

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM
KHOA CƠ KHÍ CHẾ TẠO MÁY**



HCMUTE

MÔN: TRÍ TUỆ NHÂN TẠO

ĐỀ TÀI: TẠO PHỤ ĐỀ CHO VIDEO CLIP

GVHD: PGS. TS. Nguyễn Trường Thịnh

SVTH: Đào Thị Cẩm Tú

MSSV: 20104009

TP. HCM, tháng 12 năm 2022

MỤC LỤC

CHƯƠNG 1: TỔNG QUAN	3
1.1. Giới thiệu	3
1.2. Ý nghĩa thực tiễn	3
1.3. Mục tiêu đề tài	3
CHƯƠNG 2: CƠ SỞ LÝ THUYẾT.....	4
2.1. Trí tuệ nhân tạo.....	4
2.1.1. Định nghĩa	4
2.1.2. Quá trình hình thành và phát triển.....	4
2.2. Sơ lược về xử lý ngôn ngữ tự nhiên	5
2.2.1. Định nghĩa	5
2.2.2. Các bài toán và ứng dụng	5
2.3. OpenAI Whisper.....	7
2.4. Thư viện FFmpeg	9
2.5. Streamlit.....	10
CHƯƠNG 3: CHƯƠNG TRÌNH VÀ KẾT QUẢ.....	12
3.1. Chương trình chính.....	12
3.2. Tạo giao diện cho phần mềm.....	13
3.3. Kết quả.....	14
CHƯƠNG 4: KẾT LUẬN.....	16
DỮ LIỆU UP LÊN GITHUB	17
TÀI LIỆU KHAM KHẢO.....	18

CHƯƠNG 1: TỔNG QUAN

1.1. Giới thiệu

AI hay còn gọi là trí tuệ nhân tạo – Artificial Intelligence, một công nghệ gọi là học máy ghi nhận, thu thập nhận thức và suy nghĩ của con người. Ứng dụng AI phổ biến hiện nay đó là Nhận dạng giọng nói. Công nghệ này không còn là một khái niệm mới mẻ, tất cả những ông lớn ngành công nghệ đều đã đang lặn sâu vào “cuộc chơi” này. Đó là trợ lý ảo Siri của Apple, Cortana của Microsoft, Alexa của Amazon, đến cả Samsung cũng chấp chững cho thai nghén trợ lý Bixby của riêng mình hay không thể không kể đến Google Assistant của Google. Việc ứng dụng Nhận diện giọng nói vào cuộc sống giúp chuyển đổi giọng nói thành văn bản một cách nhanh chóng, hiệu quả. Đây là một thuận lợi đối với người khuyết tật khi họ không thể dùng chuột hay bàn phím, nhưng có thể dùng giọng nói để hệ thống chuyển thành văn bản, giúp nhập liệu hay điều khiển một cách dễ dàng.

Bên cạnh đó, việc chuyển đổi giọng nói thành văn bản có thể giúp ích cho mọi người khi xem phim, video trong môi trường nhạy cảm với âm thanh như nơi công cộng, những địa điểm hạn chế tiếng ồn bằng việc tạo phụ đề cho video. Tạo video có phụ đề chi tiết sẽ cho phép người xem vẫn xem được nội dung cần nghe, vì phụ đề truyền tải chính xác những gì đang được cập nhật trong video.

1.2. Ý nghĩa thực tiễn

Ngày nay, nhu cầu giải trí là một phần không thể thiếu trong đời sống của chúng ta. Chuẩn bị phụ đề cho video có nghĩa là một người trên xe bus, một bà mẹ đang ru con ngủ và thậm chí một người trên máy bay quên tai nghe vẫn có thể xem video. Từ hiệu quả và tính chính xác của công nghệ Nhận diện giọng nói thì nó đã được áp dụng vào việc tạo phụ đề cho những chiếc video.

1.3. Mục tiêu đề tài

Mục tiêu của đề tài là ứng dụng công nghệ nhận diện giọng nói Speech to text để chuyển lời thoại trong video thành văn bản. Từ văn bản đã được tạo, chèn lên phía dưới video tạo thành những dòng chữ thể hiện nội dung lời thoại, đó là phụ đề.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

2.1. Trí tuệ nhân tạo

2.1.1. Định nghĩa

Trí tuệ nhân tạo (TTNT), tiếng Anh là artificial intelligence hay chữ viết tắt được dùng phổ biến là AI, còn có thể hiểu bình dân hơn là ‘thông minh nhân tạo’, tức là sự thông minh của máy móc do con người tạo ra, đặc biệt tạo ra cho máy tính, robot, hay các máy móc có các thành phần tính toán điện tử.

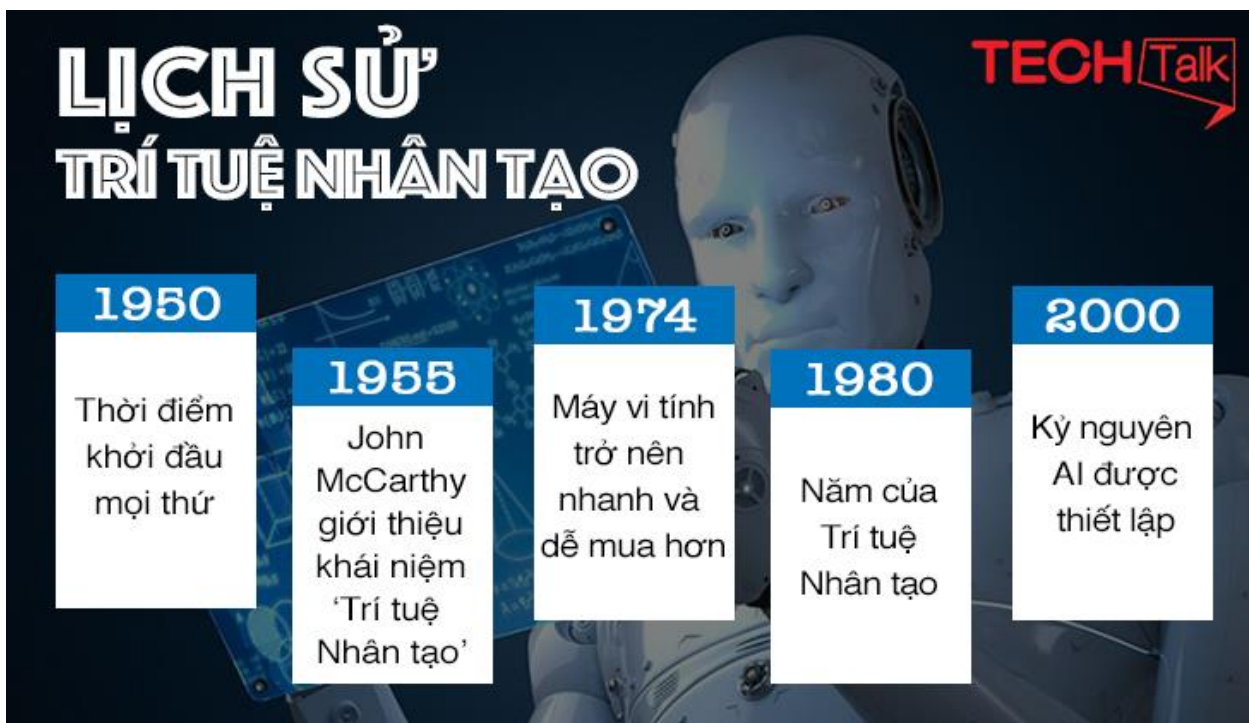
Trí tuệ nhân tạo là một lĩnh vực của khoa học và công nghệ nhằm làm cho máy có những khả năng của trí tuệ và trí thông minh của con người, tiêu biểu như biết suy nghĩ và lập luận để giải quyết vấn đề, biết giao tiếp do hiểu ngôn ngữ và tiếng nói, biết học và tự thích nghi.

Trí tuệ nhân tạo này do con người lập trình ra với mục đích tự động hóa các hành vi thông minh như con người, từ đó cắt giảm bớt nhân công là con người và có tính chuẩn xác cao hơn.

Sự khác biệt của trí tuệ nhân tạo so với các lập trình logic trước kia chính là khả năng suy nghĩ độc lập của chúng, thay vì việc mọi thứ được lập trình sẵn và cỗ máy đó sẽ thực hiện các thao tác theo logic được con người đặt ra, AI - Trí Tuệ Nhân Tạo sẽ tự xem xét tình huống và đưa ra phương án tối ưu nhất, qua đó tiết kiệm chi phí cũng như vận hành cho công việc hiệu quả hơn.

2.1.2. Quá trình hình thành và phát triển

Mong muốn làm cho máy có những khả năng của trí thông minh con người đã có từ nhiều thế kỷ trước, tuy nhiên Trí tuệ nhân tạo chỉ xuất hiện khi con người sáng tạo ra máy tính điện tử. Vào năm 1950, Alan Turing - một nhà đa khoa người Anh đã xuất bản “Computing Machinery and Intelligence” - Máy tính và trí thông minh. Ông được xem là người đặt tiền đề cho AI bằng cách đề xuất thử nghiệm Turing - một phương pháp xác định máy tính có suy nghĩ và thực sự thông minh hay không.



Hình 1. Sự phát triển của Trí tuệ nhân tạo qua các năm

2.2. Sơ lược về xử lý ngôn ngữ tự nhiên

2.2.1. Định nghĩa

Xử lý ngôn ngữ tự nhiên (natural language processing - NLP) là một nhánh nghiên cứu của trí tuệ nhân tạo, trong đó phát triển các thuật toán, xây dựng các chương trình máy tính có khả năng phân tích, xử lý, và hiểu ngôn ngữ của con người. NLP tập trung vào việc nghiên cứu sự tương tác giữa máy tính và ngôn ngữ tự nhiên của con người, dưới dạng tiếng nói (speech) hoặc văn bản (text).

Mục tiêu của lĩnh vực này là giúp máy tính hiểu và thực hiện hiệu quả những nhiệm vụ liên quan đến ngôn ngữ của con người như: tương tác giữa người và máy, cải thiện hiệu quả giao tiếp giữa con người với con người, hoặc đơn giản là nâng cao hiệu quả xử lý văn bản và lời nói.

2.2.2. Các bài toán và ứng dụng

Nhận dạng chữ viết: Có hai kiểu nhận dạng, thứ nhất là nhận dạng chữ in, ví dụ nhận dạng chữ trên sách giáo khoa rồi chuyển nó thành dạng văn bản điện tử như dưới định dạng

doc của Microsoft Word chẳng hạn. Phức tạp hơn là nhận dạng chữ viết tay, có khó khăn bởi vì chữ viết tay không có khuôn dạng rõ ràng và thay đổi từ người này sang người khác. Với chương trình nhận dạng chữ viết in có thể chuyển hàng ngàn đầu sách trong thư viện thành văn bản điện tử trong thời gian ngắn. Nhận dạng chữ viết của con người có ứng dụng trong khoa học hình sự và bảo mật thông tin (nhận dạng chữ ký điện tử).

Nhận dạng tiếng nói: Nhận dạng tiếng nói rồi chuyển chúng thành văn bản tương ứng. Giúp thao tác của con người trên các thiết bị nhanh hơn và đơn giản hơn, chẳng hạn thay vì gõ một tài liệu nào đó bạn đọc nó lên và trình soạn thảo sẽ tự ghi nó ra. Đây cũng là bước đầu tiên cần phải thực hiện trong ước mơ thực hiện giao tiếp giữa con người với robot. Nhận dạng tiếng nói có khả năng trợ giúp người khiếm thị rất nhiều.

Tổng hợp tiếng nói: Từ một văn bản tự động tổng hợp thành tiếng nói. Thay vì phải tự đọc một cuốn sách hay nội dung một trang web, nó tự động đọc cho chúng ta. Giống như nhận dạng tiếng nói, tổng hợp tiếng nói là sự trợ giúp tốt cho người khiếm thị, nhưng ngược lại nó là bước cuối cùng trong giao tiếp giữa robot với người.

Dịch tự động (machine translate): Như tên gọi đây là chương trình dịch tự động từ ngôn ngữ này sang ngôn ngữ khác. Một phần mềm điển hình về tiếng Việt của chương trình này là Evtrans của Softex, dịch tự động từ tiếng Anh sang tiếng Việt và ngược lại, phần mềm từng được trang web vdict.com mua bản quyền, đây cũng là trang đầu tiên đưa ứng dụng này lên mạng. Tháng 10 năm 2008 có hai công ty tham gia vào lĩnh vực này cho ngôn ngữ tiếng Việt là công ty Lạc Việt (công ty phát hành từ điển Lạc Việt) và Google, một thời gian sau đó Xalo_vn cũng đưa ra dịch vụ tương tự.

Tìm kiếm thông tin (information retrieval): Đặt câu hỏi và chương trình tự tìm ra nội dung phù hợp nhất. Thông tin ngày càng đầy lên theo cấp số nhân, đặc biệt với sự trợ giúp của internet việc tiếp cận thông tin trở lên dễ dàng hơn bao giờ hết. Việc khó khăn lúc này là tìm đúng nhất thông tin mình cần giữa bề bộn tri thức và đặc biệt thông tin đó phải đáng tin cậy. Các máy tìm kiếm dựa trên giao diện web như Google hay Yahoo hiện nay chỉ phân tích nội dung rất đơn giản dựa trên tần suất của từ khoá và thứ hạng của trang và một số tiêu chí đánh giá khác để đưa ra kết luận, kết quả là rất nhiều tìm kiếm không nhận được câu trả lời phù hợp, thậm chí bị dẫn tới một liên kết không liên quan gì do thủ thuật đánh

lừa của các trang web nhằm giới thiệu sản phẩm (có tên tiếng Anh là SEO viết tắt của từ *search engine optimization*). Thực tế cho đến bây giờ chưa có máy tìm kiếm nào hiểu được ngôn ngữ tự nhiên của con người trừ trang www.ask.com được đánh giá là "hiểu" được những câu hỏi có cấu trúc ở dạng đơn giản nhất. Mới đây cộng đồng mạng đang xôn xao về trang Wolfram Alpha, được hứa hẹn là có khả năng hiểu ngôn ngữ tự nhiên của con người và đưa ra câu trả lời chính xác^[1]. Lĩnh vực này hứa hẹn tạo ra bước nhảy trong cách thức tiếp nhận tri thức của cả cộng đồng.

Tóm tắt văn bản: Từ một văn bản dài tóm tắt thành một văn bản ngắn hơn theo mong muốn nhưng vẫn chứa những nội dung thiết yếu nhất.

Khai phá dữ liệu (data mining) và phát hiện tri thức: Từ rất nhiều tài liệu khác nhau phát hiện ra tri thức mới. Thực tế để làm được điều này rất khó, nó gần như là mô phỏng quá trình học tập, khám phá khoa học của con người, đây là lĩnh vực đang trong giai đoạn đầu phát triển. Ở mức độ đơn giản khi kết hợp với máy tìm kiếm nó cho phép đặt câu hỏi để từ đó công cụ tự tìm ra câu trả lời dựa trên các thông tin trên web mặc cho việc trước đó có câu trả lời lưu trên web hay không (giống như trang Yahoo! hỏi và đáp, nơi chuyên đặt các câu hỏi để người khác trả lời), nói một cách nôm na là nó đã biết xử lý dữ liệu để trả lời câu hỏi của người sử dụng, thay vì máy móc đáp trả những gì chỉ có sẵn trong bộ nhớ.

2.3. OpenAI Whisper

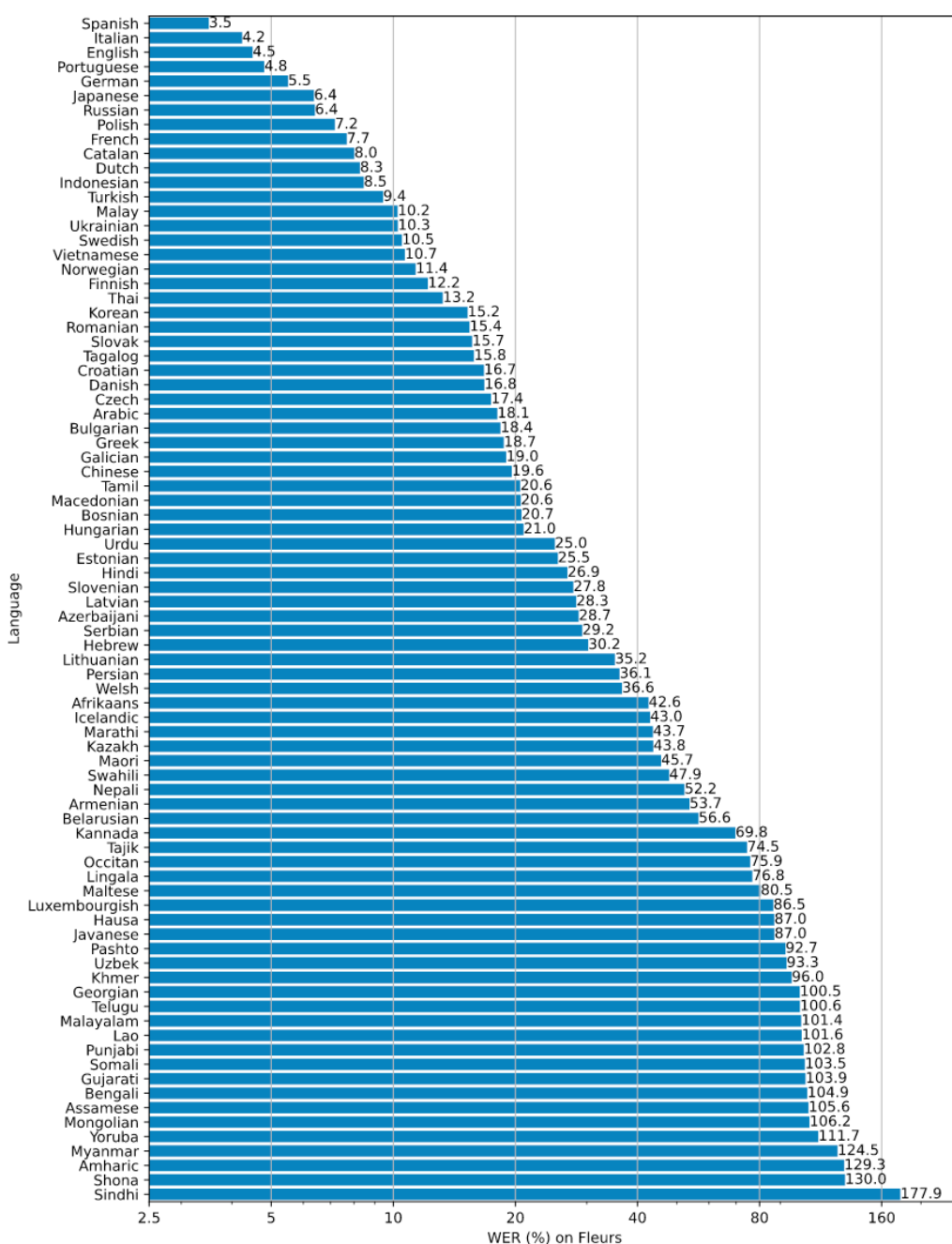
Whisper là hệ thống nhận dạng giọng nói tự động (ASR) được đào tạo dựa trên 680.000 giờ dữ liệu, giám sát đa ngôn ngữ, đa tác vụ thu thập từ web. Nó cho phép phiên âm bằng nhiều ngôn ngữ, cũng như dịch từ các ngôn ngữ đó sang tiếng Anh. Đây là một mô hình có nguồn mở làm nền tảng cho việc xây dựng các ứng dụng hữu ích và cho việc nghiên cứu sâu hơn xử lý giọng nói.

OpenAI mô tả Whisper như một mô hình biến đổi mã hóa-giải mã, một loại mạng thần kinh có thể sử dụng ngữ cảnh thu thập được từ dữ liệu đầu vào, tìm hiểu các liên kết tình huống, sau đó có thể được dịch sang ngôn ngữ cần thiết ở đầu ra của mô hình.

Phương thức hoạt động của Whisper: Âm thanh đầu vào được chia thành các đoạn dài 30 giây, chuyển đổi thành biểu đồ quang phổ log-Mel và sau đó chuyển vào bộ mã hóa. Một bộ giải mã được đào tạo để dự đoán văn bản chú thích tương ứng, trộn lẫn với các mã

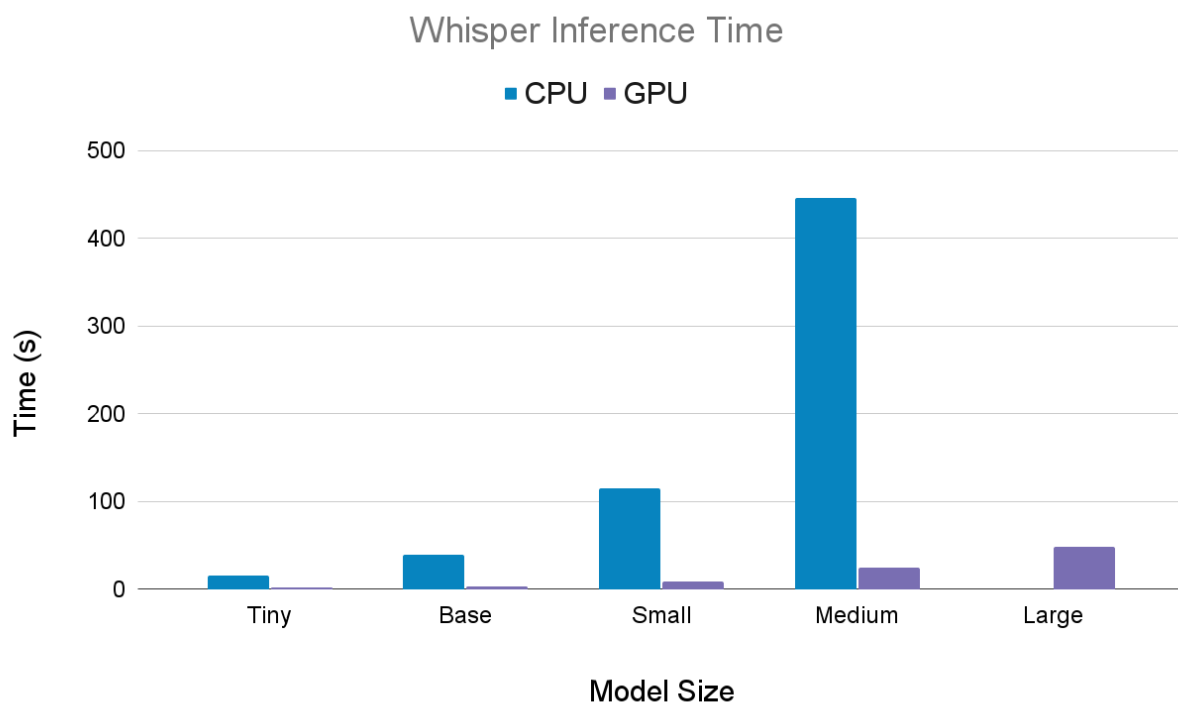
thông báo đặc biệt, chỉ đạo mô hình duy nhất thực hiện các tác vụ như nhận dạng ngôn ngữ, khoảng thời gian của các cụm từ, phiên âm giọng nói đa ngôn ngữ và dịch giọng nói sang tiếng Anh.

Mặc dù Whisper đạt được kết quả tiên tiến trên một số ngôn ngữ như tiếng Anh, tiếng Đức, tiếng Nhật, v.v., hiệu suất tương đối thấp đối với các ngôn ngữ khác.



Hình 2. Tỷ lệ lỗi của Whiper khi nhận dạng các ngôn ngữ

Whisper có năm mô hình: Tiny, Base, Small, Medium, Large với mức độ chính xác từ nhỏ đến lớn. Do đó, mô hình Large có độ chính xác tốt nhất. Whisper có thể được sử dụng trên cả CPU và GPU; tuy nhiên, thời gian triển khai trên CPU cực kỳ chậm khi sử dụng các mô hình có độ chính xác lớn, vì vậy chỉ nên chạy chúng trên GPU.



Hình 3. Đồ thị biểu diễn thời gian chạy của các mô hình trên CPU và GPU

2.4. Thư viện FFmpeg

FFmpeg là một dự án phần mềm miễn phí bao gồm một bộ phần mềm khổng lồ gồm các thư viện và chương trình để xử lý video, audio, multimedia files và streams. Cốt lõi của nó là chính chương trình FFmpeg, được thiết kế để xử lý các video và audio dựa trên command line và được sử dụng rộng rãi để chuyển đổi định dạng, chỉnh sửa cơ bản (cắt và ghép), chia tỷ lệ video, hiệu ứng hậu kỳ video. FFmpeg được xuất bản theo Giấy phép GNU Lesser General Public License 2.1+ hay GNU General Public License 2+ (dựa theo option nào được enabled).

FFmpeg được ứng dụng rất nhiều trong các phần mềm xử lý audio và video, thư viện của nó là thành phần core của ứng dụng media player như VLC, cũng là thành phần core trong quá trình xử lý video của YouTube và iTunes khi lưu trữ file.

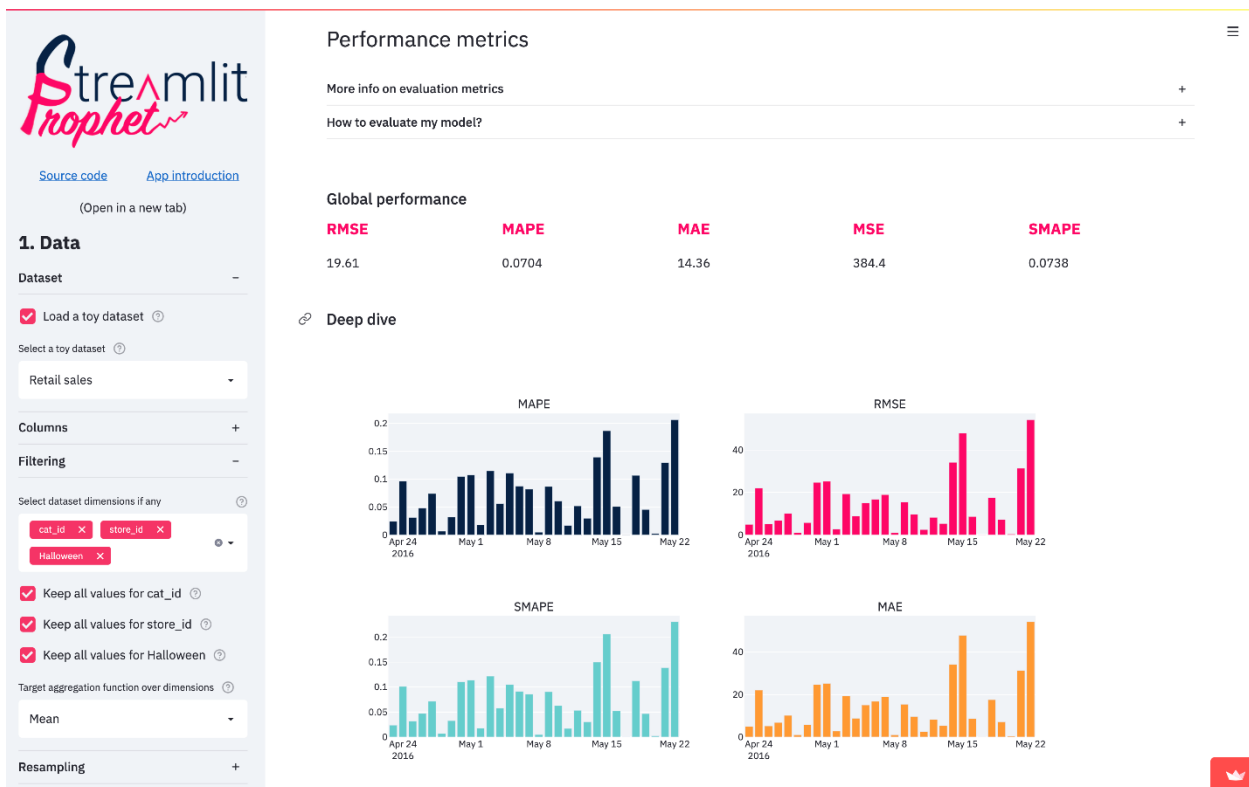
Các ứng dụng cơ bản của FFMpeg trong xử lý âm thanh, hình ảnh, video:

- Xem thông tin file
- Convert file sang định dạng khác
- Convert video sang audio
- Thay đổi độ phân giải video (kích thước video)
- Giảm dung lượng, chất lượng video
- Giảm dung lượng, chất lượng audio
- Xóa audio stream (xóa nhạc)
- Xóa video stream (xuất ảnh từ video)
- Thay đổi tỷ lệ khung hình
- Convert một phần cụ thể của video
- Tạo video với ảnh và file audio
- Trim video/audio
- Chia video thành nhiều phần
- Nối nhiều video thành một
- Thêm subtitle cho video
- Xem, test file video và audio
- Tăng, giảm tốc độ phát video
- Live stream video
- Xem thông tin về phiên bản FFMpeg, hướng dẫn về các tham số

2.5. Streamlit

Streamlit là một khung ứng dụng web dựa trên ngôn ngữ Python để trực quan hóa dữ liệu và phân tích kết quả một cách hiệu quả và linh hoạt hơn. Đây là một thư viện nguồn mở hỗ trợ các nhà khoa học dữ liệu, nhà phân tích dữ liệu phát triển bảng điều khiển trực quan hóa Machine Learning (ML) nhanh chóng. Chúng có thể xây dựng và triển khai các ứng dụng dữ liệu mạnh mẽ chỉ với một vài dòng code.

Streamlit là một open-source Python, nó giúp ta dễ dàng tạo một web app cho MachineLearning và Data Science. Đối với một người có rất ít kiến thức về HTML, CSS, JavaScript thì đây có lẽ là công cụ rất phù hợp để demo các sản phẩm AI. Chúng ta có thể test cục bộ, sau đó muốn deploy lên internet có thể dùng Heroku, Streamlit hay Ngrok.



Hình 4. Một web app được tạo nên từ Streamlit

CHƯƠNG 3: CHƯƠNG TRÌNH VÀ KẾT QUẢ

3.1. Chương trình chính

Khai báo thư viện

Model được sử dụng trong bài là “base” nhằm giúp chương trình triển khai nhanh hơn. Tuy nhiên điểm yếu của mô hình này là lỗi khi nhận dạng ngôn ngữ khá cao. Ngoài ra nếu sử dụng GOOGLE COLAB chúng ta có thể thay đổi sang mô hình “medium” để quá trình nhận dạng ngôn ngữ chính xác hơn.

```
import os
import sys
import subprocess
import whisper
from whisper.utils import write_vtt
model = whisper.load_model("base")
```

Chương trình con chuyển file video ‘mp4’ thành file audio định dạng ‘mp3’

```
def video2mp3(video_file, output_ext="mp3"):
    filename, ext = os.path.splitext(video_file)
    subprocess.call(["ffmpeg", "-y", "-i", video_file, f"{filename}.{output_ext}"],
                    stdout=subprocess.DEVNULL,
                    stderr=subprocess.STDOUT)
    return f"{filename}.{output_ext}"
```

Chương trình con chuyển đổi file audio thành văn bản và tạo phụ đề cho video

```
def transcribe(input_video,nn):
    audio_file = video2mp3(input_video)
    options = dict(language=nn,beam_size=5, best_of=5,fp16=False)
    transcribe_options = dict(task="transcribe", **options)
    result = model.transcribe(audio_file,**transcribe_options)
    output_dir = ''
    audio_path = audio_file.split(".")[0]
    with open(os.path.join(output_dir, audio_path + ".vtt"), "w",encoding='utf-8') as vtt:
        write_vtt(result["segments"], file=vtt)
    subtitle = audio_path + ".vtt"
    output_video = audio_path + "_subtitled.mp4"
    subprocess.call(["ffmpeg", "-i", input_video, "-vf", f"subtitles={subtitle}:force_style='FontName=Arial,FontSize=24'",
                    f"{output_video}"],
                    stdout=subprocess.DEVNULL,
                    stderr=subprocess.STDOUT)
    return output_video
```

3.2. Tạo giao diện cho phần mềm

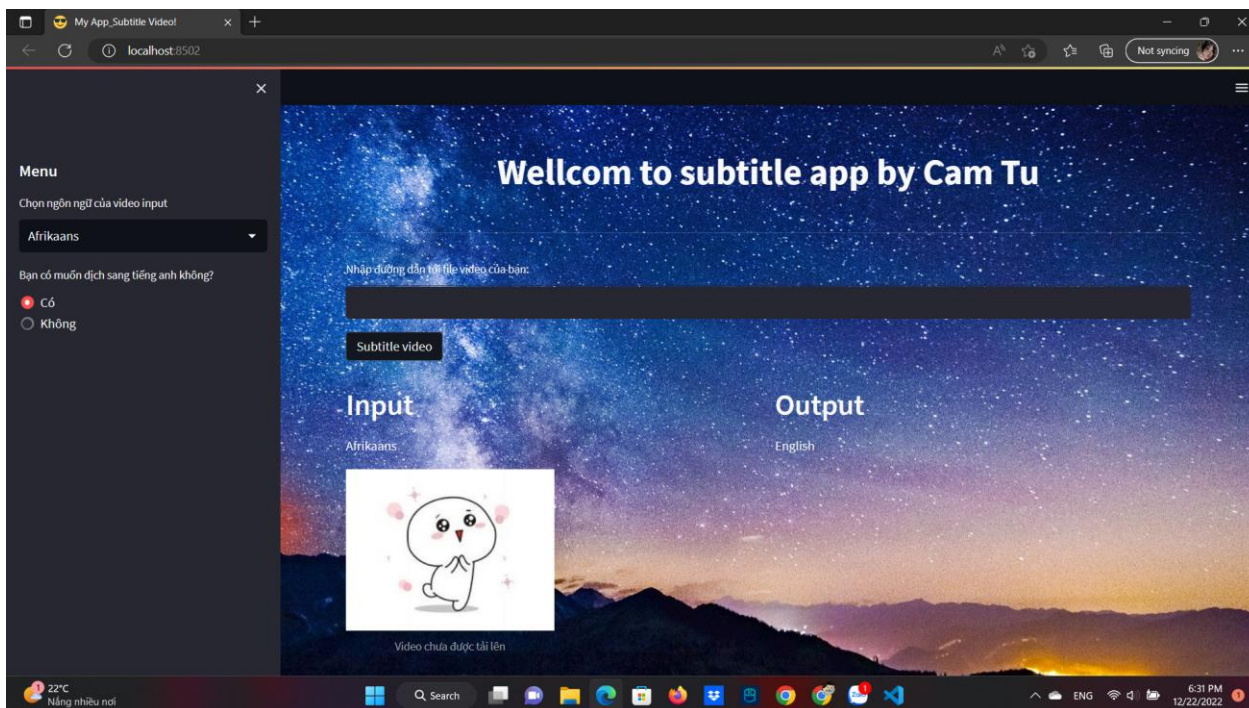
```
import streamlit as st
import time
from PIL import Image
st.set_page_config(
    page_title="My App_Subtitle Video!",
    page_icon="🎥",
    layout="wide"
)

page_bg_img = """
<style>
[data-testid="stAppViewContainer"]{
background-image: url("https://thuthuatnhanh.com/wp-content/uploads/2019/04/hinh-nen-thien-ha-milky-way.jpg");
background-size: cover;
background-position: top left;
background-repeat: no-repeat;
background-attachment: local;
}
</style>
"""
st.markdown(page_bg_img, unsafe_allow_html=True)
```

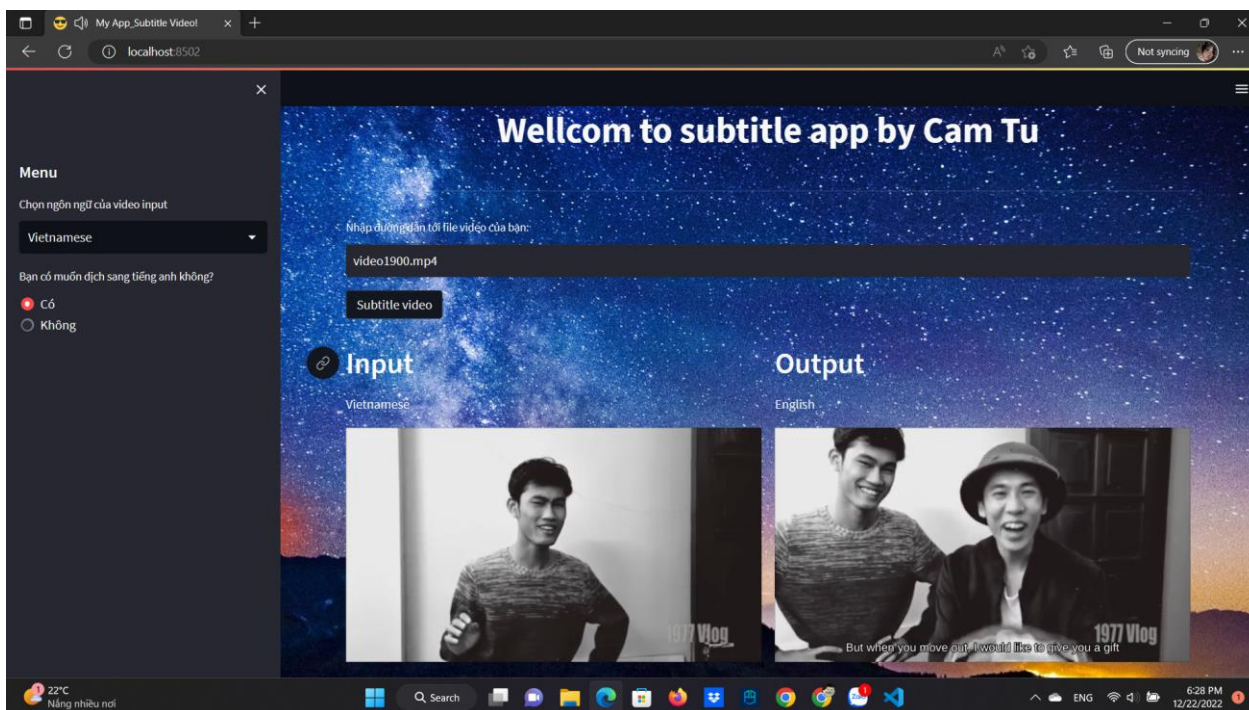
```
st.markdown("<h1 style='text-align: center;'>Wellcom to subtitle app by Cam Tu</h1>", unsafe_allow_html=True)
st.markdown("---", unsafe_allow_html=True)
url = st.text_input("Nhập đường dẫn tới file video của bạn:")
convert = st.button('Subtitle video')
with st.sidebar:
    st.header('Menu')
    language = st.selectbox("Chọn ngôn ngữ của video input", ('Afrikaans', ...
    choice = st.radio('Bạn có muốn dịch sang tiếng anh không?', ('Có', 'Không'))
col1, col2 = st.columns(2)
with col1:
    st.header('Input')
    st.write(language)
    if url == '':
        image = Image.open('icon.jpg')
        st.image(image, caption='Video chưa được tải lên')
    else:
        video_file = open(url, 'rb')
        video_input = video_file.read()
        st.video(video_input)
with col2:
    st.header('Output')
    if choice == 'Có':
        st.write('English')
    else:
        st.write(language)
    if convert:
        with st.spinner('Đang chạy'):
            time.sleep(10)
            if choice == 'Có':
                output = translate(url)
            else:
                output = transcribe(url, language.lower())
            video_file = open(output, 'rb')
            video_input = video_file.read()
            st.video(video_input)
```

3.3. Kết quả

Giao diện chính của web app



Kết quả chuyển đổi



Kết quả chuyển đổi từ tiếng việt sang tiếng anh



Kết quả phụ đề tiếng việt không chuyển đổi



CHƯƠNG 4: KẾT LUẬN

Ngày nay, cùng với sự phát triển của công nghệ thông tin cuộc sống cũng ngày càng trở nên dễ dàng hơn nhất là đối với người khuyết tật. Tạo phụ đề cho video có thể mang lại sự thuận tiện cho mọi người nhất là những nơi đông người. Ứng dụng của phần mềm tạo phụ vào giải trí, học tập, lẫn tìm tòi những điều mới sẽ giúp mọi người có thể tiếp thu nhiều điều mới hơn hoặc đơn giản là học tiếng anh. Phần mềm này có thể giúp bạn chuyển âm thanh trong video thành văn bản. Ngoài ra, phần mềm này cũng có thể dịch được sang tiếng anh.

Bên cạnh sự thuận lợi cho việc tạo phụ đề video thì web app này còn có một vài hạn chế. Khi sử dụng các mô hình có mức độ chính xác cao, thì thời gian xử lý video của ứng dụng này khá lâu, tiêu tốn nhiều thời gian của người sử dụng. Để tránh làm tiêu tốn thời gian của người dùng, ứng dụng này sử dụng mô hình “base” của Whisper với mức độ chính xác không cao. Ngoài ra, ứng dụng chỉ dịch được sang ngôn ngữ tiếng anh mà chưa có thể dịch sang bất kỳ ngôn ngữ nào khác.

DỮ LIỆU UP LÊN GITHUB

Github: https://github.com/CamTu22/PROJECT_CUOI_KI_AI.git

QRCODE:



TÀI LIỆU KHAM KHẢO

1. Công ty cổ phần công nghệ Viindoo, *Tìm hiểu về lịch sử phát triển của trí tuệ nhân tạo và những cột mốc đáng nhớ*, <https://viindoo.com/vi/blog/cong-nghe-16/lich-su-phat-trien-cua-tri-tue-nhan-tao-va-nhung-cot-moc-413>
2. Ryan O’Conner, *How to run OpenAI’s Whisper Speech Recognition Model*, AssemblyAI, <https://www.assemblyai.com/blog/how-to-run-openais-whisper-speech-recognition-model/>
3. FPT.Digital, *Xử lý ngôn ngữ tự nhiên: Công nghệ giúp máy tính hiểu và giao tiếp với con người*, [https://digital.fpt.com.vn/dxarticles/xu-ly-ngon-ngu-tu-nhien.html#:~:text=X%E1%BB%AD%20l%C3%BD%20ng%C3%B4n%20ng%E1%BB%AF%20t%E1%BB%B1%20nh%C3%AAn%20\(natural%20language%20processing\)%2C,th%E1%BB%B1c%20h%C3%B3a%20m%E1%BB%A5c%20ti%C3%AAn%20n%C3%A0y.](https://digital.fpt.com.vn/dxarticles/xu-ly-ngon-ngu-tu-nhien.html#:~:text=X%E1%BB%AD%20l%C3%BD%20ng%C3%B4n%20ng%E1%BB%AF%20t%E1%BB%B1%20nh%C3%AAn%20(natural%20language%20processing)%2C,th%E1%BB%B1c%20h%C3%B3a%20m%E1%BB%A5c%20ti%C3%AAn%20n%C3%A0y.)
4. Wikipedia, *Xử lý ngôn ngữ tự nhiên*, <https://vi.wikipedia.org/wiki/X%E1%BB%AD%20l%C3%BD%20ng%C3%B4n%20ng%E1%BB%AF%20t%E1%BB%B1%20nh%C3%AAn>
5. Nguyễn Thịnh Hưng, *FFmpeg và 20 câu lệnh cơ bản xử lý âm thanh, hình ảnh và video*, Viblo.asia, <https://viblo.asia/p/ffmpeg-va-20-cau-lenh-co-ban-xu-ly-am-thanh-hinh-anh-va-video-naQZRYBAKvx>
6. Streamlit.io, *A faster way to build and share data apps*, <https://streamlit.io/>