# Convert YouTube Videos into Subtitled Text with Whisper AI and Pytube

Anvi Kohli · Follow

3 min read · Aug 31, 2023



In this article, we'll dive into a project to convert YouTube videos into neatly subtitled text using the OpenAI's Whisper AI and Pytube. Whether you're a content enthusiast, a curious learner, or someone who loves to have subtitles on videos, this project empowers you to effortlessly transform spoken words into insightful, timestamped text. Let's Go!

## Project Overview

The project involves the following steps:

1. Installing Dependencies: We'll install the Whisper AI library using the `git` command and the pip command for Pytube library for downloading YouTube videos.
2. Converting YouTube Video to Audio: Using Pytube, we'll download a YouTube video and convert it into an audio file ( `.mp3` format).
3. Transcribing Audio: We'll use the Whisper AI tool to transcribe the audio file, generating a list of segments with timestamps and corresponding text.
4. Formatting and Saving: We'll format the transcribed segments with timestamps in the `[hh:mm:ss.sss --> hh:mm:ss.sss]` format and save the result in a text file.

## Technologies and Libraries Used

- Whisper AI: Whisper AI is a state-of-the-art automatic speech recognition (ASR) system developed by OpenAI. It's designed to convert spoken language into written text.
- Pytube: The Pytube library provides a simple and efficient way to interact with YouTube videos, enabling us to download content for

further processing.

- Torch: Torch is used for neural network computations and forms an integral part of the Whisper AI framework.

## Step-by-Step

## Step 1: Installing Dependencies

Before we proceed, we need to install the necessary libraries:

```
!pip -qqq install git+https://github.com/openai/whisper.git
!pip -qqq install pytube
```

## Step 2: Importing Libraries and Loading the Model

```
from pytube import YouTube
import whisper
```

```
import torch
import os

device = "cuda" if torch.cuda.is_available() else "cpu"
whisper_model = whisper.load_model("large", device=device)
```

## Step 3: Converting YouTube Video to Audio

In this step, we'll use the Pytube library to download the YouTube video and convert it to audio:

```
def video_to_audio(video_URL, destination, final_filename):

  # Get the video
  video = YouTube(video_URL)

  # Convert video to Audio
  audio = video.streams.filter(only_audio=True).first()

  # Save to destination
  output = audio.download(output_path = destination)

  _, ext = os.path.splitext(output)
  new_file = final_filename + '.mp3'

  # Change the name of the file
  os.rename(output, new_file)
```

```
def convert(url):
  # Video to audio
  video_URL = url
  destination = "."
  final_filename = "audio_file_to_convert"
  video_to_audio(video_URL, destination, final_filename)
```

## Step 4: Transcribing the Audio

We'll leverage Whisper AI to transcribe the audio from the downloaded video:

```
def transcribe():
  audio_file = "audio_file_to_convert.mp3"
  result = whisper_model.transcribe(audio_file)
  result_segments = result['segments']
  print(result_segments)
  return format_segments(result_segments)
```

## Step 5: Formatting and Saving

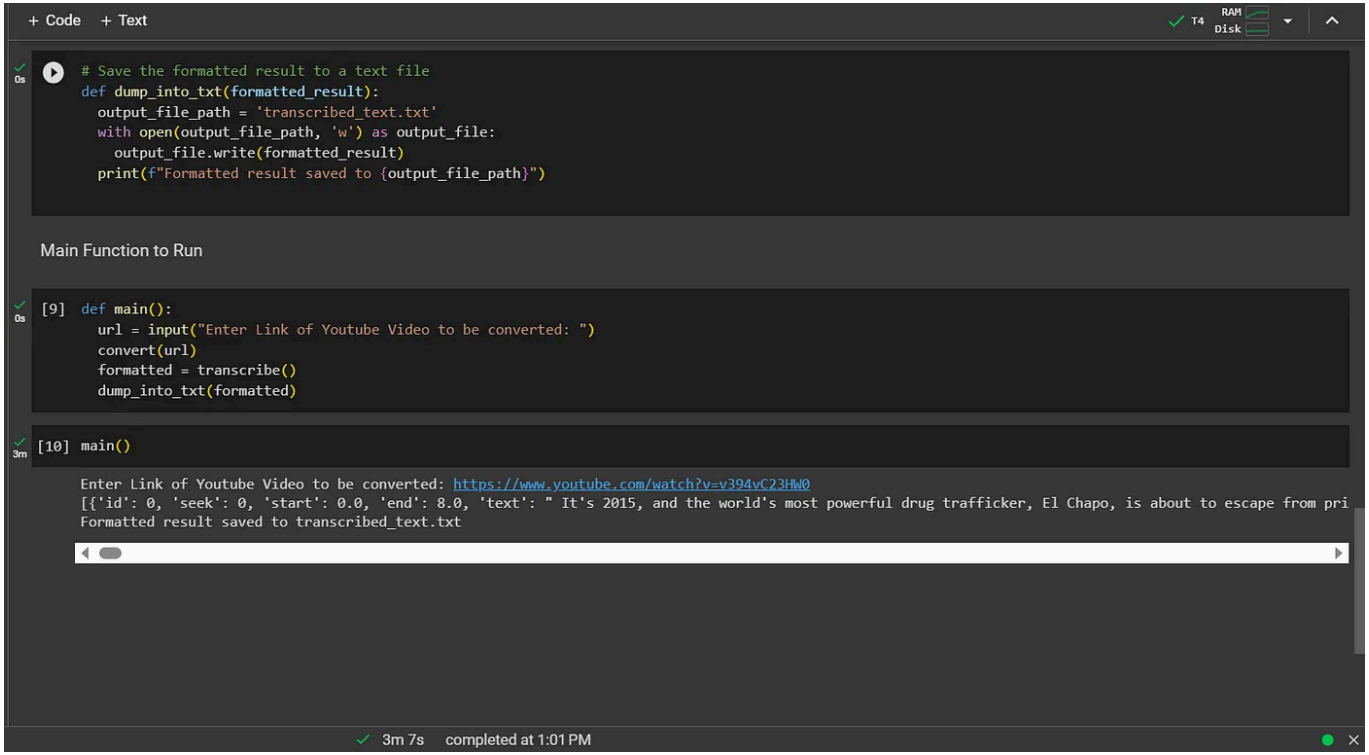
We'll format the transcribed segments and save them to a text file:

```
def format_segments(result_segments):
    formatted_output = []

    for segment in result_segments:
        start_time = segment['start']
        end_time = segment['end']
        text = segment['text']

        formatted_text = f"[{format_time_milliseconds(start_time)} --> {format_t
        formatted_output.append(formatted_text)

    return "\n".join(formatted_output)

def format_time_milliseconds(seconds):
    minutes, seconds = divmod(seconds, 60)
    hours, minutes = divmod(minutes, 60)
    milliseconds = int((seconds - int(seconds)) * 1000)
    return f"{int(hours):01}:{int(minutes):01}:{int(seconds):02}.{milliseconds:0
```

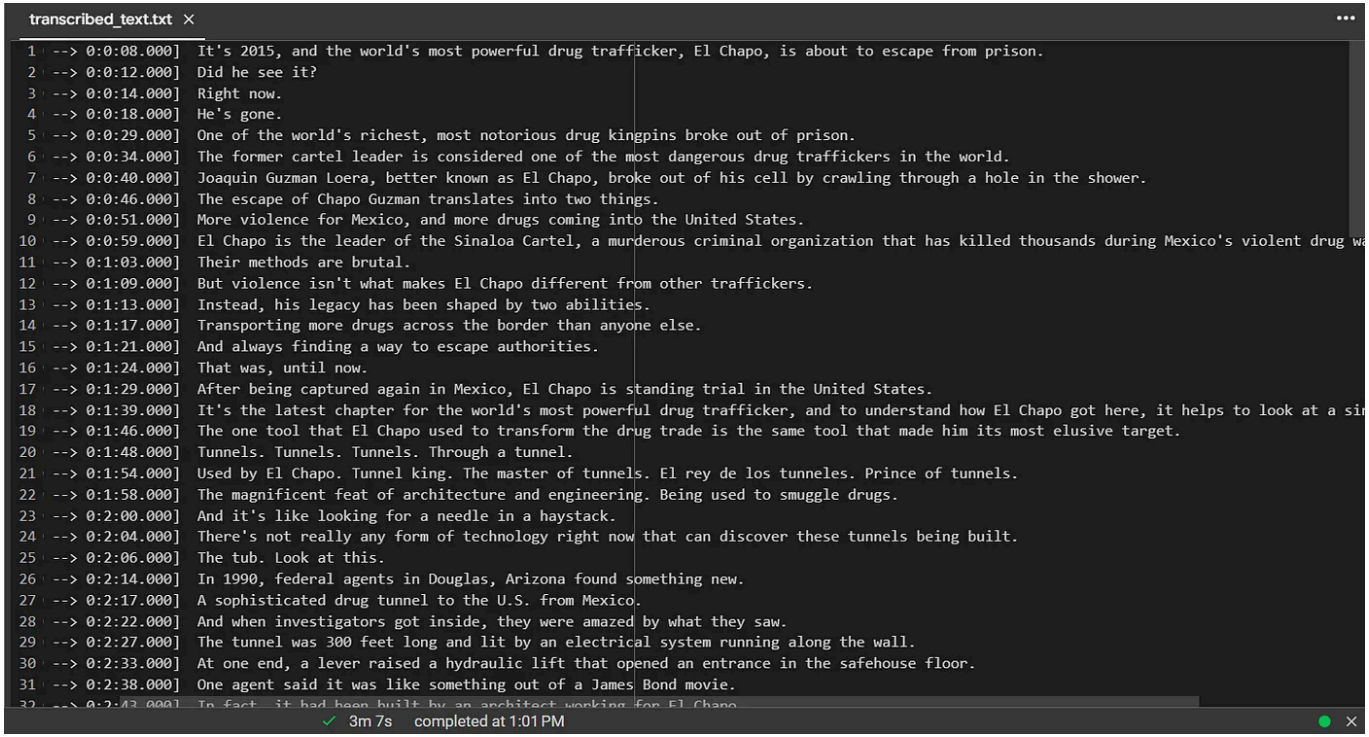
## Function to save formatted text into a txt file

```
# Save the formatted result to a text file
def dump_into_txt(formatted_result):
  output_file_path = 'transcribed_text.txt'
  with open(output_file_path, 'w') as output_file:
    output_file.write(formatted_result)
  print(f"Formatted result saved to {output_file_path}")
```

And that's it. Here is the simple way you can transcribe any video from YouTube and generate its transcripts. Not just YouTube, but the WhisperAI can be used to transcribe any video in any format.




Main Function




Generated Text File using the code

The code of the above article can be found here: anvichip/yt-video-audio2text-whisper (github.com)

Happy Learning!

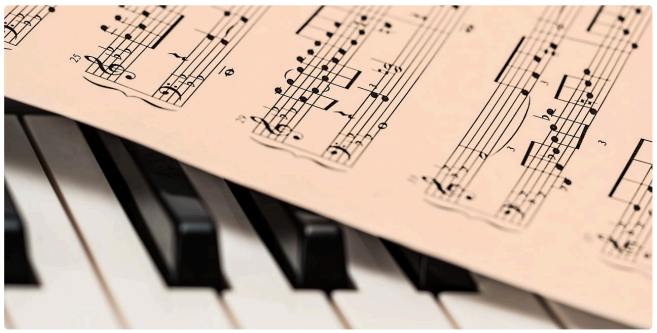
## Written by Anvi Kohli

Follow

2 Followers

Techie | Budding AI/ML scientist | Healthcare

### More from Anvi Kohli



Anvi Kohli

## Get Images & Videos for free from keywords using Pixabay API

Visual content plays a pivotal role in captivating audiences and conveying...

Mar 26

Anvi Kohli

## Navigating the Cloud: A Journey through AWS Solutions Architect...

Empowering Architects to Design Resilient Cloud Solutions

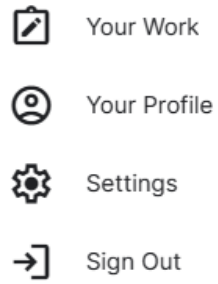
Dec 17, 2023 1

Anvi Kohli

## Generate Images using Stable Diffusion XL model

Welcome to the Part—1 of the Stable Diffusion Series where we today we will chec...

May 2 1




Anvi Kohli

## How to Load Kaggle Datasets directly onto Google Colab—3...

Tired of downloading large datasets locally to run large AI models? Well, thankfully for us ...

Mar 30

See all from Anvi Kohli

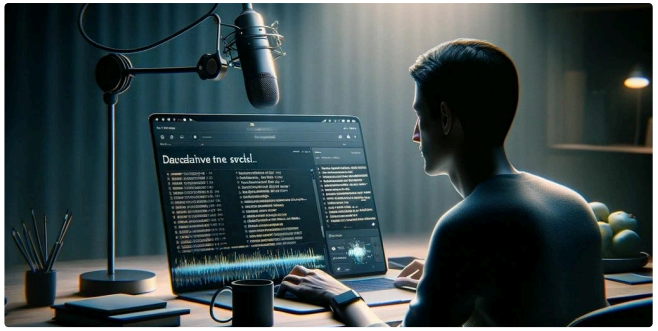
## Recommended from Medium

Mark OBrien

## Whisper & Python for Video Transcription

Explore the Whisper Model from OpenAI for audio transcription made easy, fully local

Mar 4 6



Jeremy Savage

## Using OpenAI's Whisper to Transcribe Real-time Audio

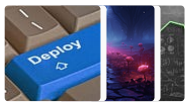
The availability of advanced technology and tools, in particular, AI is increasing at an ever...

Apr 12 72

### Lists

**Coding & Development**

11 stories · 694 saves




**Predictive Modeling w/ Python**

20 stories · 1364 saves

**Practical Guides to Machine Learning**

10 stories · 1645 saves

**ChatGPT**

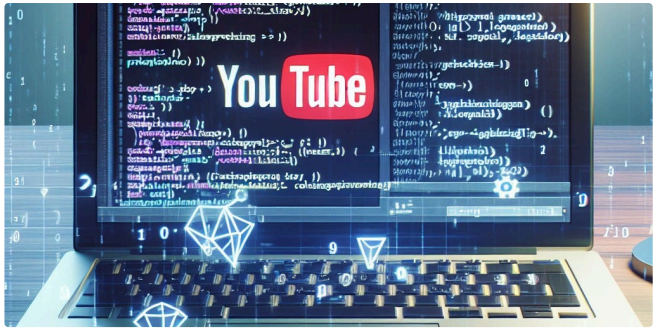
21 stories · 711 saves

Dean Martin

## Text to Speech from C# using and XTTS v2 (Python), with Chains &...

In this post we build on the previous post to build a completely free and state-of-the-art...

Jan 22 21

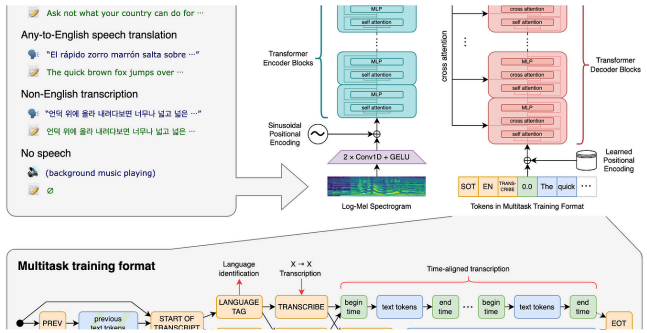



Imane Momayiz

## Scraping Youtube video transcripts

How to fetch video transcripts from any Youtube channel with YouTube Data API and...

Mar 14




Benjamin Consolvo in Intel Analytics Software

## Automatic Speech Recognition using OpenAI Whisper without a...


Be Tech! with Santander in Be Tech! with Santander

## Guide: The Rise of Voice Cloning Technology

Easy Step-by-Step Guide to English and French Transcription and Translation on CPUs

Mar 14 211 4

By Vicentes Motos. Discover how to generate realistically human voices using a variety of...

Feb 22 57

See more recommendations