. Esto será un poco divertido. Me encantaría estar en la nube 9 como un pony único que no dolería una mosca. Sería como un pez fuera del agua y tan en forma como un violín para estar bajo el clima. Guardemos esto. Guardemos como una ola. Tienen instrucciones sobre cómo podríamos ejecutar esto simplemente directamente desde la línea de comando una vez que está instalada. Te mostraré cómo usar la API de Python, que muestran aquí. Entonces es realmente simple. Solo importamos susurros. Entonces vamos a crear nuestro modelo, que vamos a cargar. modelo que se llama base. Y luego solo usando este objeto modelo, ejecutamos Transcribe en nuestro archivo de audio. Entonces lo llamé modismos. Usemos la versión de onda. Queremos que esto devuelva el resultado. Ahora, me di cuenta de que cuando corrí esto antes, recibo este error debido a la mitad tensor de Kuda y el tensor flotante. Pude resolver esto. Entonces eso es algo a tener en cuenta. Si no funciona para usted, es posible que deba establecer el punto flotante 16 para caer. Y puede ver después de que se ejecute aquí, detectó el idioma ya como inglés y luego este objeto de resultado tiene algunos métodos diferentes en ellos, pero lo que queremos dentro de esto es solo el texto y podríamos ver que parece que se ve. Al igual que el resultado, es bueno, me encantaría estar

en la nube nueve como un pony de un truco que no lastimaría a una mosca, sería como un pez fuera del agua y esto se arruinó un poco de este pez de agua enTan en forma como un violín y tal vez no lo dije con claridad, otra cosa es saber es cuando ejecute esto por primera vez, tendrá que descargar el modelo base.Por lo tanto, es posible que vea una barra de progreso y tendrá que descargar ese modelo.Y dice que cuando ejecuta esta transcripción, en realidad está tomando 30 segundos de su archivo de audio y ejecutando predicciones en él.Ahora también hay otro enfoque que puede adoptar, que es un enfoque de nivel inferior, donde realmente crea el modelo y luego crea el objeto de audio y el patrón recorta esto.Entonces, solo asegúrese de que este fragmento de audio sea de solo 30 segundos.segundos o lo dará una palmada con 30 segundos, ya que esa es la longitud que el modelo espera tener como entrada.Luego está haciendo un espectrograma del mouse log.Está detectando el lenguaje y podemos decodificar aquí y proporcionar muchas más opciones si quisiéramos.Si ejecuto esta celda, nuevamente obtenga este error, que ahora puedo configurar en las opciones de decodificación, FP16 es igual a fallas.Y en realidad esta vez parece que todo tiene todo correcto.Sería como un pez fuera del agua.y está en forma como un violín.Así que eso es todo para susurros.Solo quiero compararlo con un tipo de modelo existente.Y una biblioteca popular para hacer esta es la biblioteca de reconocimiento de voz.La forma en que ejecutamos la biblioteca de reconocimiento de voz es que la importamos y luego creamos este objeto reconocedor, con el que luego podemos cargar nuestro archivo de audio.Después de eso, puede tomar el objeto reconocedor y hay algunos métodos de reconocimiento diferentes para eso.Y vamos a usar el reconocimiento de Google y veamos cuál es el resultado.Por lo tanto, parece que no agregé ninguna puntuación, y la nube nueve es diferente.Me encantaría estar en la nube nueve como un pony único que no dolería una mosca.Pero lo único a tener en cuenta es que esto realmente está usando la API de reconocimiento de voz de Google.La biblioteca Whisper, en realidad, se descarga el modelo y es tuyo para usar.También le recomiendo que eche un vistazo al documento Whisper, que se lanzó con este código.También entran en detalles sobre cómo se entrenó el modelo y la

arquitectura que se usa. Whisper trabaja en un montón de idiomas diferentes. El rendimiento que dicen varía según el idioma. Así que puedes ir aquí en el repositorio de GitHub, donde tienen una trama que muestra qué idiomas realmente funcionan mejor para los bares aquí. Más pequeño es mejor y más grande significa que funciona peor. Así que sigue siendo bastante impresionante la cantidad de idiomas en los que este modelo funciona.