

Hey YouTube, in diesem Video werde ich Ihnen zeigen, wie Sie jeden Audio mit dem kostenlosen Open -Source -Paket in Python namens Whisper schnell in Text umwandeln können. Ich werde zeigen, dass ich es installiert habe, ein Beispiel dafür zeige, wie ich es ausgeführt habe und es mit einer vorhandenen Bibliothek vergleiche. Wenn Sie also anfangen, möchten Sie wahrscheinlich zum Whisper Get Hub -Repository gehen, das wir uns hier ansehen, und sie geben Anweisungen dazu, wie Sie es installieren können. Nun eine Sache, die Sie beachten sollten, wenn Sie nur den Namen Whisper installieren. Es wird nicht die richtige Version installieren. Wir möchten aus diesem Git -Repository installieren. Nehmen Sie also diesen PIP -Installationsbefehl und führen Sie ihn in Ihrer Umgebung aus, in der Sie Python ausführen. Und sie erwähnten auch hier, dass Sie FFMPEG installiert haben. Es gibt einige Anweisungen dazu, aber ich hatte das bereits auf meinem Computer installiert. Jetzt, da ich Whisper Installation habe, lassen Sie uns einfach ein Audio erstellen, auf dem ich dies testen kann. Also werde ich ein paar Redewendungen sagen. Redewendungen sind normalerweise schwer für Modelle zu verstehen. Auch wenn dies nur eine Sprache zum Text ist. Das wird Spaß machen. Ich würde gerne als One Trick Pony auf Cloud 9 sein, das keine Fliege verletzen würde. Ich wäre wie ein Fisch aus Wasser und so fit wie eine Geige, um unter dem Wetter zu sein. Lassen Sie uns dies speichern. Lassen Sie es uns als Welle retten. Sie haben Anweisungen, wie wir dies nur direkt aus der Befehlszeile ausführen können, sobald sie installiert ist. Ich werde Ihnen zeigen, wie Sie die Python -API verwenden, die sie hier zeigen. Es ist also wirklich einfach. Wir importieren nur flüsterlich. Dann werden wir unser Modell erstellen, nämlich werden wir laden. Modell, das als Basis bezeichnet wird. Und dann verwenden wir nur dieses Modellobjekt, und wir führen Transkription unserer Audio -Datei aus. Also habe ich es Redewendungen genannt. Verwenden wir die Wellenversion. Wir möchten, dass dies das Ergebnis zurück gibt. Jetzt, als ich das schon einmal lief, erhalte ich diesen Fehler wegen Kudas Halbzensor- und Schwimmer -Tensor. Ich konnte das lösen. Das ist also etwas, das man beachten sollte. Wenn es für Sie nicht funktioniert, müssen Sie möglicherweise den schwimmenden Punkt 16 festlegen, um zu fallen. Und Sie können sehen, nachdem es hier ausgeführt wurde, die Sprache bereits als Englisch entdeckt hat, und dann enthält dieses Ergebnisobjekt einige verschiedene Methoden, aber was wir in diesen Text einbeziehen möchten, ist nur der Text und wir können sehen, dass es aussieht Als ob das Ergebnis gut ist, würde ich gerne als One -Trick -Pony auf Cloud Nine sein, das eine Fliege nicht schaden würde, würde ich wie ein Fisch aus dem Wasser sein, und dies hat diesen Fisch ein wenig durcheinander gebracht So fit wie eine Geige und vielleicht habe ich es nicht klar genug gesagt. Eine andere Sache, um zu wissen, wenn Sie dies zum ersten Mal ausführen, muss es das Basismodell herunterladen. Vielleicht sehen Sie eine Fortschrittsleiste und Sie m

üssen dieses Modell herunterladen. Und wenn Sie diesen Transkriber ausführen, benötigt er tatsächlich 30 Sekunden - Teile Ihrer Audio -Datei und leitet Vorhersagen darauf aus. Jetzt gibt es auch einen weiteren Ansatz, den Sie verfolgen können. Dies ist ein Ansatz mit niedrigerer Ebene, bei dem Sie das Modell tatsächlich erstellen und dann das Audioobjekt erstellen und diese Muster transportieren. Sie stellen also nur sicher, dass dieser Audio -Stück nur 30 Sekunden beträgt. Sekunden oder es wird es mit 30 Sekunden, da dies die Länge ist, die das Modell als Eingabe erwartet. Dann macht es ein Protokoll -Mausspektrogramm. Es erkennt die Sprache und wir können hier dekodieren und viel mehr Optionen bieten, wenn wir wollten. Wenn ich diese Zelle ausführe, erhalte ich diesen Fehler, den ich jetzt in den Dekodierungsoptionen einstellen kann, FP16 entspricht Fehlern. Und diesmal sieht es so aus, als hätte es alles richtig. Ich wäre wie ein Fisch aus Wasser. und ist fit wie eine Geige. Das ist es für Whisper. Ich möchte es nur mit einem vorhandenen Modelltyp vergleichen. Und eine beliebte Bibliothek dafür ist die Spracherkennungsbibliothek. Die Art und Weise, wie wir die Spracherkennungsbibliothek ausführen, importieren wir sie und erstellen dann dieses Erkennungsobjekt, mit dem wir unsere Audio -Datei laden können. Danach können Sie das Erkennungsobjekt aufnehmen und es gibt einige unterschiedliche Erkennungsmethoden dafür. Und wir werden das Google -Erkennen und Mal sehen, was das Ergebnis ist. Es sieht also so aus, als hätte es keine Interpunktion hinzugefügt, und die Cloud Nine ist anders. Ich würde gerne als One Trick Pony auf Cloud Nine sein, das keine Fliege verletzen würde. Das einzige, was man beachten sollte, ist, dass dies tatsächlich die Google -Spracherkennungs -API verwendet. In der Whisper -Bibliothek haben Sie tatsächlich das Modell herunterladen und es gehört Ihnen. Ich empfehle auch, dass Sie sich das Whisper -Papier ansehen, das mit diesem Code veröffentlicht wurde. Sie gehen auch in Details ein, wie das Modell trainiert wurde und welche Architektur es verwendet wird. Whisper arbeitet an einer Reihe verschiedener Sprachen. Die Leistung, die sie sagen, variiert je nach Sprache. Sie können also hier auf dem Github -Repo gehen, wo eine Handlung zeigt, die zeigt, welche Sprachen hier tatsächlich am besten für die Bars erfolgen. Kleiner ist besser und größer und bedeutet, dass es schlechter funktioniert. Die Anzahl der Sprachen, an denen dieses Modell funktioniert, ist also immer noch ziemlich beeindruckend.