LLM, RAG – Video Transcript Retrieval
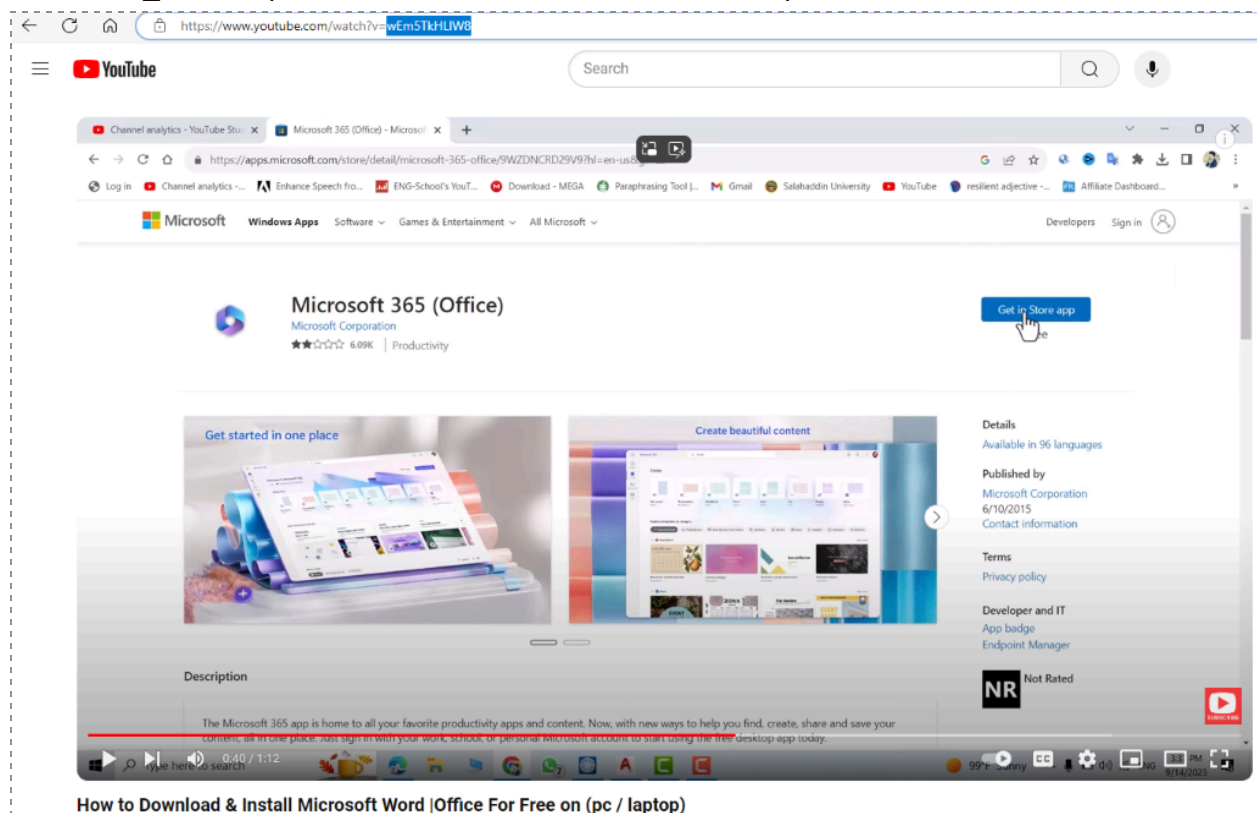
First install the YouTube transcript api by typing the following command into the terminal.

pip install youtube-transcript-api

Then enter the following command to retrieve the transcript of the video:

youtube_transcript_api <video_id>

Your video_id is the part after ?v= in the video url, see example below:

In this sample video your video_id would be what is highlighted in blue on the url, "wEm5TkHLlW8"

Then format the transcript into one JSON file and save this file by typing the following command:

youtube_transcript_api video_id ... --languages de en --format json > transcripts.json

Finally check to make sure the JSON file is in your present working directory by typing ls.