



Transforming audio and video into useable data

Jonathan Soma

Knight Chair in Data Journalism, Columbia University

Abraji Congress 2024

Sunday, July 14

js4571@columbia.edu

<https://bit.ly/abraji24-audio-video>

Hi, I'm Soma!

- Professor at Columbia University's Journalism School, where I teach with two data journalism programs
 - Data Journalism MS, a 12-month degree
 - Lede Program, 10-week summer intensive
- Teaching journalists machine learning and artificial intelligence for ten years, long before ChatGPT!
- You can just ignore me and watch YouTube videos

YouTube BE

Search

Home

Shorts

Subscriptions

You >

Your channel

History

Playlists

Your videos

Watch later

Liked videos

Your clips

Dearest friend, I daresay I have not partaken of food in ages. I'm positively famished

Practical AI for Investigative Journalism

Jonathan Soma

Playlist • Public • 6 videos • 669 views

A six session series held in April 2024 about real-life use cases for journalism in (mostly investigative) jour...more

▶ Play all

Sort

Dearest friend, I daresay I have not partaken of food in ages. I'm positively famished 2:09:33

Sorting documents (Practical AI for Investigative Journalism, Session 1)

Jonathan Soma • 1.1K views • Streamed 2 months ago

Large language models don't understand facts or concepts, they only know statistical probability 2:14:32

Structured, validated data from LLMs (Practical AI for Investigative Journalism, Session 2)

Jonathan Soma • 743 views • Streamed 2 months ago

Evaluating Verifiability in Generative Search Engines 1:45:26

Why generative AI is a dead end for responsible journalism (Practical AI for Journalism, Session 3)

Jonathan Soma • 581 views • Streamed 2 months ago

AI, Hugging Face and non-chatbot models (Practical AI for Journalism, Session 4)

Jonathan Soma • 836 views • Streamed 2 months ago

Context window 2:01:02

Local models/private AI (Practical AI for Investigative Journalism, Session 5)

Jonathan Soma • 451 views • Streamed 2 months ago

Transcription and audio models (Practical AI for Investigative Journalism, Session 6)

Jonathan Soma • 332 views • Streamed 2 months ago

ars TECHNICA

BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE

SIGN IN

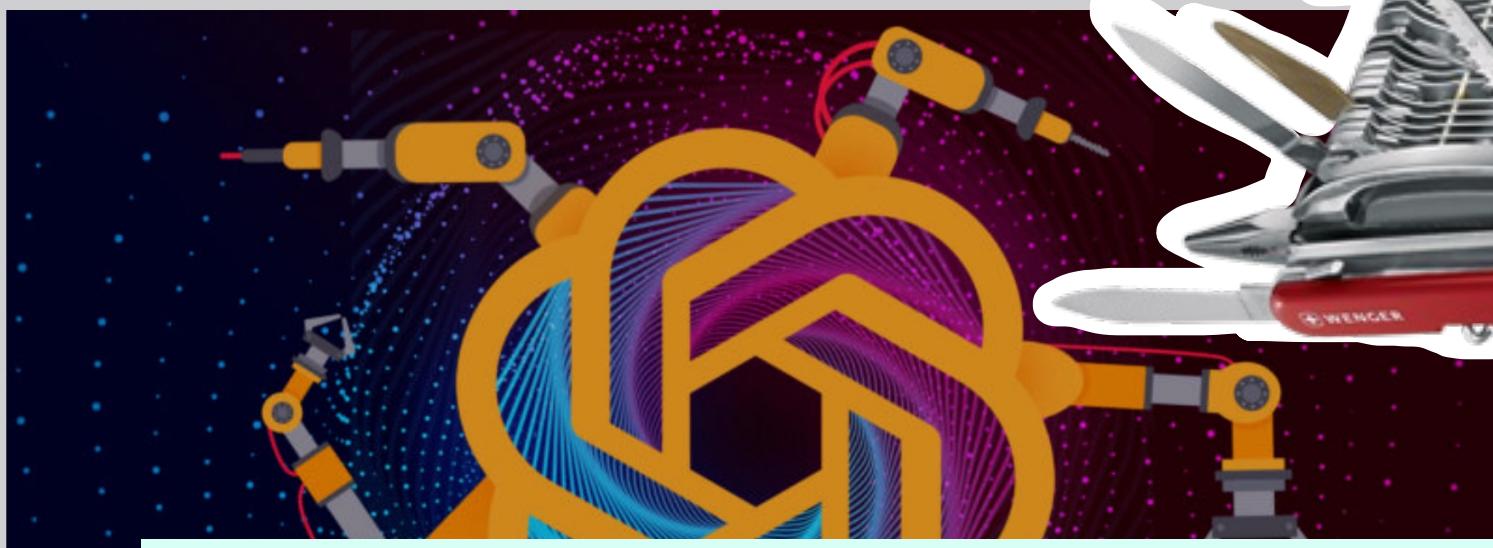
DO ANYTHING —

How ChatGPT turned generative AI into an anything tool

poor discoverability of potential

Until recently, AI models were specialized tools. Modern LLMs are different.

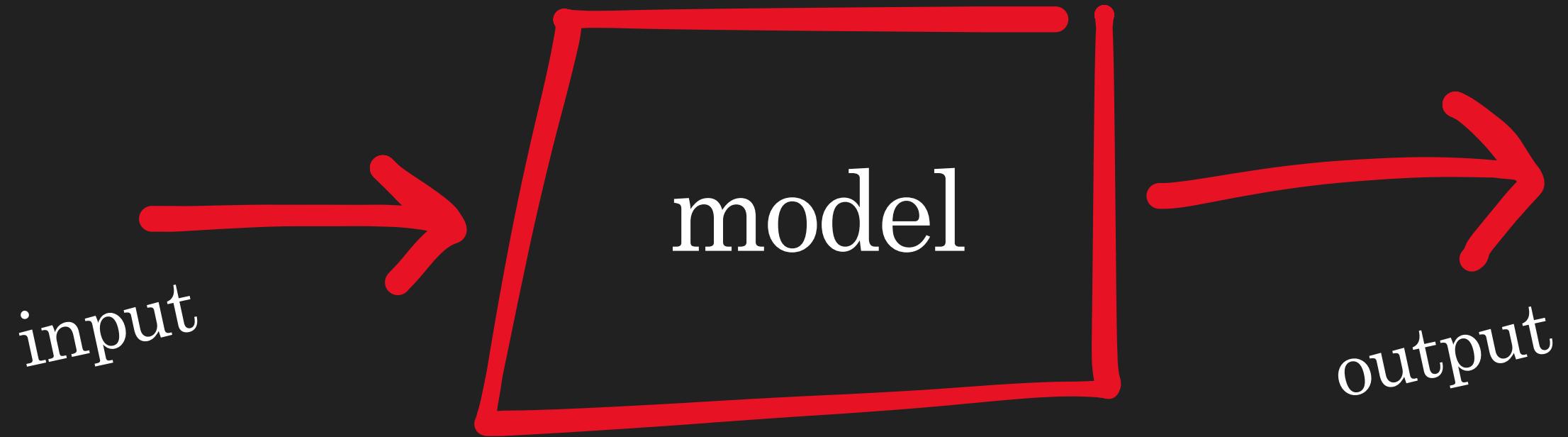
HAOMIAO HUANG - 8/23/2023, 7:30 AM



<https://arstechnica.com/ai/2023/08/how-chatgpt-turned-generative-ai-into-an-anyt...>

models
"tools"





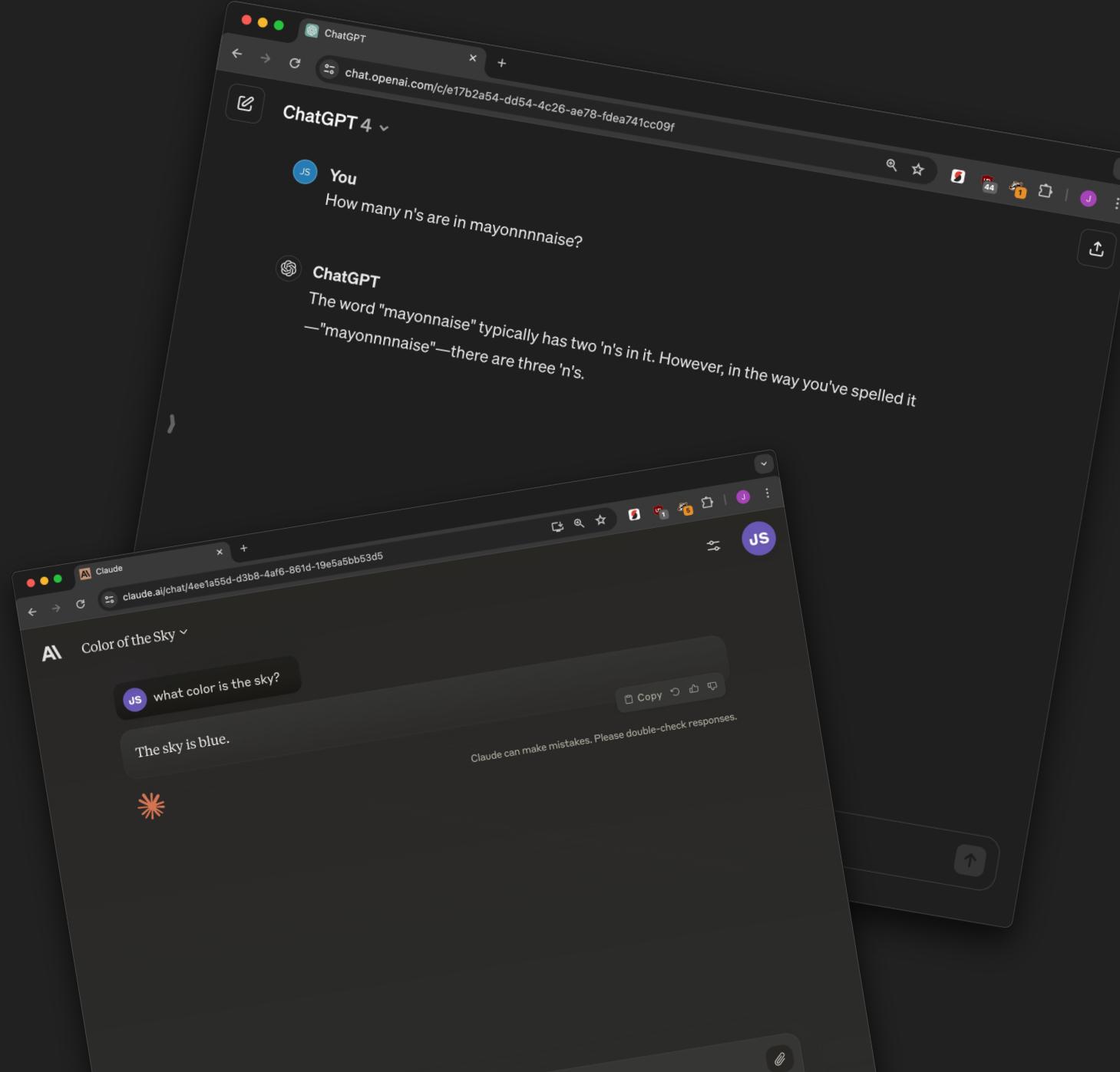
prompt



chat model



response



bill text



model



category

Louisiana state legislation (completed in class)

File Edit View Insert Format Data Tools Extensions Help

A1 A bill text

A	B	C	D	E
1 bill text > HLS 24RS-94 **[ENGROSSED](.underline)** > > 2024 Regular Session > > HOUSE BILL NO. 8 > > BY REPRESENTATIVES MELERINE AND HUGHES > > SCHOOLS/HIGH SCHOOL: Prohibits the use of an appeals proc > certain students > > HLS 24RS-88 **[REENGROSSED](.underline)** > > 2024 Regular Session > > HOUSE BILL NO. 9 > > BY REPRESENTATIVE ILLG > > RETIREMENT/DISTRICT ATTY: Provides for survivor benefits o > the > > HLS 24RS-53 **[REENGROSSED](.underline)** > > 2024 Regular Session > > HOUSE BILL NO. 12 > > BY REPRESENTATIVE JORDAN CRIME: Provides relative to the crime of nonconsensual disclosure private image > > 2024 Regular Session **[ENROLLED](.underline)** SENATE RE 27 > BY SENATORS PRESSLY, BASS, JENKINS AND SEABAUGH	ai_category education	education_prompt no, this legislation prohibits an appeals process for high school graduation for students who have not passed state assessments, so it is primarily related to education standards and requirements rather than issues that would directly impact crime reporting	education	about retirement? no
3	retirement	no, this bill relates to technical changes in retirement benefits for surviving spouses and children of District Attorneys and does not appear to have direct implications for crime reporting or related industries.	retirement, employment	yes
4	CRIME	maybe, this legislation deals with expanding the definition of nonconsensual disclosure of private images, which could potentially affect crime related to that offense.		

Claude found 2 cells update

bill text



Text
classification
model



category

Louisiana state legislation (completed in class)

File Edit View Insert Format Data Tools Extensions Help

A1 A bill text

A	B	C	D	E
	ai_category	education_prompt		about retirement?
1 > HLS 24RS-94 **[ENGROSSED](.underline)** > > 2024 Regular Session > > HOUSE BILL NO. 8 > > BY REPRESENTATIVES MELERINE AND HUGHES > > SCHOOLS/HIGH SCHOOL: Prohibits the use of an appeals proc > certain students > > HLS 24RS-88 **[REENGROSSED](.underline)** > > 2024 Regular Session > > HOUSE BILL NO. 9 > > BY REPRESENTATIVE ILLG > > RETIREMENT/DISTRICT ATTY: Provides for survivor benefits o > the > > HLS 24RS-53 **[REENGROSSED](.underline)** > > 2024 Regular Session > > HOUSE BILL NO. 12 > > BY REPRESENTATIVE JORDAN CRIME: Provides relative to the crime of nonconsensual disclosure private image > > 2024 Regular Session **[ENROLLED](.underline)** SENATE RE > 27 > BY SENATORS PRESSLY, BASS, JENKINS AND SEABAUGH	education	no, this legislation prohibits an appeals process for high school graduation for students who have not passed state assessments, so it is primarily related to education standards and requirements rather than issues that would directly impact crime reporting	education	no
2	retirement	no, this bill relates to technical changes in retirement benefits for surviving spouses and children of District Attorneys and does not appear to have direct implications for crime reporting or related industries.	retirement, employment	yes
3	CRIME	maybe, this legislation deals with expanding the definition of nonconsensual disclosure of private images, which could potentially affect crime related to that offense.		
4				

Claude found 2 cells update

Working

text

classification

summarization

entity extraction

poems about cats

sentiment analysis

audio

transcribing all of
your interviews



transcribing... other things?

audio

text

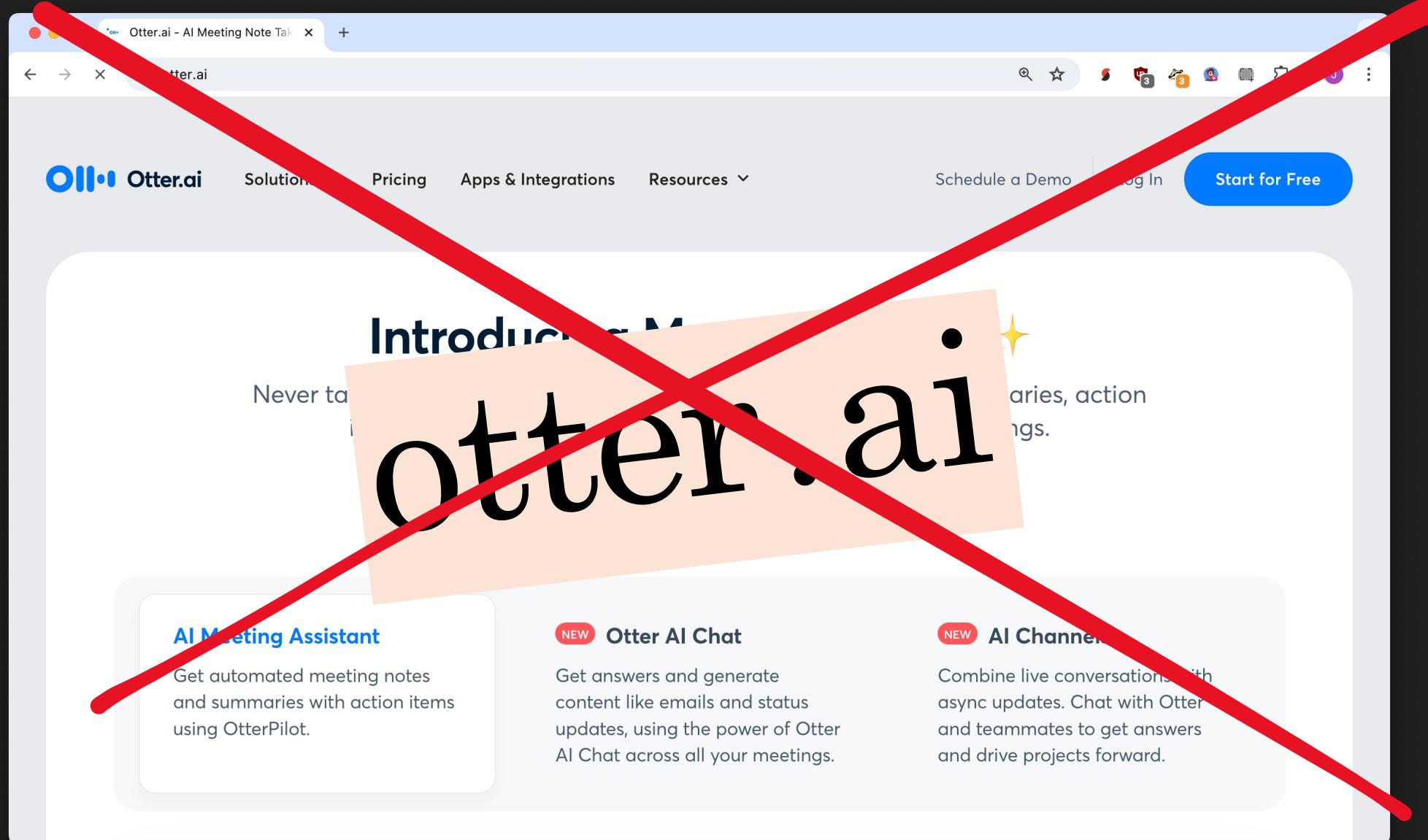
classification

summarization

entity
extraction

sentiment
analysis







Tasks Libraries Datasets Languages Licenses
Other

Filter Tasks by name

Multimodal

Image-Text-to-Text
Visual Question Answering
Document Question Answering

Computer Vision

Depth Estimation Image Classification
Object Detection Image Segmentation
Text-to-Image Image-to-Text
Image-to-Image Image-to-Video
Unconditional Image Generation
Video Classification Text-to-Video

Natural Language Processing

Text Classification Token Classification
Table Question Answering
Question Answering
Zero-Shot Classification Translation
Summarization Feature Extraction

Models 620,116

Filter by name

new Full-text search

↑↓ Sort: Trending

meta-llama/Meta-Llama-3-8B

Text Generation • Updated about 12 hours ago • ↓ 172k • 22

meta-llama/Meta-Llama-3-70B

Text Generation • Updated about 12 hours ago

meta-llama/Meta-Llama-3-70B

Text Generation • Updated about 12 hours ago

microsoft/Phi-3-mini-128k-instruct

Text Generation • Updated about 6 hours ago • 631

meta-llama/Meta-Llama-3-70B

Text Generation • Updated about 12 hours ago • ↓ 116k • 443

microsoft/Phi-3-mini-4k-instruct

Text Generation • Updated about 6 hours ago • 266

Instruct-v0.1

Text Generation • ↓ 17.5k • 453

dolphin-2.9-llama3-8b

Text Generation • Updated about 12 hours ago • ↓ 625 • 184

shenzi-wang/Llama3-8B-Chinese-Chat

Text Generation • Updated about 23 hours ago • 174

microsoft/Phi-3-mini-4k-instruct-gguf

Text Generation • Updated 1 day ago • 166

Hugging Face

Models 764,112

Multimodal

-  Image-Text-to-Text
 -  Visual Question Answering
 -  Document Question Answering
- Computer Vision
-  Depth Estimation
 -  Image Classification
 -  Object Detection
 -  Image Segmentation
 -  Text-to-Image
 -  Image-to-Text
 -  Image-to-Image
 -  Image-to-Video
 -  Unconditional Image Generation
 -  Video Classification
 -  Text-to-Video
 -  Zero-Shot Image Classification
 -  Mask Generation
 -  Zero-Shot Object Detection
 -  Text-to-3D
 -  Image-to-3D
 -  Image Feature Extraction

Natural Language Processing

-  Text Classification
 -  Token Classification
 -  Table Question Answering
 -  Question Answering
 -  Zero-Shot Classification
 -  Translation
 -  Summarization
 -  Feature Extraction
 -  Text Generation
 -  Text2Text Generation
 -  Fill-Mask
 -  Sentence Similarity
- Audio
-  Text-to-Speech
 -  Text-to-Audio
 -  Automatic Speech Recognition
 -  Audio-to-Audio
 -  Audio Classification
 -  Voice Activity Detection
- Tabular
-  Tabular Classification
 -  Tabular Regression

Audio models and ASR



audio



Automatic speech
recognition model



transcription

audio goes in, text comes out!

Elon Musk Sues OpenAI... nytimes.com/2024/03/01/technology/elon-musk-openai-sam-altman-la...

The New York Times

Harvesting Data for A.I. A.I. Data Race, Explained What Is 'Synthetic Data?' Key Figures in the Field A.I. Faces Quiz

Elon Musk Sues OpenAI and Sam Altman for Violating the Company's Principles

Musk said the prominent A.I. start-up had put profits and commercial interests ahead of seeking to benefit humanity.

Share full article    328



ChatGPT 4

JS You
How many n's are in mayonnnnase?

ChatGPT
The word "mayonnaise" typically has two 'n's in it. However, in the way you've spelled it—"mayonnnnase"—there are three 'n's.

AI

ChatGPT...
ChatGPT can make mistakes. Consider checking important information.

The image shows two browser windows side-by-side. The left window is the OpenAI website at openai.com/index/whisper, featuring the 'Introducing Whisper' section. The right window is the Hugging Face platform at huggingface.co/models?sort=trending&search=whisper, displaying a list of Whisper models.

OpenAI Landing Page (Left Window):

- Page Title: Introducing Whisper | OpenAI
- Page URL: openai.com/index/whisper
- Header Navigation: Research, Products, Safety, Company
- Date: September 21, 2022
- Main Section Header: Introducing Whisper
- Text: We've trained and are open-sourcing a neural net called Whisper that approaches human level robustness and accuracy on English speech recognition.
- Buttons: Read paper ↗, View code ↗, View model card ↗

Hugging Face Model Library (Right Window):

- Page Title: Models - Hugging Face
- Page URL: huggingface.co/models?sort=trending&search=whisper
- Header: Search models, datasets, users...
- Section: Models 292
- Search Bar: openai/whisper
- Buttons: new Full-text search, Add
- List of models:

 - openai/whisper-large-v3: Automatic Speech Recognition • Updated Feb 8 • ↓ 2.01M • ❤ 2.41k
 - openai/whisper-large-v2: Automatic Speech Recognition • Updated Feb 29 • ↓ 306k • ❤ 1.56k
 - openai/whisper-tiny: Automatic Speech Recognition • Updated Feb 29 • ↓ 216k • ❤ 180
 - openai/whisper-base: Automatic Speech Recognition • Updated Feb 29 • ↓ 359k • ❤ 159
 - openai/whisper-small: Automatic Speech Recognition • Updated Feb 29 • ↓ 368k • ❤ 152
 - openai/whisper-medium: Automatic Speech Recognition • Updated Feb 29 • ↓ 188k • ❤ 174
 - openai/whisper-tiny.en: Automatic Speech Recognition • Updated Jan 22 • ↓ 170k • ❤ 77

What is “open”?

- Weights: Your ability to run the model
- License: Restrictions on use of model
- Data available: What was the model trained on?

A screenshot of a GitHub search results page for repositories containing the word "whisper". The search bar at the top shows "whisper". The results are filtered by "Repositories" and show 12k results. The first result is "openai/whisper" with a description: "Robust Speech Recognition via Large-Scale Weak Supervision". It has 1.6M stars and was updated yesterday. The second result is "m-bain/whisperX" with a description: "WhisperX: Automatic Speech Recognition with Word-level Timestamps (& Diarization)". It has 12k stars and was updated 3 days ago. The third result is "ggerganov/whisper.cpp" with a description: "Port of OpenAI's Whisper model in C/C++". It has 31k stars and was updated 13 hours ago. The fourth result is "Const-me/Whisper" with a description: "High-performance GPGPU inference of OpenAI's Whisper automatic speech recognition (ASR) model". It has 7.2k stars and was updated on Oct 17, 2023.

A screenshot of a GitHub search results page for "Models" containing the word "whisper". The search bar at the top shows "whisper". The results show 10,649 models. The top result is "openai/whisper-large-v3" with a description: "Automatic Speech Recognition • Updated Feb 8 • 2.01M • 2.41k". The second result is "distil-whisper/distil-large-v3" with a description: "Automatic Speech Recognition • Updated 28 days ago • 199k • 106". The third result is "openai/whisper-large-v2" with a description: "Automatic Speech Recognition • Updated Feb 29 • 306k • 1.56k". The fourth result is "ggerganov/whisper.cpp" with a description: "Updated Dec 10, 2023 • 570". The fifth result is "BELLE-2/Belle-whisper-large-v3-zh" with a description: "Automatic Speech Recognition • Updated Mar 11 • 1.5k • 52". The sixth result is "kotoba-tech/kotoba-whisper-v1.0" with a description: "Automatic Speech Recognition • Updated 2 days ago • 3.26k • 35". The seventh result is "distil-whisper/distil-large-v2" with a description: "Automatic Speech Recognition • Updated Mar 21 • 25.6k • 484".

MacWhisper

goodsnooze.gumroad.com/l/mac... ☆

Private, high-quality transcription

100K+ users

Featured by AI

Whisper JAX - a Hugging Face

huggingface.co/spaces/sanchit-gandhi/whisper-jax

App Files Community 80

Spaces sanchit-gandhi/whisper-jax like 1.88k Running

Microphone Audio File YouTube

Whisper JAX: The Fastest Whisper API ⚡

potential? Cause it feels to me like previous generation learned. Yeah, of self-awareness, competing prior being clear about what you're optimizing one's out there trying to make a point their time or be ineffective. I think it we were previously optimizing for the moment. Or maybe we've entered uptake. We kind of are still inheriting following and we haven't upgraded our really want is a little bit different to be praised for or what the people are

MacWhisper

Jordi Bruin ★★★★★ 1268

Quickly and easily transcribe audio files into transcription technology Whisper. Whether other important audio, MacWhisper quickly into text.

€0 ↵

I want this!

EasyWhisper | Easily transcript

easywhisper.io

EASY WHISPER

Advantages Key Features Prices FAQs

ibe audio

for Mac OS mac OS

Enter YouTube URL...

My files



<https://bit.ly/abraj24-audio-video>

Activity: try an online
version of Whisper called
Whisper JAX

<https://huggingface.co/spaces/sanchit-gandhi/whisper-jax>



MacWhisper



Jordi Bruin

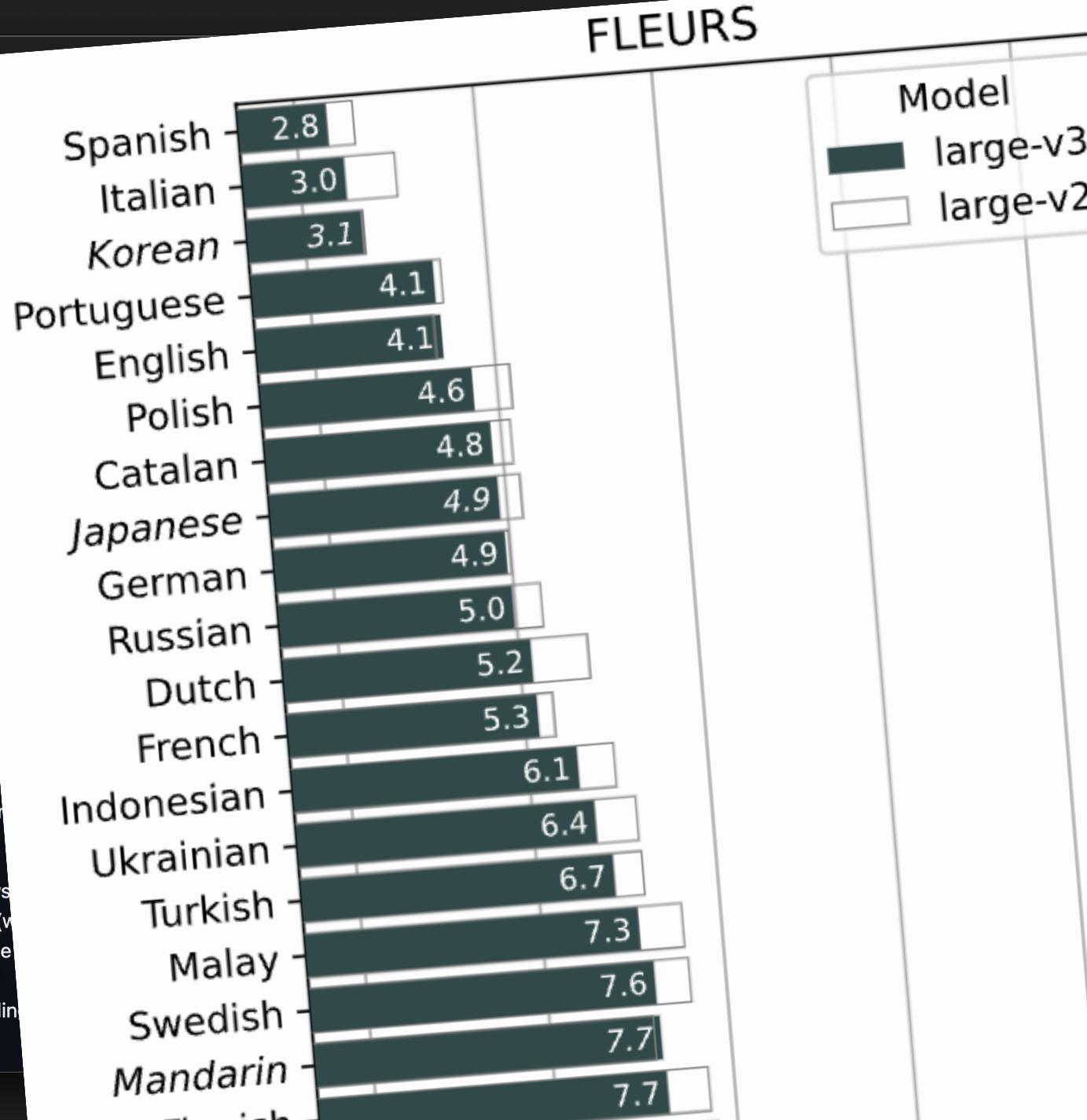
★★★★★ 1268 ratings

Quickly and easily transcribe audio files into text with OpenAI's state-of-the-art transcription technology Whisper. Whether you're recording a meeting, lecture, or other important audio, MacWhisper quickly and accurately transcribes your audio files into text.

Features

- Easily record and transcribe audio files on your Mac
- Just drag and drop audio files to get a high quality transcription
- Supported formats: mp3, wav, m4a, ogg, opus, mov and mp4 videos.
- Adjust whisper settings (beam search / greedy, beam size etc)
- Supports Tiny, Base and Small transcription qualities (and you can compare the transcription quality between all quality levels)





Available models and languages

There are five model sizes, four with English-only versions, offering speed and accuracy tradeoffs. Below are the names of the available models and their approximate memory requirements and inference speed relative to the large model; actual speed may vary on many factors including the available hardware.

Size	Parameters	English-only model	Multilingual model	Required VRAM	RAM
tiny	39 M	tiny.en	tiny	~1 GB	~100 MB
base	74 M	base.en	base	~1 GB	~100 MB
small	244 M	small.en	small	~2 GB	~100 MB
medium	769 M	medium.en	medium	~5 GB	~100 MB
large	1550 M	N/A	large	~10 GB	~100 MB

The .en models for English-only applications tend to perform better, especially for the tiny.en and base.en models. We observed that the difference becomes less significant for the small.en and medium.en models.

Whisper's performance varies widely depending on the language. The figure below shows the performance breakdown of large-v3 and large-v2 models by language, using WERs (word error rates) or CER (character error rates, shown in *italic*) evaluated on the Common Voice and Fleurs datasets. Additional WER/CER metrics corresponding to the other models and datasets can be found in Appendix D.1, D.2, and D.4 of [the paper](#), as well as the BLEU (Bilingual Evaluation Understudy) scores for translation in Appendix D.3.

Activity: I'll show you
MacWhisper

Using whisper from code

There are many, many, many ways

openai/whisper: Robust Speech-to-Text in Python · GitHub

github.com/openai/whisper

openai / whisper

Code Pull requests 56 Discussions Actions Security Insights

whisper Public

main Go to file + <> Code

ryanheisse and jongwook Skip silence arou... ba3f3cd · 5 months ago 136 Commits

.github/workflows Update test.yml 8 months ago

data initial commit 2 years ago

notebooks Use ndimage.median_filter inste... last year

tests large-v3 (#1761) 6 months ago

whisper Skip silence around hallucination... 5 months ago

.flake8 apply formatting with black (#10... last year

.gitattributes fix github language stats getting ... last year

.gitignore initial commit 2 years ago

.pre-commit-config.yaml Add .pre-commit-config.yaml (#1... 8 months ago

Watch 513 Fork

import whisper

```
model = whisper.load_model("base")
result = model.transcribe("audio.mp3")
print(result["text"])
```

60.6K stars

513 watching

7k forks

Report repository

Releases 10

v20231117 Latest on Nov 17, 2023

+ 9 releases

```
from faster_whisper import WhisperModel

device = "cuda" if torch.cuda.is_available() else "cpu"
compute_type = "int8"

model = WhisperModel("large-v3",
                     device=device,
                     compute_type=compute_type)

segments, info = model.transcribe("audio.mp3")
```

SYSTRAN/faster-whisper: Fas x
github.com/SYSTRAN/faster-whisper/

README MIT license

CI passing pypi package 1.0.1

Faster Whisper transcription with CTranslate2

Whisper model using [CTranslate2](#), which is a fast alternative to [openai/whisper](#) for the same accuracy while using less memory and with 8-bit quantization on both CPU and GPU.

It is required to transcribe [13 minutes](#) of audio

Used by 1.2k

C2A + 1,157

Contributors 31

+ 17 contributors

Languages

Python 100.0%

	Model	Format	size	Time	Max. GPU memory	Max. CPU memory
	openai/whisper	fp16	5	4m30s	11325MB	9439MB
	faster-whisper	fp16	5	54s	4755MB	3244MB
	faster-whisper	int8	5	59s	3091MB	3117MB

Vaibhavs10/insanely-fast-whisper · GitHub

github.com/Vaibhavs10/insanely-fast-whisper?tab=readme-ov-file

pyproject.toml Update pyproject.toml 4 months ago

README Apache-2.0 license

Insanely Fast Whisper

An opinionated CLI to transcribe Audio files w/ Whisper on-device! Powered by 😊 *Transformers*, Optimum & flash-attn

TL;DR - Transcribe 150 minutes (2.5 hours) of audio in less than 98 seconds - with [OpenAI's Whisper Large v3](#). Blazingly fast transcription is now a reality! ⚡



Not convinced? Here are some benchmarks we ran on a Nvidia A100 - 80GB 🤗

Optimisation type	Time to Transcribe (150 mins of Audio)
large-v3 (Transformers) (fp32)	~31 (31 min 1 sec)
large-v3 (Transformers) (fp16 + batching [24] +	5 (5 min 0 sec)

```
import torch
from transformers import pipeline
from transformers.utils import is_flash_attn_2_available

attn = "flash_attention_2" if is_flash_attn_2_available() else "sdpa"
pipe = pipeline(
    "automatic-speech-recognition",
    model="openai/whisper-large-v3",
    torch_dtype=torch.int8,
    device="mps",
    model_kwargs={"attn_implementation": attn},
)

outputs = pipe(
    "audio.mp3",
    batch_size=24,
    return_timestamps=True,
)

outputs
```



```
from whisper_jax import FlaxWhisperPipline
import jax.numpy as jnp

# instantiate pipeline with bfloat16 and enable batching
pipeline = FlaxWhisperPipline("openai/whisper-large-v2",
                               dtype=jnp.bfloat16,
                               batch_size=16)

# transcribe and return timestamps
outputs = pipeline("audio.mp3",
                   task="transcribe",
                   return_timestamps=True)
```

The screenshot shows a web browser window displaying the GitHub repository for "Whisper JAX" by sanchit-gandhi. The repository URL is github.com/sanchit-gandhi/whisper-jax. The page includes the README, Apache-2.0 license, and a Packages section indicating no packages have been published. The Contributors section lists four individuals: sanchit-gandhi, pcuenca, eltoclear, and tdzz1102. The Languages section shows that the code is primarily written in Jupyter Notebook (98.2%) and Python (1.8%). The README contains instructions for running the pipeline, installing via pip, and updating the package.

README Apache-2.0 license

Whisper JAX

This repository contains optimised JAX code for OpenAI's [Whisper Model](#), largely built on the 🤗 Hugging Face Transformers library. Compared to OpenAI's PyTorch code, Whisper JAX is faster and more memory efficient. It is currently available on Cloud TPUs, and can be run standalone (see [Pipeline Usage](#)).

To run the pipeline on a Cloud TPU, refer to the following Kaggle notebook. The transcription process takes approx 30 sec:

on the Hugging Face Hub:

version 0.4.5. Installation assumes that you already have Python installed on your device. You can do so using the official [installation](#) guide. Once you have Python installed, Whisper JAX can be installed through pip:

```
git clone https://github.com/sanchit-gandhi/whisper-jax.git
```

To update the Whisper JAX package to the latest version, simply run:

Packages
No packages published

Contributors 4

- sanchit-gandhi Sanchit Gandhi
- pcuenca Pedro Cuenca
- eltoclear Ikko Eltoclear Ashimine
- tdzz1102

Languages

- Jupyter Notebook 98.2%
- Python 1.8%

[m-bain/whisperX](https://github.com/m-bain/whisperX) · [WhisperX](#)

git clone https://github.com/m-bain/whisperX.git

github.com/m-bain/whisperX

[Readme](#) [BSD-4-Clause license](#)

WhisperX

Stars 9.1k Issues 398 open License BSD-4-Clause Arxiv 2303.00747 Tweet

This repository provides fast automatic speech recognition (70x realtime with large-v2) with word-level timestamps and speaker diarization.

- ⚡ Batched inference for 70x realtime transcription using whisper large-v2
- ✍ [faster-whisper](#) backend, requires <8GB gpu memory for large-v2 with beam_size=5
- ⌚ Accurate word-level timestamps using wav2vec2 alignment
- 👯 Multispeaker ASR using speaker diarization from [pyannote-audio](#) (speaker ID labels)
- 🗣 VAD preprocessing, reduces hallucination & batching with no WER degradation

Whisper is an ASR model [developed by OpenAI](#), trained on a large dataset of diverse audio. It does produce highly accurate transcriptions, the corresponding timestamps are at the utterance level, not per word, and can be inaccurate by several seconds. OpenAI's whisper does not natively support batching.

Phoneme-Based ASR A suite of models finetuned to recognise the smallest unit of speech distinguishing one word from another, e.g. the element p in "tap". A popular example model is [wav2vec2.0](#).

Forced Alignment refers to the process by which orthographic transcriptions are aligned to audio recordings to automatically generate phone-level segmentation.

+ 6 releases

Sponsor this project

<https://www.buyameacoffee.com/maxh...>

```
import whisperx
import torch

device = "cuda" if torch.cuda.is_available() else "cpu"
compute_type = "int8"

model = whisperx.load_model("base",
                            device,
                            compute_type=compute_type)
audio = whisperx.load_audio("audio.mp3")
result = model.transcribe(audio, batch_size=4)
```

```
git clone https://github.com/ggerganov/whisper.cpp.git
bash ./whisper.cpp/models/download-ggml-model.sh large-v3
cd whisper.cpp && make
ffmpeg -i ./audio.mp3 -acodec "pcm_s16le" -ac 1 -ar 16000
./audio.wav
./whisper.cpp/main \
  -m whisper.cpp/models/ggml-large-v3.bin \
  -f ./adr-11yo-bf-dc.wav \
  -of whispercpp.large-v3 --output-txt
```

```
submodule update --init --recursive
```

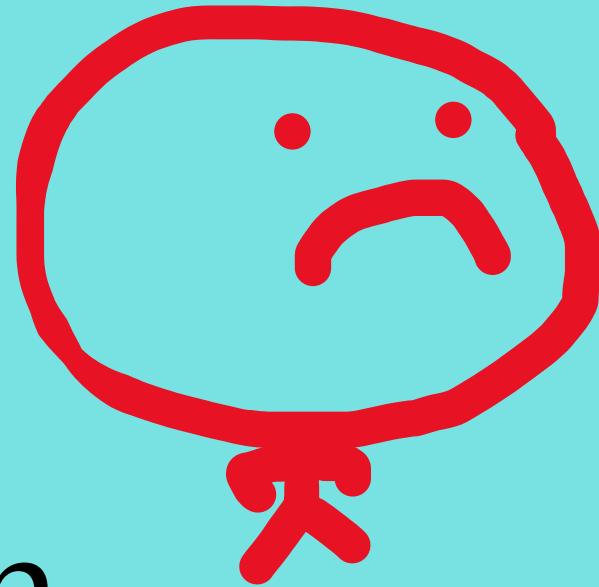
Build the wheel:

```
# Option 1: using pypa/build
python3 -m build -w

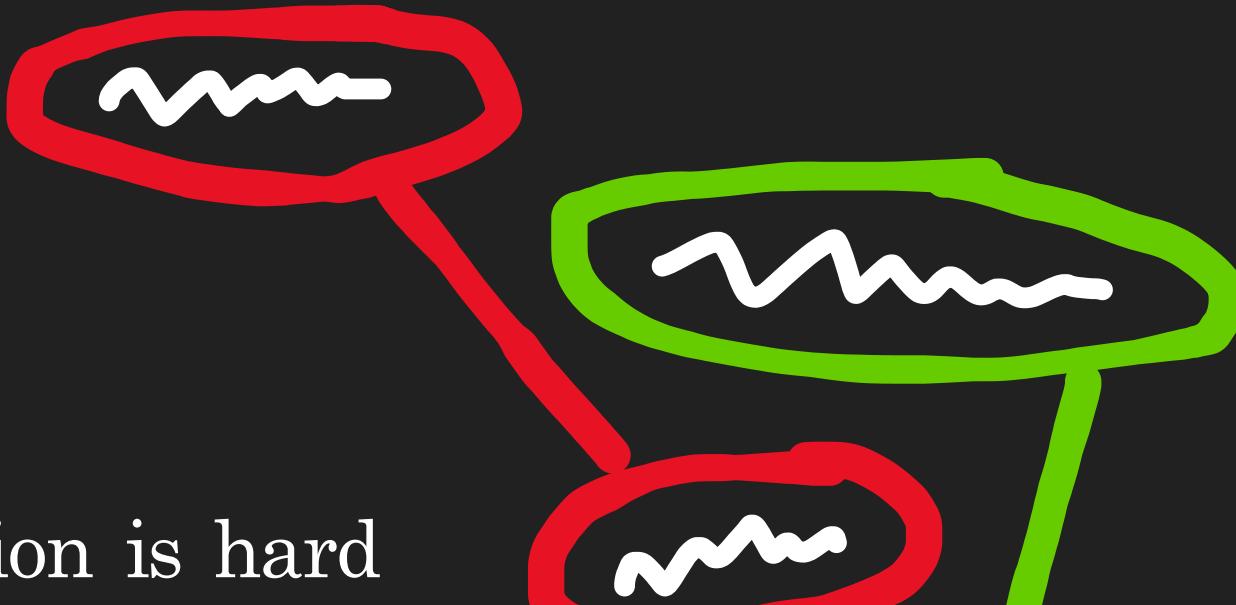
# Option 2: using bazel
./tools/bazel build //:whispercpp_wheel
```

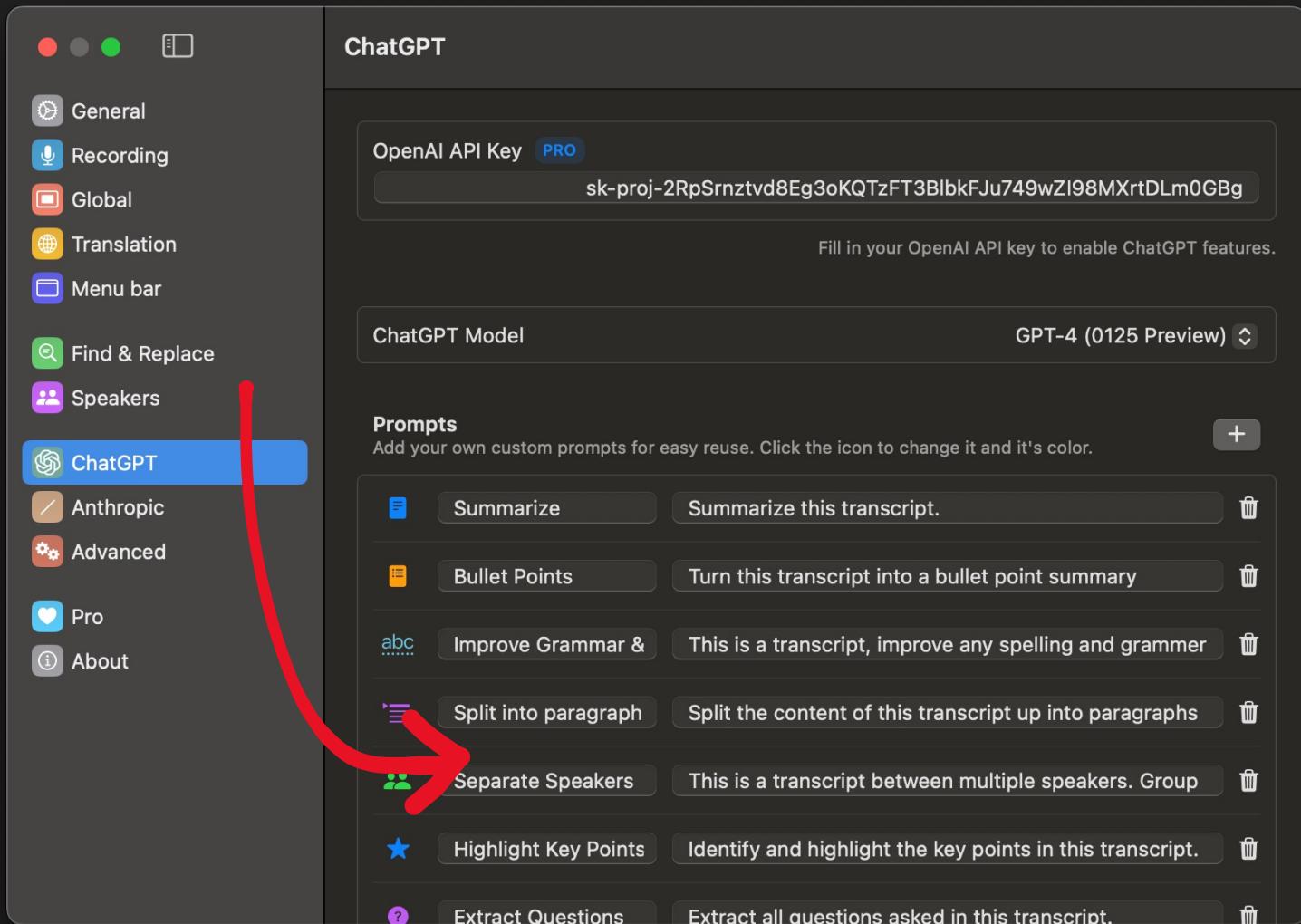
Speaker diarization

A fancy name for “separating individual speakers”



Diarization is hard





This is a transcript between multiple speakers. Group sentences said by the same person together and separate each section with a new line.

pyannote/pyannote-audio: Ne +

github.com/pyannote/pyannote-audio

README Code of conduct MIT license

Using `pyannote.audio` open-source toolkit in production? Make the most of it thanks to our [consulting services](#).

pyannote.audio speaker diarization toolkit

`pyannote.audio` is an open-source toolkit written in Python for speaker diarization. Based on [PyTorch](#) machine learning framework, it comes with state-of-the-art [pretrained models and pipelines](#), that can be further finetuned to your own data for even better performance.

What is speaker diarization?

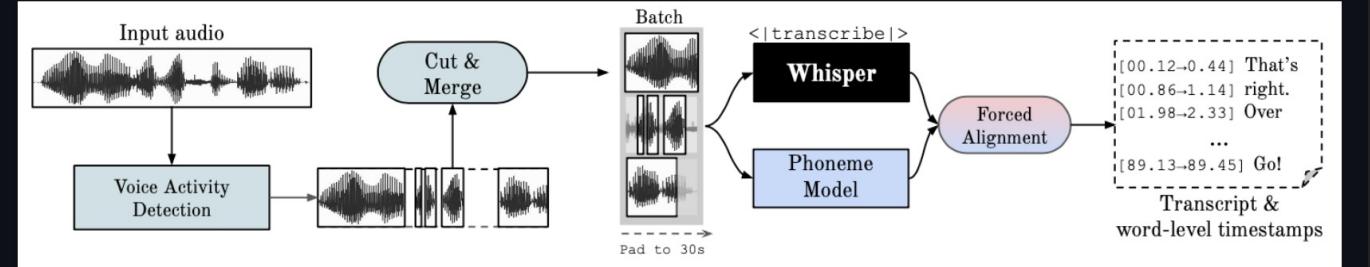


TL;DR

1. Install `pyannote.audio` with `pip install pyannote.audio`
2. Accept `pyannote/segmentation-3.0` user conditions
3. Accept `pyannote/speaker-diarization-3.1` user conditions

WhisperX

Stars 9.1k issues 398 open license BSD-4-Clause Arxiv 2303.00747 Tweet



Speaker Diarization

To enable Speaker Diarization, include your Hugging Face access token (read) that you can generate from [Here](#) after the `--hf_token` argument and accept the user agreement for the following models: [Segmentation](#) and [Speaker-Diarization-3.1](#) (if you choose to use Speaker-Diarization 2.x, follow requirements [here](#) instead.)

Note

As of Oct 11, 2023, there is a known issue regarding slow performance with pyannote/Speaker-Diarization-3.0 in whisperX. It is due to dependency conflicts between faster-whisper and pyannote-audio 3.0.0. Please see [this issue](#) for more details and potential workarounds.

Hugging Face Search models, datasets, users...

pyannote/segmentation-3.0 like 140

Voice Activity Detection, pyannote.audio, PyTorch, pyannote, pyannote-audio-model, audio, voice, speech, speaker, speaker-diarization, speaker-change-detection, speaker-segmentation, overlapped-speech-detection, resegmentation, License: mit

Use in pyannote.audio

Model card Files Community 1

You need to agree to share your contact information to access this model

This repository is publicly accessible, but you have to accept the conditions to access its files and content.

The collected information will help acquire a better knowledge of pyannote.audio userbase and help its maintainers improve it further. Though this model uses MIT license and will always remain open-source, we will occasionally email you about premium models and paid services around pyannote.

Log in or Sign Up to review the conditions and access this model content.

Downloads last month 8,142,219

Inference API (serverless) has been turned off for this model.

Spaces using pyannote/segmentation-3.0 11

- avans06/whisper-webui-translate
- Valor/whisper-webui-translate
- yvankob/whisper-webui-translate
- Maximofn/subtify
- SoybeanMilk/whisper-webui-translate
- Nymbo/whisper-webui-translate
- dandrade/whisper-webui-translate

Using this open-source model in production?

Hugging Face Search models, datasets, users...

pyannote/speaker-diarization-3.1 like 205

Automatic Speech Recognition, pyannote.audio, pyannote, pyannote-audio-pipeline, audio, voice, speech, speaker, speaker-diarization, speaker-change-detection, voice-activity-detection, overlapped-speech-detection, Inference Endpoints, arxiv:2111.14448, arxiv:2012.01477

License: mit

Deploy Use in pyannote.audio

Model card Files

You need to agree to share your contact information to access this model

This repository is publicly accessible, but you have to accept the conditions to access its files and content.

The collected information will help acquire a better knowledge of pyannote.audio userbase and help its maintainers improve it further. Though this pipeline uses MIT license and will always remain open-source, we will occasionally email you about premium pipelines and paid services around pyannote.

Log in or Sign Up to review the conditions and access this model content.

Downloads last month 5,683,669

Inference API Automatic Speech Recognition

Browse for file or Record from browser Compute

This model could not be loaded on Inference API (serverless). JSON Output Maximize

The screenshot shows the Hugging Face settings/tokens page. On the left, a sidebar lists various account settings, with 'Access Tokens' selected. The main area displays 'User Access Tokens' with three entries: 'myself' (READ), 'another' (READ), and 'roboflow' (WRITE). A warning message at the top right of this section advises against posting tokens publicly. Two red arrows point from the 'Manage' dropdowns of the first two token entries to the corresponding lines of code in the code editor on the right.

User Access Tokens

Access tokens programmatically authenticate your identity to the Hugging Face Hub, allowing applications to perform specific actions specified by the scope of permissions granted. Visit the [documentation](#) to discover how to use them.

⚠ Make sure to never post your tokens publicly

User	Scope	Action
myself	READ	Manage Show
another	READ	Manage Show
roboflow	WRITE	Manage Show

```
# https://huggingface.co/settings/tokens
# https://huggingface.co/pyannote/segmentation-3.0
# https://huggingface.co/pyannote/speaker-diarization-3.1

diarize_model = whisperx.DiarizationPipeline(use_auth_token='hf_ABCD1234',
                                             device=device)

# add min/max number of speakers if known
diarize_segments = diarize_model(audio)

result = whisperx.assign_word_speakers(diarize_segments, result)
```

Which to pick? It depends.

- CPU vs Apple Metal (M1/M2/M3) vs NVIDIA GPU (fancy video card)
 - ``cuda`` will be NVIDIA, ``cpu`` or ``mps`` will be Apple
 - ``float16`` will be NVIDIA, ``int8`` will be Apple
 - *And so many other settings to change*
- Diarization
 - No diarization: whisper, faster-whisper
 - pyannote: WhisperX, insanely-fast-whisper
 - tinydiarize and stereo diarization: Whispercpp

I like insanely-fast-whisper and WhisperX

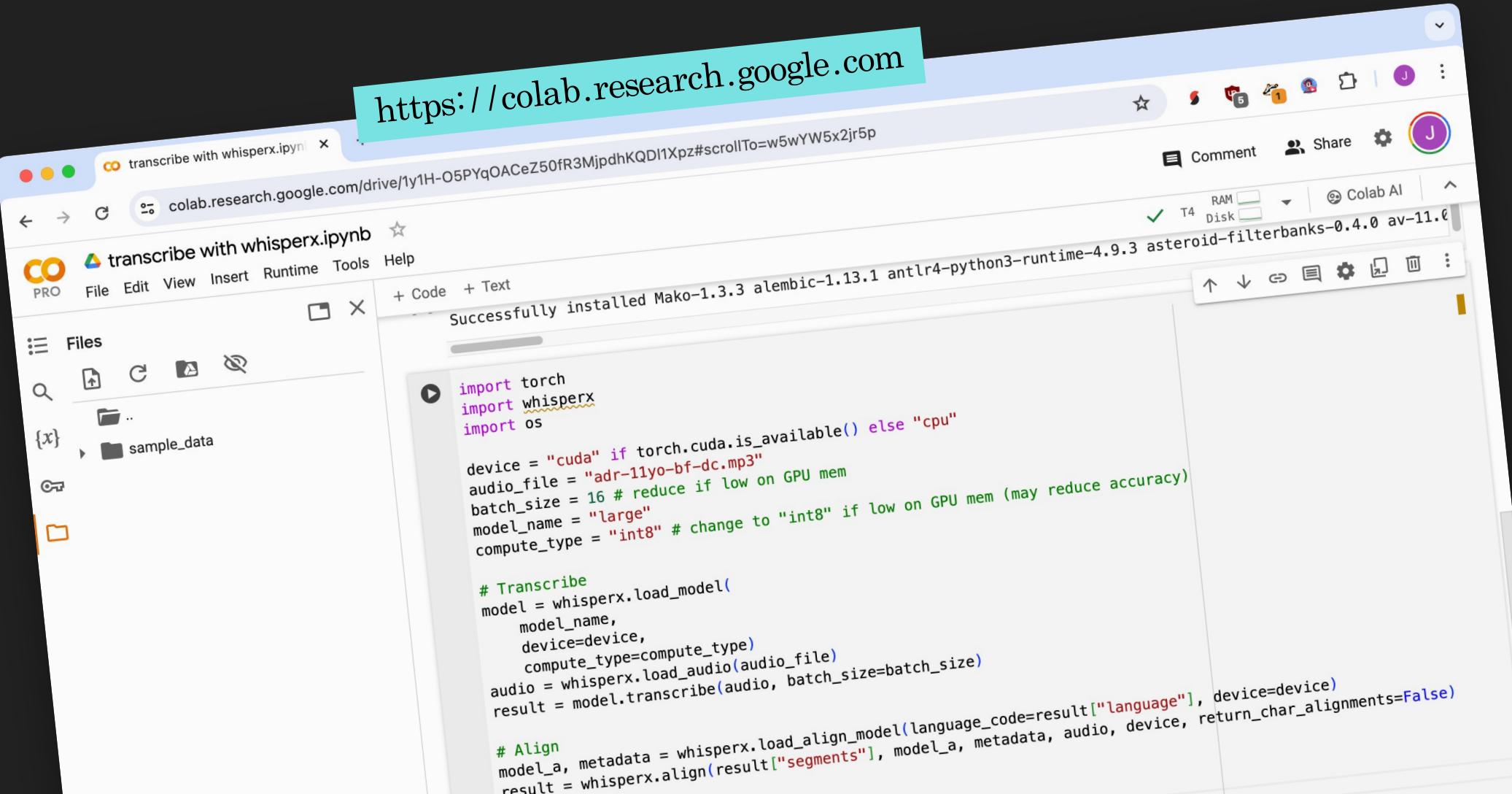
and MacWhisper!



```
!insanely-fast-whisper \
--file-name audio.mp3 \
--device-id mps \
--transcript-path output.json \
--hf_token hf_HUGGINGFACE_TOKEN
```



Or use the cloud



The screenshot shows a Google Colab notebook titled "transcribe with whisperx.ipynb". The URL in the address bar is <https://colab.research.google.com>. The notebook interface includes a sidebar with "Files" and a toolbar with "Comment", "Share", and "Colab AI" buttons. A status bar at the top right indicates "T4 RAM" and "Disk" usage. The main code cell contains the following Python code:

```
import torch
import whisperx
import os

device = "cuda" if torch.cuda.is_available() else "cpu"
audio_file = "adr-11yo-bf-dc.mp3"
batch_size = 16 # reduce if low on GPU mem
model_name = "large"
compute_type = "int8" # change to "int8" if low on GPU mem (may reduce accuracy)

# Transcribe
model = whisperx.load_model(
    model_name,
    device=device,
    compute_type=compute_type)
audio = whisperx.load_audio(audio_file)
result = model.transcribe(audio, batch_size=batch_size)

# Align
model_a, metadata = whisperx.load_align_model(language_code=result["language"], device=device)
result = whisperx.align(result["segments"], model_a, metadata, audio, device, return_char_alignments=False)
```

The code cell output shows the message: "Successfully installed Mako-1.3.3 alembic-1.13.1 antlr4-python3-runtime-4.9.3 asteroid-filterbanks-0.4.0 av-11.0".

Models 10,696

whisper

new Full-text search

Add filters Sort: Trending

openai/whisper-large-v3
Automatic Speech Recognition • Updated Feb 8 • 2.06M • 2.41k

distil-whisper/distil-large-v3
Automatic Speech Recognition • Updated 29 days ago • 2021k • 108

ggerganov/whisper.cpp
Updated Dec 10, 2023 • 571

BELLE-2/Belle-whisper-large-v3-zh
Automatic Speech Recognition • Updated Mar 11 • 1.44k • 53

openai/whisper-large-v2
Automatic Speech Recognition • Updated Feb 29 • 321k • 1.56k

distil-whisper/distil-large-v2
Automatic Speech Recognition • Updated Mar 21 • 25.4k • 484

Systran/faster-whisper-large-v3
Automatic Speech Recognition • Updated Nov 23, 2023 • 272k • 121

kotoba-tech/kotoba-whisper-v1.0
Automatic Speech Recognition • Updated 3 days ago • 3.26k • 35

You can also pick other models

6.3x faster than large-v3,

Model	Params / M	Rel. Latency	Short-Form	Sequential Long-Form	Chunked Long-Form
<u>large-v3</u>	1550	1.0	8.4	10.0	11.0
<u>distil-large-v3</u>	756	6.3	9.7	10.8	10.9
<u>distil-large-v2</u>	756	5.8	10.1	15.6	11.6



<https://bit.ly/abraj24-audio-video>

Activity: Let's try it out
on Google Colab

What else is out there?

A screenshot of a web browser displaying a YouTube video player. The video shows two women speaking at a podium with an American flag in the background. A transcript is overlaid on the right side of the screen, showing automatically generated subtitles for the video. The transcript includes timestamps and text such as "what they did on the weekends they went and they peel potatoes and". Below the video player, the YouTube channel page for "Louisiana Special School District Special Board Meeting 01/24/2023" is visible, showing 180 views and a timestamp of 1:27:20 / 3:28:30. The transcript overlay has a dark background and white text, with a play button icon indicating it is paused.

YouTube transcripts

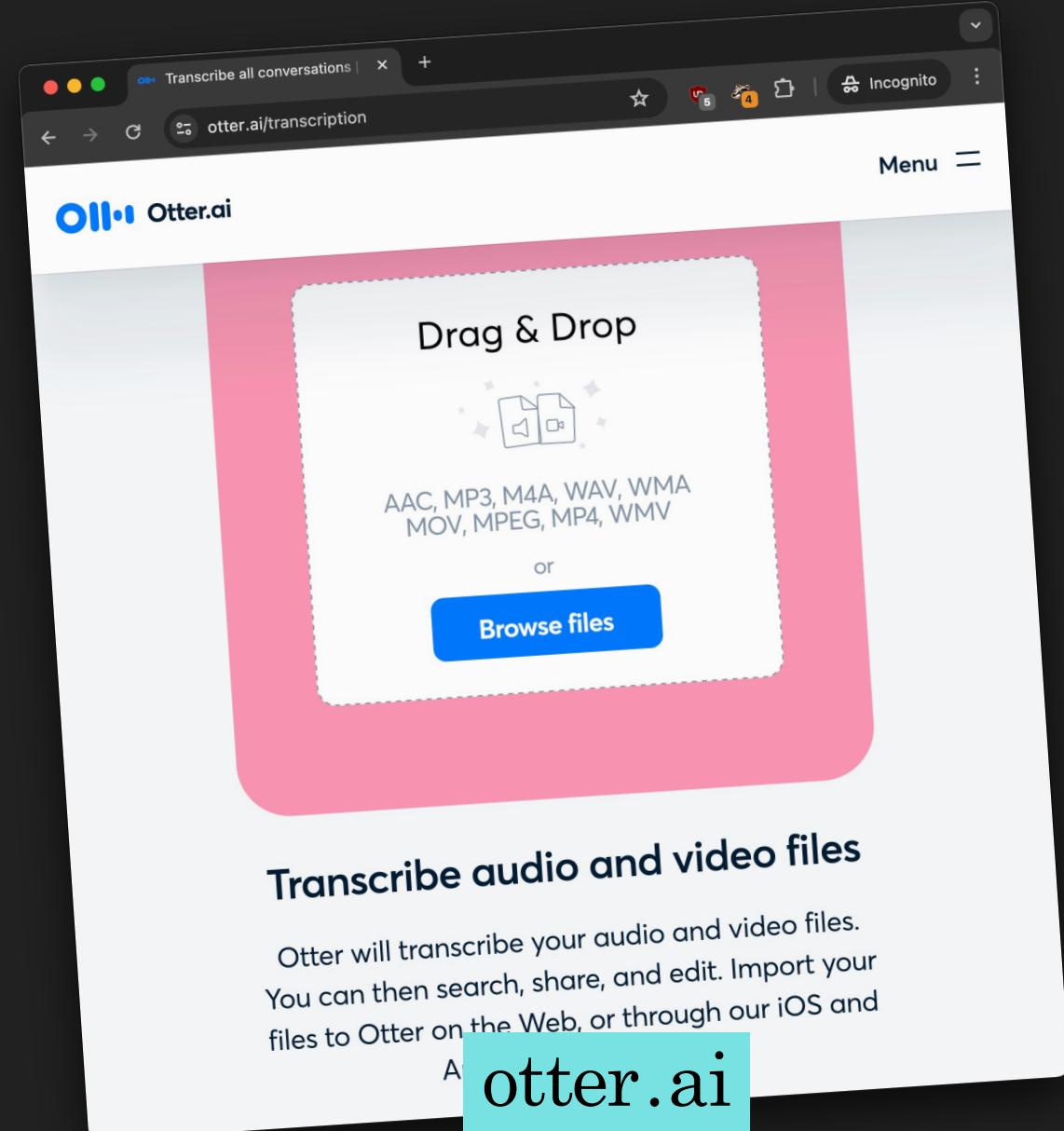
Free
Automatically
generated
Easy to
download

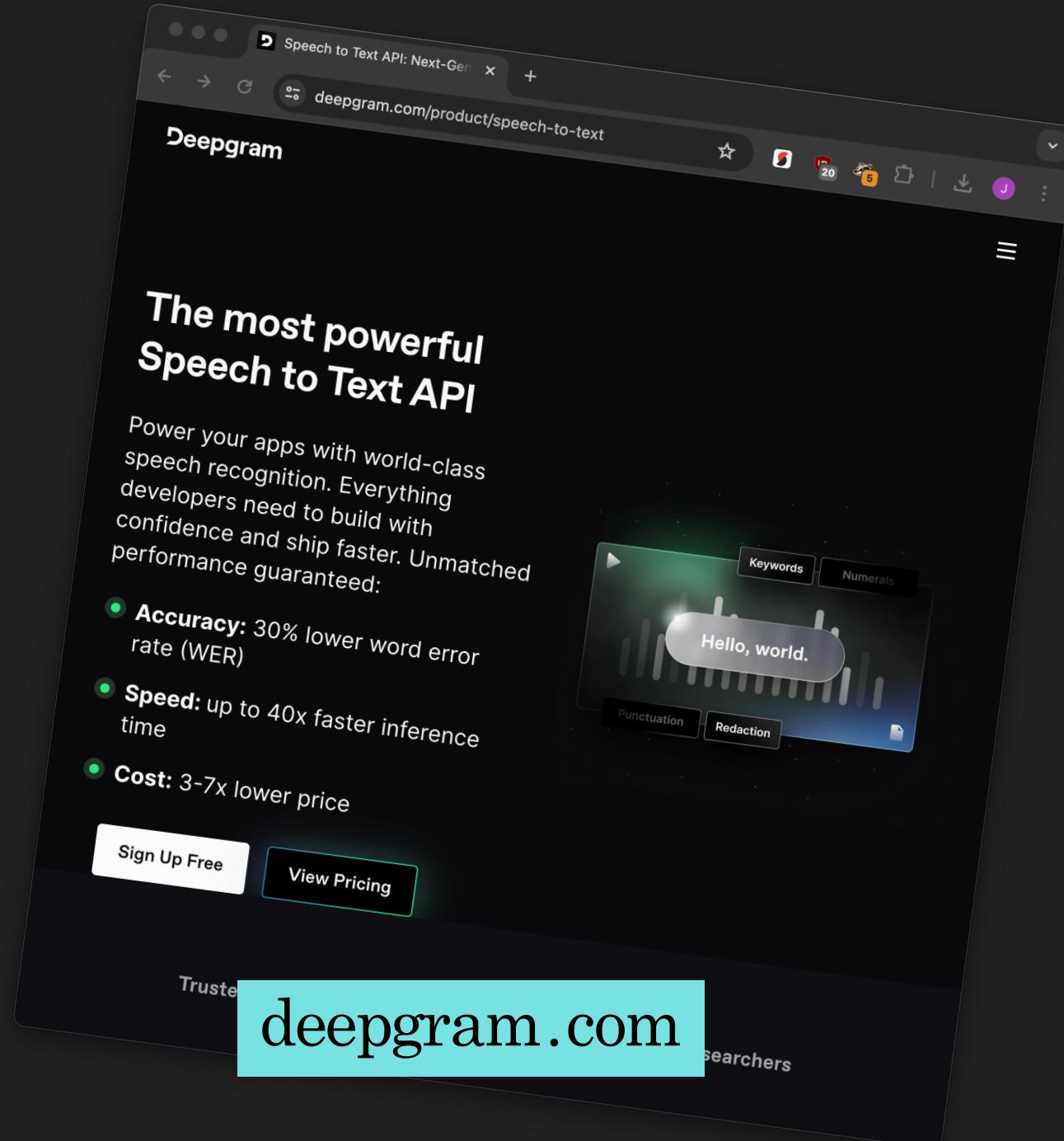
So easy to use

Already incorporated
in journalist's life in
ten thousand ways

Costs some money

Diarization





All sorts of extras

API available

Supposedly state of
the art???

Diarization

First, we'll use them…

Citizen DJ / American English x +

citizen-dj.labs.loc.gov/loc-american-english-dialect-recordings/use/

Citizen DJ

Collection: American English Dialect Recordings ▾

American English Dialect Recordings

The Center for Applied Linguistics Collection contains 118 hours of recordings documenting North American English dialects. The recordings include speech samples, linguistic interviews, oral histories, conversations, and excerpts from public speeches. They were drawn from various archives, and from the private collections of fifty collectors, including linguists, dialectologists, and folklorists.

[View this collection on the Library of Congress website](#)

Download files in bulk

A set of **4389 audio samples** have been automatically generated from this collection for use in your favorite music production software. It contains **293 audio segments** and **4096 one-shot audio clips**, ideal for use in samplers.

- [Download in 192 Kbps .mp3 format](#) (111.8mb) - Ideal if you would just like to preview and browse this collection
- [Download in 16-bit .wav format](#) (505.3mb) - Ideal for production

Rights & access

The sound recordings from this collection that are part of the Citizen DJ project are free to use and reuse. You can copy, modify, distribute and perform the works, even for commercial purposes, all without asking permission or paying fees.

About this Collection | Amer... x +

loc.gov/collections/american-english-dialect-recordings-from-the-center-for-applied-linguistics/about-this-colle... ☆

LIBRARY LIBRARY OF CONGRESS

This Collection | 🔍 Share

Library of Congress » Digital Collections » American English Dialect Recordings: The Center for Applied Linguistics Collection » About this Collection

COLLECTION

American English Dialect Recordings: The Center for Applied Linguistics Collection

[About this Collection](#) [Collection Items](#) [Articles and Essays](#)

Featured Content

Speech by Amelia Earhart

Speech by Fiorello H. La Guardia, New York

Speech by Alben Barkley, Kentucky

Speech by Isaac Asimov, New York, New York

Interview with H. L. Mencken

[Listen to this page](#)

About this Collection

North American English dialects. The recordings include speech samples, linguistic interviews, oral histories, conversations, and excerpts from public speeches. They were drawn from various archives, and from the private collections of fifty collectors, including linguists, dialectologists, and folklorists.

Citizen DJ / American English x LOC Conversation with 11 year old x +

loc.gov/item/afccal000390/

AUDIO RECORDING

Conversation with 11 year old black female, Washington, DC

Conversation with 11 year old black female, Washington, DC

00:00 09:02

Share

Download: MP3 Go

About this Item

Title
Conversation with 11 year old black female, Washington, DC

Names
Fasold, Ralph W. (Collector)
Scott, Phyllis Marie (Speaker)

Created / Published
October 30, 1968

Headings

- Games
- English language--Dialects--Washington (D.C.)
- Children, Black--Washington (D.C.)--Language
- African American girls--Washington (D.C.)--Language
- Jump rope rhymes
- Kickball
- Rope skipping
- Monopoly (Game)
- Football
- Interviews
- United States -- District of Columbia



The next excerpt is taken from Original Tape 296.
What kind of games do you play?
Sometimes I play kickball, jump rope. Sometimes I just be at home. That's all.
Well, how do you play kickball?
Team, two teams. You have to have two teams.
You have boys on one side, girls on the other.
Then they roll the ball and you kick it.
And if they get three outs, if you kick the ball and they catch it, that's one out.
There's things called fouls, and that's when they roll over first base.
You know, that don't mean nothing. You have to kick over.
If they get you out, if they tell you with the ball or touch the base, then you're out.
And you said jump rope. What kind of jump rope games do you play?
Double dutch.
Oh, what's that?
Two people be at each end, turn them up like this,
and a person or two, how many people want to be in, get in, and they jump rope.
Are there any things that you say while you're jumping rope?
One, two, three, four, like that.
Oh, you count.
Do you have any sort of rhymes or anything that you say while you're doing it? Nothing at all?
Okay. Can you think of any other games that you play now or used to play a long time ago?
You never play very much
Game that you buy. You know, like a game that you buy downtown like that? Some kind of game.
Like what?

Racial disparities in automated speech recognition

Allison Koenecke^a , Andrew Nam^b, Emily Lake^c, Joe Nudell^d, Minnie Quartey^e, Zion Mengesha^c, Connor Toups^c, John R. Rickford^c, Dan Jurafsky^{c,f}, and Sharad Goel^{d,1} 

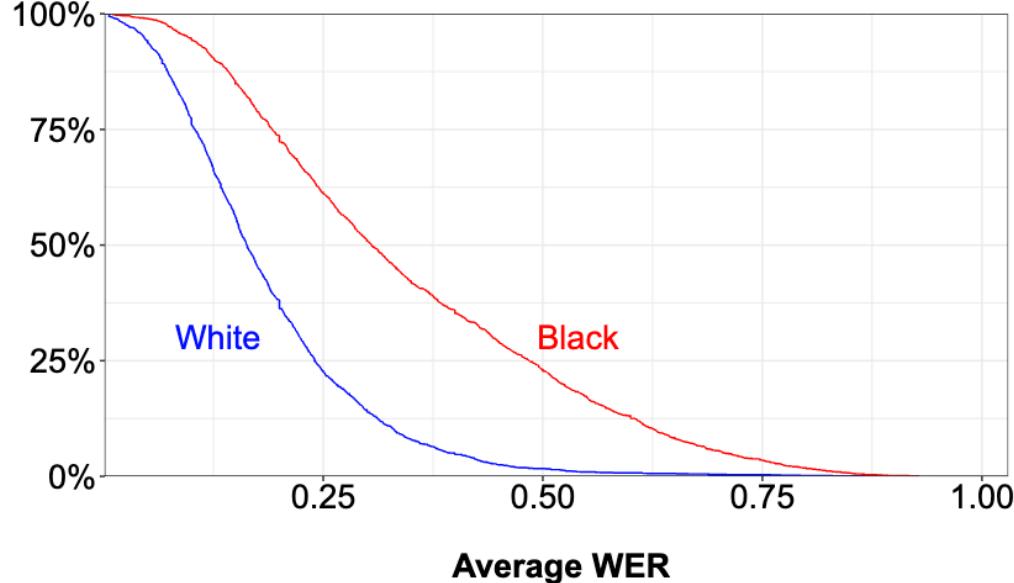


Fig. 2. The CCDF denotes the share of audio snippets having a WER greater than the value specified along the horizontal axis. The two CCDFs shown for audio snippets by white speakers (blue) versus those by black speakers (red) use the average WER across the five ASR services tested. If we assume that a $\text{WER} > 0.5$ implies a transcript is unusable, then 23% of audio snippets of black speakers result in unusable transcripts, whereas only 1.6% of audio snippets of white speakers result in unusable transcripts.

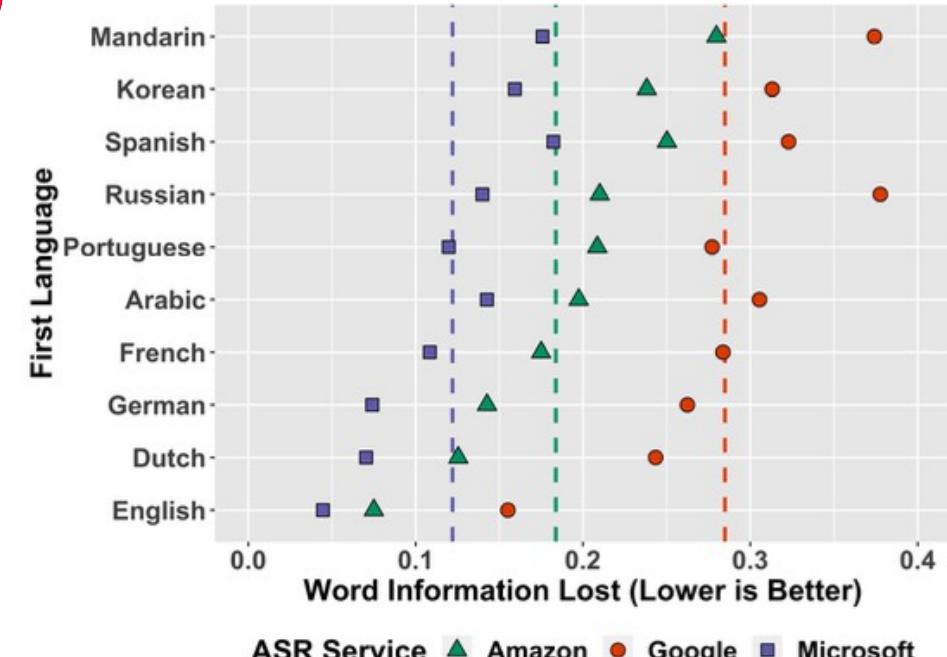
Performance Disparities Between Accents in Automatic Speech Recognition

Alex DiChristofano
a.dchristofano@wustl.edu
Division of Computational & Data Sciences
Washington University in St. Louis
St. Louis, Missouri, USA

Shefali Chandra
sc23@wustl.edu
Department of Women, Gender and Sexuality Studies and
Department of History
Washington University in St. Louis
St. Louis, Missouri, USA

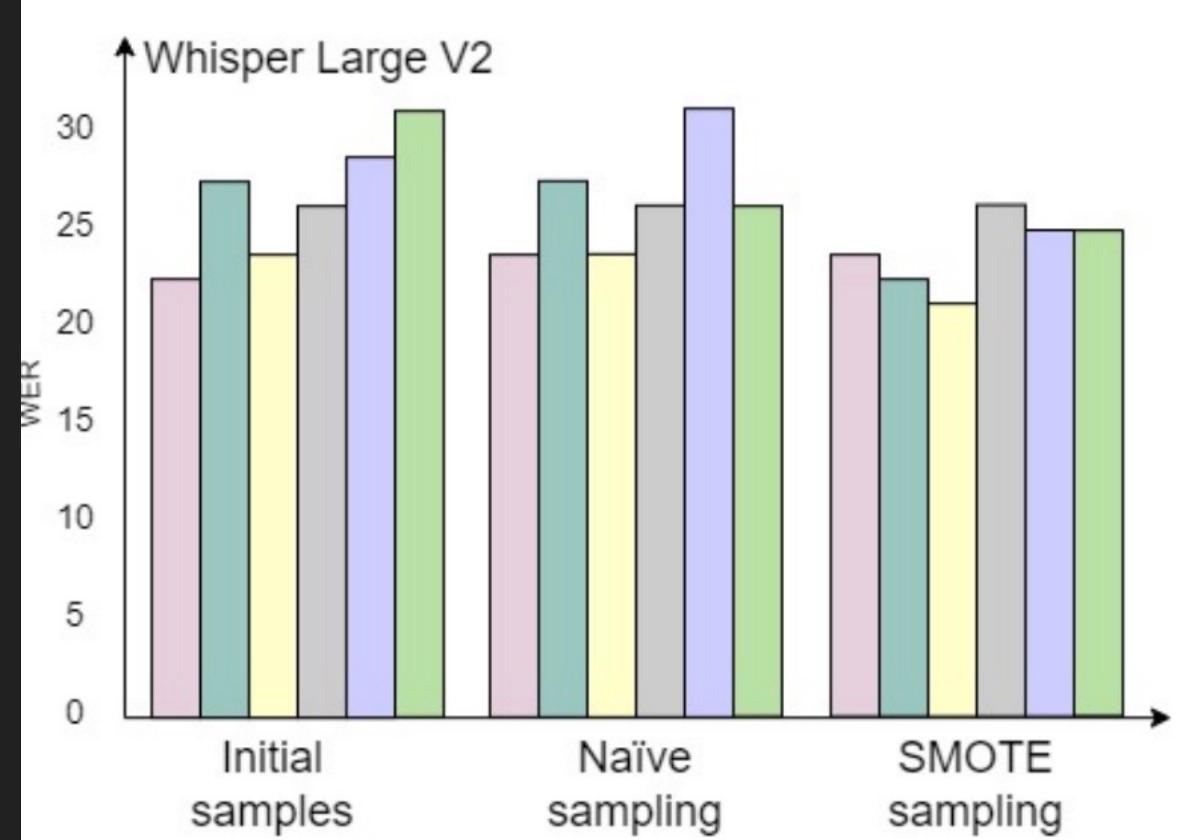
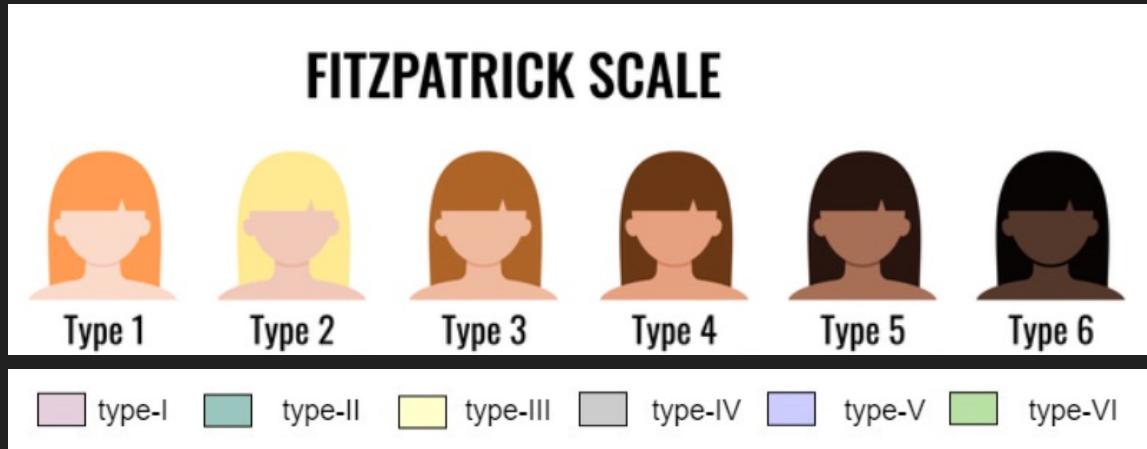
Henry Shuster
henryshuster@wustl.edu
Department of Computer Science & Engineering
Washington University in St. Louis
St. Louis, Missouri, USA

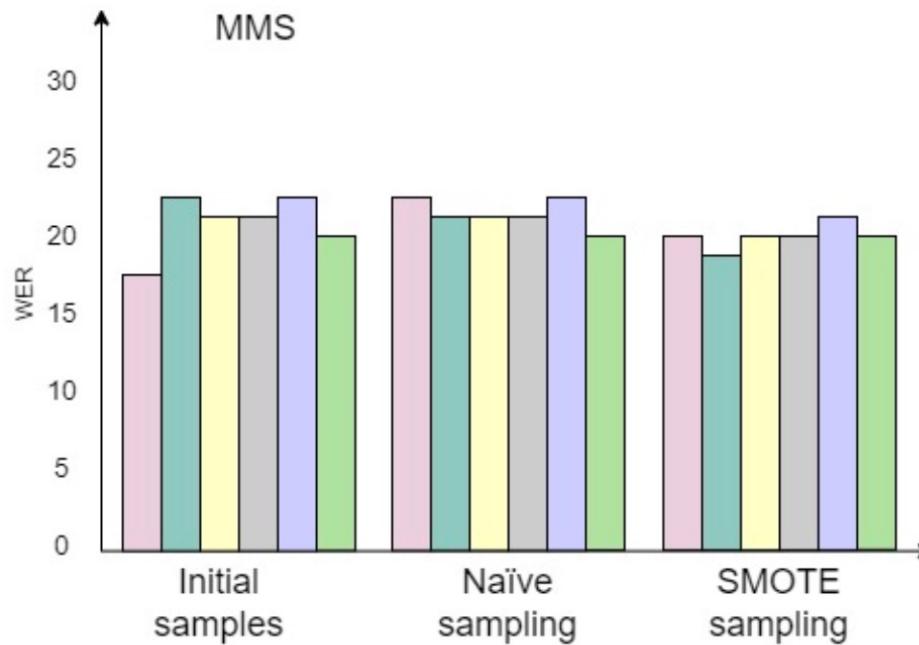
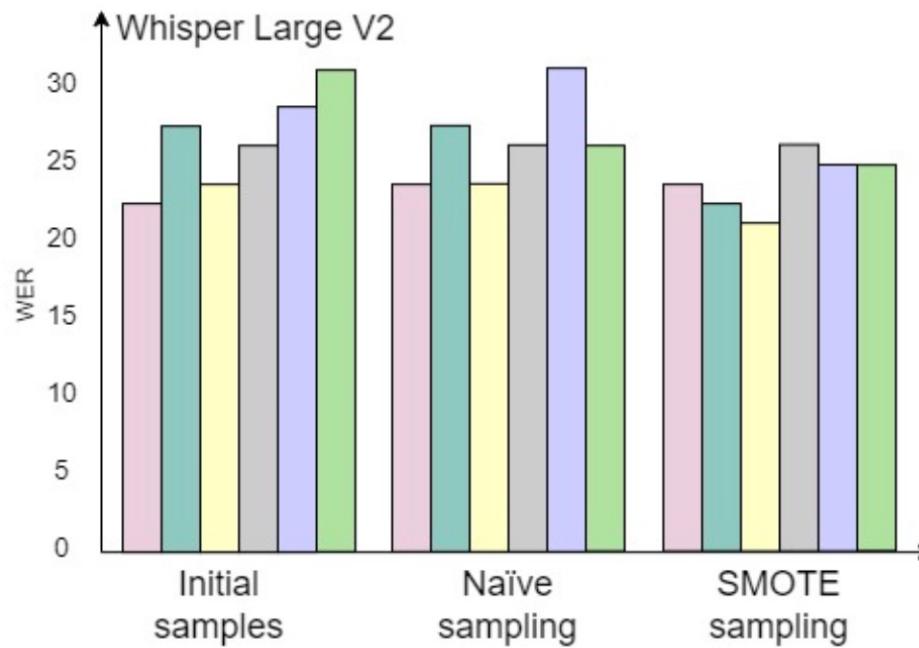
Neal Patwari
npatwari@wustl.edu
Department of Electrical & Systems Engineering and
Department of Computer Science & Engineering
Washington University in St. Louis
St. Louis, Missouri, USA



The Balancing Act: Unmasking and Alleviating ASR Biases in Portuguese

Ajinkya Kulkarni, Anna Tokareva, Mohammed Rameez Qureshi, Miguel Couceiro





facebook/mms-1b-all

Massively Multilingual Speech (MMS) - Finetuned ASR - ALL

This checkpoint is a model fine-tuned for multi-lingual ASR and part of Facebook's [Massive Multilingual Speech project](#). This checkpoint is based on the [Wav2Vec2](#) architecture and makes use of adapter models to transcribe 1000+ languages. The checkpoint consists of **1 billion parameters** and has been fine-tuned from [facebook/mms-1b](#) on 1162 languages.

Table Of Content

- [Example](#)
- [Supported Languages](#)
- [Model details](#)
- [Additional links](#)

Facebook's MMS model is better??



<https://bit.ly/abraj24-audio-video>

Activity: Let's try Facebook's model

<https://huggingface.co/spaces/mms-meta/MMS>



Simon Willison @simonw · Apr 21

It's quite fun leaving ChatGPT Voice running and later seeing how it hallucinated increasingly weird fragments of conversation based on snippets of background noise

Thanks for watching. Bye bye.



ChatGPT

You're welcome! Take care!



You

Thank you.



ChatGPT

You're welcome!



You

Thank you, and have a good day.



ChatGPT

Thanks, you too! Have a great day!



You

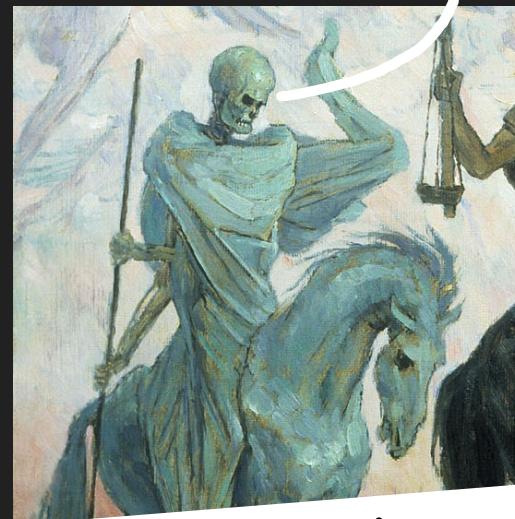
If you have any questions or comments,
please feel free to email me at chatgpt.com.

Thank you for watching.



ChatGPT

like and subscribe



Hallucinations

...then, evaluate!

CASE STUDY

Newsroom: KSAT-TV

About the newsroom

KSAT-TV is an ABC-affiliated local television station in San Antonio, Texas owned by Graham Media. The station also publishes on a variety of digital platforms. Among a multitude of duties, the station's digital producers repurpose existing content produced for television broadcasts. Separately, the station's photographers process much of their locally shot video through a commercial transcription solution.

Project goals

The goal of the project was to create a system that would transcribe, categorize, and summarize video input of live news events such as locally shot interviews or press conferences and deliver that summary as a prewrite of a story to the digital CMS.

Major challenges

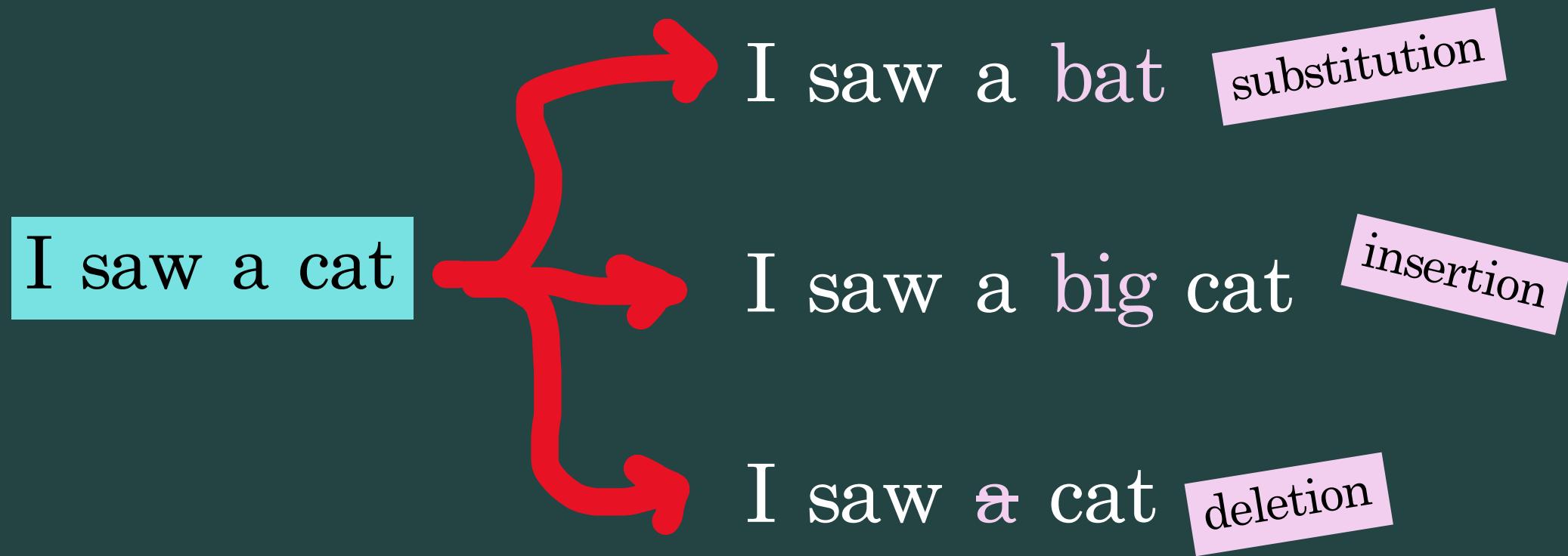
The accuracy of the transcripts was a major issue to be addressed as an inaccurate transcript leads to an inaccurate summary. As technology stands today, there is no such thing as a perfect AI-powered transcript. It was determined that a human would need to be in the loop in between the transcription and summarization processes. In the developed system, a journalist will review the accuracy of a transcript before allowing it to proceed to summarization.

Additionally, the accuracy of the summarizations provided by the generative AI system were questionable, and the system would sometimes make up quotes to use in the summary. However, the development team worked on this issue by testing out different prompts to see which would yield an accurate and comprehensive summary of the transcript, eventually landing on a few prompts that accomplished this.

The generative AI summarization system could also only intake a limited size clip at a time, so longer city council meetings could not all be summarized at once. But the team used a text chunker to split arbitrary length transcripts into chunks that the AI could process.

1. There were delays in getting documentation for the Arc

Word error rate



I saw a cat

substitution

Who saw a big bat?

substitution

insertion



2 substitutions + 1 insertion + 0 deletions

4 total words

$$\frac{3}{4} = 0.75 \text{ or } 75\% \text{ WER}$$



Reference: I'm a five-year-old kid
ASR: I am a 5 year old kid

S = 2, I = 3, D = 0
WER = 5/4 = 125%



Reference: I have sent a message
ASR: I haven't sent a message

S = 1, I = 0, D = 0
WER = 1/5 = 20%

Other measurements

- CER: character error rate (awful)
- WER: word error rate
- SER: sentences with at least one mistake
- MER: Match error rate, a more advanced version of WER
- WIL: Word information lost, fanciest approach

The screenshot shows a Mac OS X desktop environment with a web browser window open to the GitHub repository for JiWER. The title bar reads "jitsi/jiwer: Evaluate your speech recognition system". The browser interface includes standard controls like back, forward, and search, along with a tab bar showing multiple open tabs.

The main content area displays the repository's README file. It starts with a section titled "JiWER" and a brief description: "JiWER is a simple and fast python package to evaluate an automatic speech recognition system. It supports the following measures:"

1. word error rate (WER)
2. match error rate (MER)
3. word information lost (WIL)
4. word information preserved (WIP)
5. character error rate (CER)

Below this, there is a note about how the measures are computed using the minimum-edit distance between reference and hypothesis sentences, mentioning the use of [RapidFuzz](#) for C++ implementation.

Two sections are present at the bottom: "Documentation" and "Installation". A link to the documentation is provided: "For further info, see the documentation at jitsi.github.io/jiwer".

The URL "https://github.com/jitsi/jiwer" is also visible at the bottom of the browser window.

```
from transformers.models.whisper.english_normalizer import BasicTextNormalizer
import jiwer

normalize = BasicTextNormalizer()

truth = open("gotranscript.txt").read()
whisper = open("whisper.txt").read()

truth = normalize(truth)
whisper = normalize(whisper)

output = jiwer.process_words(truth, whisper)
print(output.wer)
```

	whisperx.large-v3	whispercpp.large-v3	insanely-fast-whisper.large-v3	macwhisper.large	youtube	deepgram	whisperx.base	insanely-fast-whisper.base	otter_ai	insanely-fast-whisper.tiny
whisperx.large-v3	0.00	0.14	0.15	0.17	0.26	0.28	0.30	0.37	0.47	0.51
whispercpp.large-v3	0.14	0.00	0.15	0.16	0.23	0.23	0.28	0.34	0.44	0.48
insanely-fast-whisper.large-v3	0.15	0.15	0.00	0.18	0.28	0.29	0.32	0.36	0.46	0.50
macwhisper.large	0.17	0.17	0.19	0.00	0.28	0.29	0.31	0.37	0.47	0.52
deepgram	0.26	0.23	0.29	0.27	0.28	0.00	0.32	0.37	0.44	0.48
youtube	0.26	0.24	0.29	0.28	0.00	0.29	0.31	0.36	0.42	0.50
whisperx.base	0.31	0.30	0.34	0.31	0.32	0.35	0.00	0.28	0.44	0.51
insanely-fast-whisper.base	0.36	0.35	0.37	0.36	0.36	0.39	0.27	0.00	0.47	0.48
insanely-fast-whisper.tiny	0.48	0.49	0.50	0.49	0.48	0.48	0.47	0.46	0.52	0.00
otter_ai	0.54	0.54	0.56	0.54	0.49	0.54	0.50	0.55	0.00	0.63

You can use the WER calculation from the machine recognition results to evaluate the quality of the model you're using with your app, tool, or product. A WER of 5-10% is considered to be good quality and is ready to use. A WER of 20% is acceptable, but you might want to consider more training. A WER of 30% or more signals poor quality and requires customization and training.

-Microsoft

normalized-whisperx.txt

- about there s another one that s almost
- like that called simple simon you know how
to play that one simple simon have you ever
played that one yeah i think we say simon
said do this oh yeah that s it that s it
that s right yeah well tell me how to play
that one yeah if you pick a person it can
be in a circle a line or like well you don
t have to be in no kind of form or nothing
like that and then she say| simon said do
this and you have to do this every time he
say| simon said if he say| i said do this and
you do it you| out what s the difference
between that one and mother may i she call|
you to do| i don t know almost like they|
almost alike if you ever been in a line of
some sort and she| call your name but simon
say you don t call your name he just say|
simon said do this like that and simon says
what do you have everybody does it all at
once and once mm hmm oh that yeah i thought
that was something like that did you ever
play when you were little did you ever play
hide and seek no you never played hide and

+ play that one yeah if you pick a person it can
+ be in a circle a line or like well you don t
have to be in no kind of form or nothing like
+ that and if he says simon said do this| you
+ have to do this every time he says| simon said
+ if he says| i said do this and you do it you| re
out what s the difference between that one and
+ mother may i she called| you to do| something i
+ don t know almost like they| re almost alike
+ they d be having a bit of| a line of some sort
+ and she| d call your name but simon| said| you
+ don t call your name he just| said| simon said
+ do this like that and simon says what do you
have everybody does it all at once or what mm
hmm oh that yeah i thought that was something
like that did you ever play when you were
little did you ever play hide and seek no you
never played high school no uh huh um you ever
play in the boys game you know| football yeah
okay tell me a girl s version of how football
is played well i can catch some boys some
balls we play uh five steps like i have a few
boys on my side and a few other girls on the
other side i have a few boys and a few girls

How can we determine the
“correct” transcription?

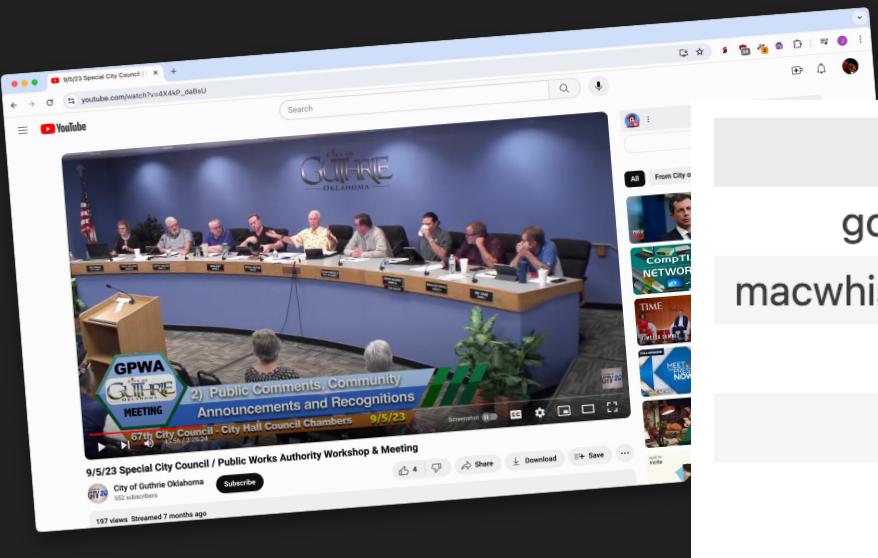
The screenshot shows the homepage of GoTranscript. At the top, there's a navigation bar with links for Services, Pricing, Enterprise, Company, and Contact. Below the navigation, a large banner features the text "100% Human-Made Transcription Services" and a 4.9/5 rating with 3547 reviews. To the right of the banner, it says "99%+ accuracy guaranteed" and "From \$0.84 /min Best industry prices". Below the banner, there are four service cards: "Transcription" (From \$0.84 per minute), "Automated Transcripts" (From \$0.20 per minute), "Captions" (From \$1.22 per minute), and "Audio Trans" (From \$8.80 per minute). Each card has a yellow "Order Transcription" button and a "Learn more" link.

The screenshot shows a web browser window displaying an article from nytimes.com/wirecutter/reviews/best-transcri... The title of the article is "The Best Transcription Services". The page includes a header with "How can we help?", the Wirecutter logo, an "Account" dropdown, and a "Subscribe" button. Below the header, there's a navigation bar with categories: Home & Garden, Kitchen, Health & Lifestyle, Tech, Baby & Kid, Style, Gifts, Shopping, and Deals. A note at the top states: "We independently review everything we recommend. When you buy through our links, we may earn a commission." with a "Learn more" link. The main content area features a large image of a person wearing a headset, likely using speech recognition software. The text on the screen in the image reads: "Speech to text transcription in 5 minutes Advanced speech recognition software Select audio/video file". The article was last updated on June 30, 2020, and includes social sharing icons for LinkedIn, Twitter, Facebook, and Email, along with a "Save" button.

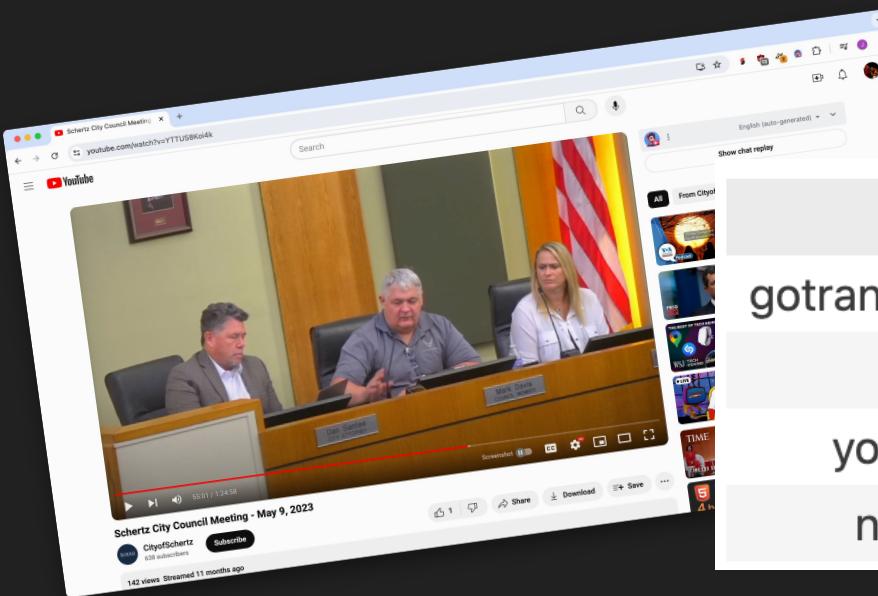
https://www.youtube.com/watch?v=4X4kP_daBsU

A screenshot of a YouTube video player. The video shows a group of people seated around a long conference table in a council chamber. The background features a blue wall with the "City of GUTHRIE OKLAHOMA" logo. A podium on the left has a microphone and a small American flag. In the foreground, a woman is seated at a desk, facing the camera. A green banner on the screen reads "GPWA CITY OF GUTHRIE MEETING" and "2) Public Comments, Community Announcements and Recognitions". Below the banner, text indicates "67th City Council - City Hall Council Chambers" and the date "9/5/23". The video player interface includes a play button, a progress bar at 43:56 / 3:26:24, and a timestamp "43:56 / 3:26:24". Below the video, the title "67th City Council / Public Works Authority Workshop & Meeting" is displayed, along with the date "9/5/23 Special City Council / Public Works Authority Workshop & Meeting". The channel information shows "City of Guthrie Oklahoma" with 552 subscribers and a "Subscribe" button. The video has 197 views and was streamed 7 months ago.

A screenshot of a YouTube video player. The video shows three people seated at a conference table in a council chamber. From left to right, they are identified by nameplates as "Dan Santee CITY ATTORNEY", "Mark Davis COUNCIL MEMBER", and "Kathy...". An American flag is visible in the background. The video player interface includes a play button, a progress bar at 55:01 / 1:24:58, and a timestamp "55:01 / 1:24:58". Below the video, the title "Schertz City Council Meeting - May 9, 2023" is displayed. The channel information shows "CityofSchertz" with 638 subscribers and a "Subscribe" button. The video has 1 view and was streamed 7 days ago. The top navigation bar of the browser shows "All", "From City of Guthrie Oklahoma", and "Meetings". The right sidebar displays a list of recommended videos, including titles like "WATCH LIVE: Buttigieg announces new airline refund...", "PBS NewsHour", "Scherz City Council Meeting", and "youtubecom/watch?v=YTTUS8Koi4k". The URL "https://www.youtube.com/watch?v=YTTUS8Koi4k" is also visible in the bottom right corner of the video frame.



	gotranscript	macwhisper-large	large	nova-2	youtube
gotranscript	0.00	0.15	0.16	0.22	0.23
macwhisper-large	0.16	0.00	0.14	0.22	0.21
large	0.16	0.14	0.00	0.17	0.16
nova-2	0.22	0.21	0.17	0.00	0.18
youtube	0.23	0.21	0.16	0.19	0.00



	gotranscript	large	youtube	nova-2
gotranscript	0.00	0.09	0.14	0.14
large	0.09	0.00	0.09	0.10
youtube	0.13	0.09	0.00	0.11
nova-2	0.14	0.10	0.11	0.00

from various city-level meetings on YouTube

WER		WER		WER		WER		WER
Model		Model		Model		Model		Model
gotranscript	0.000	gotranscript	0.000	gotranscript	0.000	gotranscript	0.000	gotranscript
large	0.089	large	0.111	large	0.130	large	0.208	macwhisper-large
youtube	0.138	nova-2	0.166	nova-2	0.191	nova-2	0.290	large
nova-2	0.145	youtube	0.195	youtube	0.201	youtube	0.317	nova-2
								youtube

You can use the WER calculation from the machine recognition results to evaluate the quality of the model you're using with your app, tool, or product. A WER of 5-10% is considered to be good quality and is ready to use. A WER of 20% is acceptable, but you might want to consider more training. A WER of 30% or more signals poor quality and requires customization and training.

-Microsoft

How do we analyze the text?

ChatGPT

chatgpt.com

ChatGPT 4o



 Python script for daily email reports

 Overcome procrastination

 Email for plumber quote

 Experience Seoul like a local

 nKksNeSBNP0.mp3

File

what is being discussed in this audio recording?

ChatGPT can make mistakes. Check important info.

nKksNeSBNP0

Macintosh HD > nKksNeSBNP0 > nKksNeSBNP0.mp3

1 of 8 selected, 1.28 TB available



ChatGPT 4o ▾



 Make me a personal webpage

 Overcome procrastination

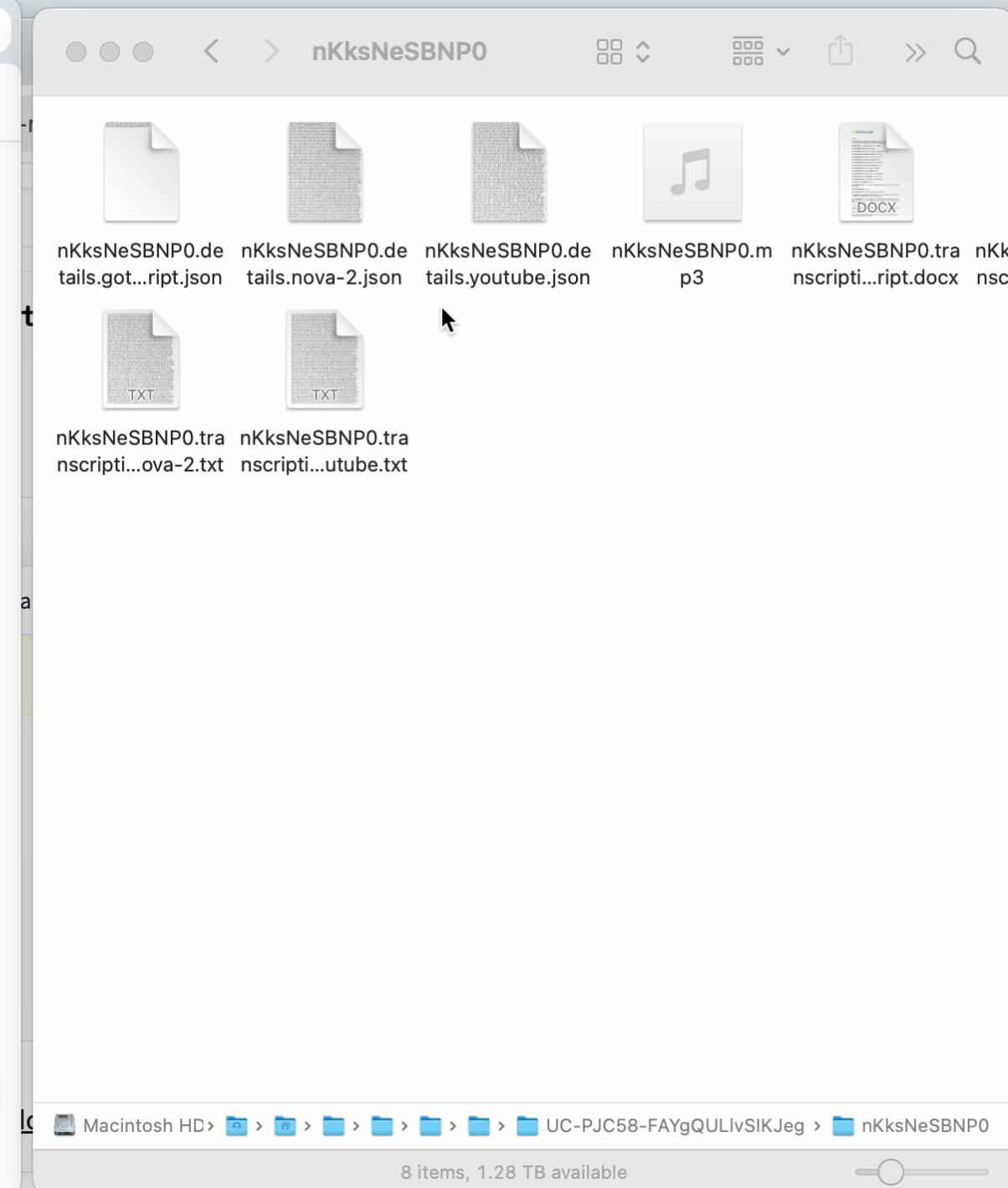
 Design a fun coding game

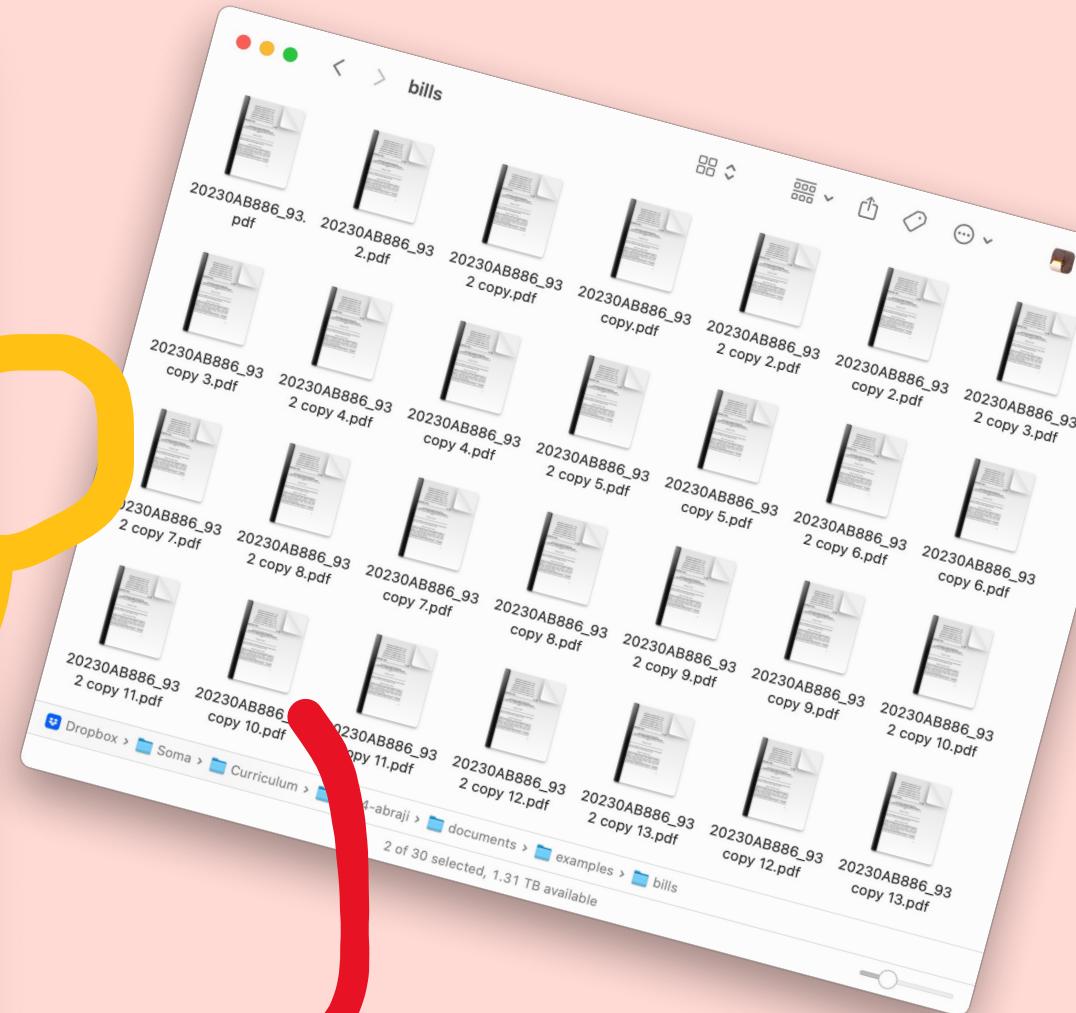
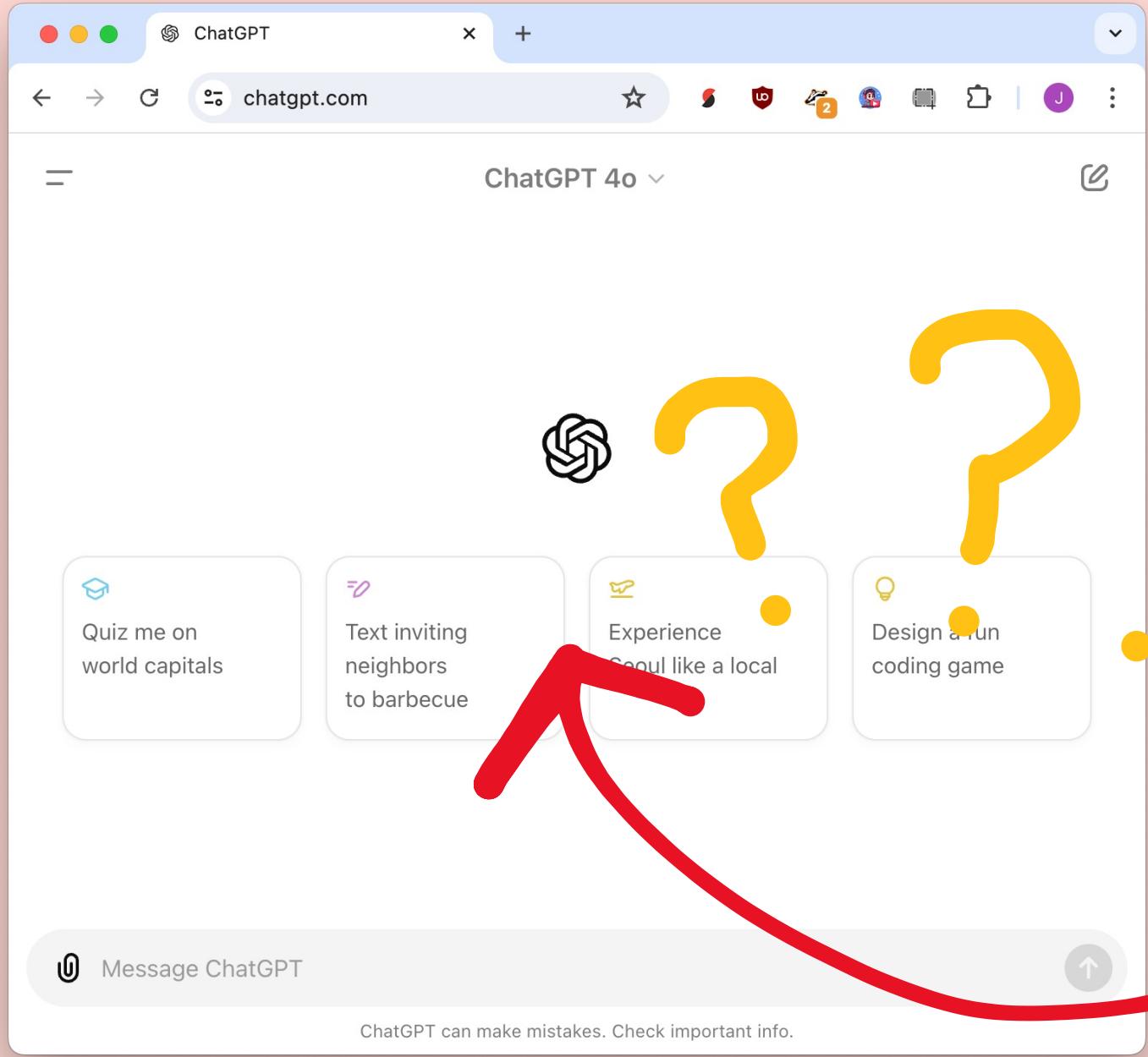
 Fun fact about the Roman Empire

Message ChatGPT



ChatGPT can make mistakes. Check important info.





Claude for Sheets™

Claude for Sheets™ brings your friendly AI assistant Claude from Anthropic right there in your spreadsheets.

By: [Anthropic](#)

Listing updated: March 11, 2024

Install

★★★★☆ 14 ⓘ ↓ 30K+

Works with:



AI in Sheets™ .Spread...

The most powerful AI tool for Google Sheets™ on the planet. Built for true scale. Safe for the enterprise. Unbeatable pricing model.

By: [Spreadsheet Daddy](#)

Listing updated: March 18, 2024

Install



GPT for Sheets™ and ...

ChatGPT in Google Sheets™ and Docs™. Get the full power of AI for inspiration, data cleaning, classification, extraction, translation, edition, summarization, writing directly in you...

By: [Talarian](#)

Listing updated: March 6, 2024

Independent security verification

Uninstall

★★★★☆ 318 ⓘ ↓ 6M+

Claude for Sheets™ - Google workspace.google.com/marketplace/app/clause_for_sheets/909417792257

Google Workspace Marketplace Search apps

Home Editor's choice Intelligent apps Work from everywhere Business essentials Apps to discover Built by Google Top charts Most popular Top rated Categories Admin managed

Uninstall

Claude for Sheets™

Claude for Sheets™ brings your friendly AI assistant Claude from Anthropic right there in your spreadsheets.

By: [Anthropic](#) Listing updated: March 11, 2024

Works with:

★ ★ ★ ★ ★ 17 i 34K+

Overview Permissions Reviews

This sheet lets you compare a set of related prompts to Claude

Step 1: type your prompts in this column

In one sentence, what is good about the color blue?
Suppose you have a table called "Customers" with the following column names:
CustomerID, CustomerName, ContactName, Address, City, PostalCode, Country
Can you write a SQL query that finds all customers in San Francisco with a customer ID greater than 1000?

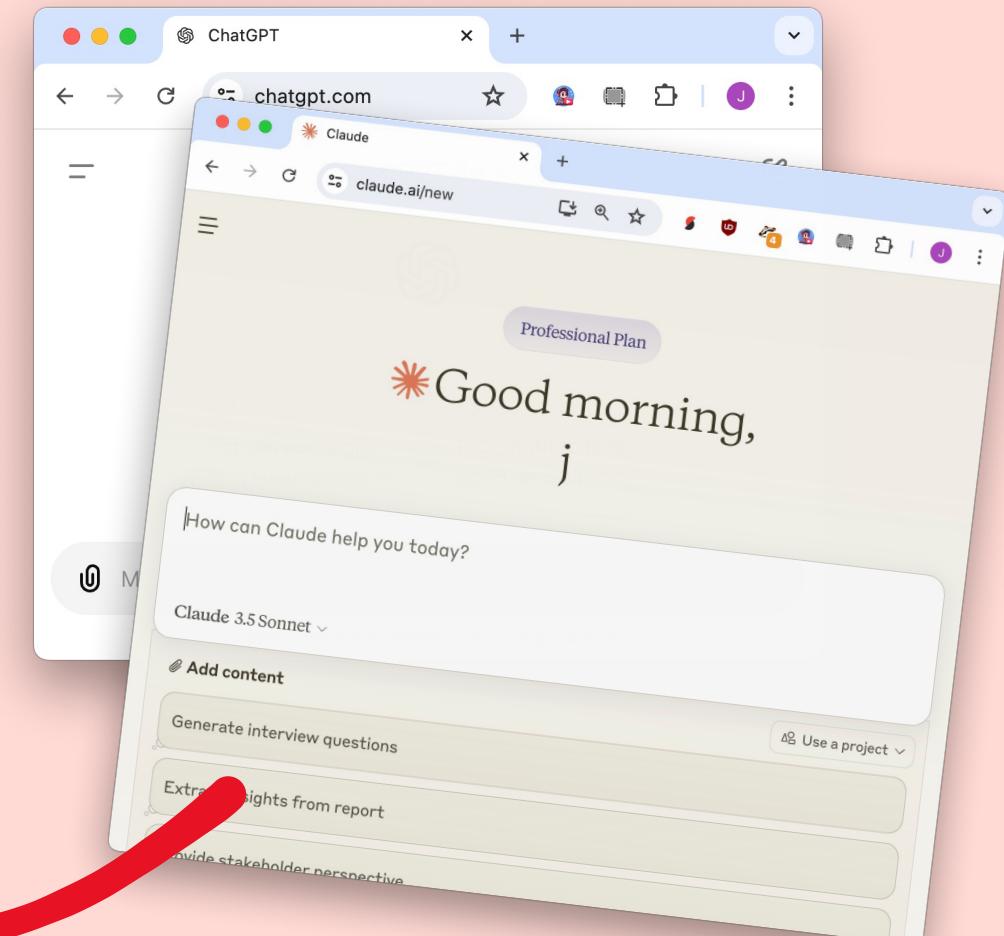
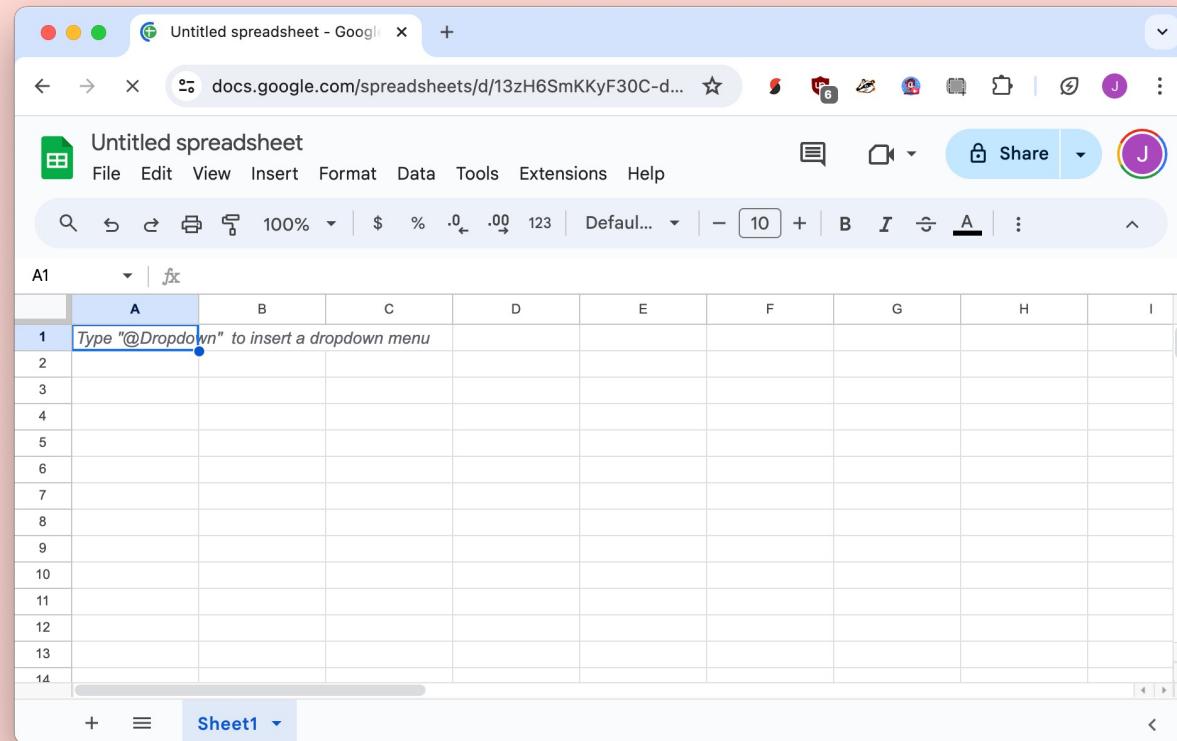
Step 2: Select your model:

claude-v1.3
claude-2

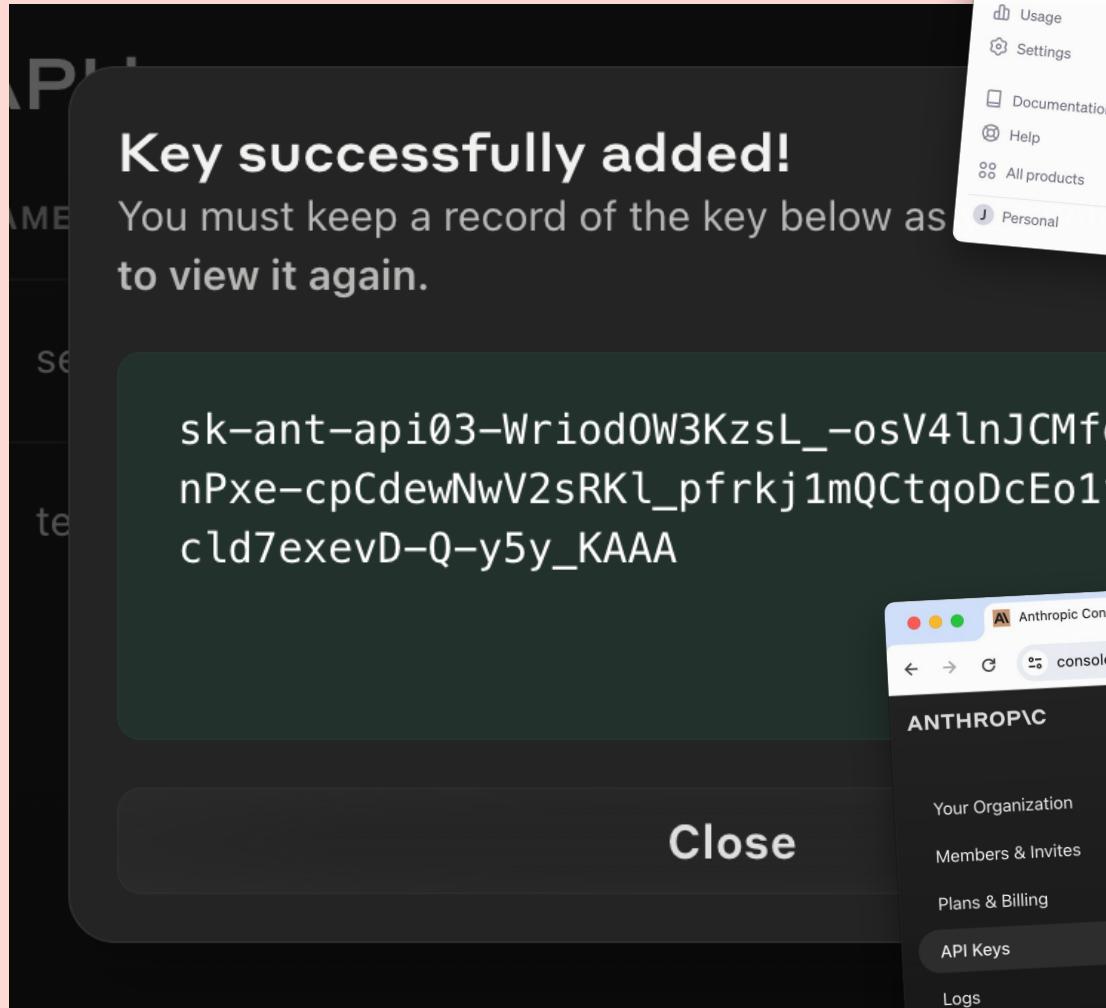
Step 3: review Claude responses in this column

Blue is a calming and soothing color.
Here is a SQL query to find all customers in San Francisco with a customer ID 1000 from the Customers table:
...sql
SELECT *
FROM Customers
WHERE City = 'San Francisco' AND CustomerID > 1000
...
This query selects all columns (*) from the Customers table where the City co

Extensions Help Add-ons Macros Apps Script AppSheet Claude in Sheets



API keys



The screenshot shows the "API keys" section of the OpenAI API platform. The sidebar on the left includes links for Playground, Assistants, Fine-tuning, API keys (which is selected and highlighted in green), Storage, Usage, Settings, Documentation, Help, All products, and Personal.

The main content area displays a table of API keys:

NAME	SECRET KEY	TRACKING	CREATED	LAST USED	PERMISSIONS
Secret key	sk-...MJ0b	+ Enable	Mar 6, 2023	Apr 13, 2023	All
Secret key	sk-...e67A	+ Enable	Mar 18, 2023	Apr 19, 2023	All
Secret key	sk-...1PWe	+ Enable	Mar 18, 2023	Mar 18, 2023	All
Secret key	sk-...M1LX	+ Enable	Mar 19, 2023	Mar 19, 2023	All
Secret key	sk-...2ABE	+ Enable	Mar 23, 2023	Never	All

The screenshot shows the "API Keys" section of the Anthropic Console. The sidebar on the left includes links for Your Organization, Members & Invites, Plans & Billing, API Keys (selected and highlighted in orange), Logs, and Usage.

The main content area displays a table of API keys:

NAME	KEY	CREATED AT	UPDATED AT
claude for sheets	sk-ant-api03-1Sk...iAAA	Apr 9, 2024	
sample class	sk-ant-api03-g0n...rgAA	Apr 10, 2024	

A "Create Key" button is located in the top right corner of the interface.

Search ↶ ↷ 🖨️ 📂 200% \$ % .0 .00 123 Default... - 10 + B I ÷ A ↳

A1

| fx

A**B**

1

2

3

4

5

6

7

8

E8

fx

A

B

C

1 object

color

2 sky

3 strawberry

4 basketball

5

6

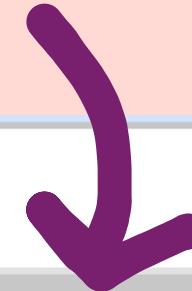
7

8

=CLAUDEEXTRACT("cor (em português)", A2)



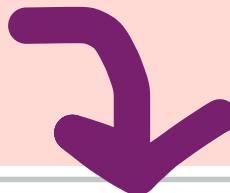
object	color
céu	céu
morango	Não há "cor (em português)" nessa string.
basquetebol	None



=CONCATENATE("Qual é a cor de '", A2, "'?
Responda em uma ou duas palavras.")



=CLAUDE(B2)



object	full prompt	response
céu	Qual é a cor de 'céu'? Responda em uma ou duas palavras.	Azul claro.
morango	Qual é a cor de 'morango'? Responda em uma ou duas palavras.	Vermelho.
basquetebol	Qual é a cor de 'basquetebol'? Responda em uma ou duas palavras.	Laranja.

Louisiana State Legislation (Completed in Class) ★ Saved to Drive

File Edit View Insert Format Data Tools Extensions Help

150% \$.00 123 Default... 10 B I A

G3 fx

	A	E	F	G
1	bill text	about retirement?	about employment?	
2	> HLS 24RS-94 **[ENGROSSED]{.underline}** > > 2024 Regular Session > > HOUSE BILL NO. 8 > > BY REPRESENTATIVES MELERINE AND HUGHES > > SCHOOLS/HIGH SCHOOL: Prohibits the use of an appeals pro > certain students > > HLS 24RS-88 **[REENGROSSED]{.underline}** > > 2024 Regular Session > > HOUSE BILL NO. 9 > > BY REPRESENTATIVE ILLG > > RETIREMENT/DISTRICT ATTY: Provides for survivor benefits c > the >			
3				



	A	E	F	G
1	bill text	about retirement?	about employment?	
	> HLS 24RS-94 **[ENGROSSED]{.underline}** > > 2024 Regular Session			
3	> certain students > > HLS 24RS-88 **[REENGROSSED]{.underline}** > > 2024 Regular Session > > HOUSE BILL NO. 9 > > BY REPRESENTATIVE ILLG > > RETIREMENT/DISTRICT ATTY: Provides for survivor benefit > the >	<input type="radio"/> no	<input type="radio"/> no	<input type="radio"/> yes

fx

=CLAUDEEXTRACT("Is this piece of legislation about retirement? Answer YES or NO", A1)

Elections européennes : Emm... +

lemonde.fr/elections-europeennes/article/2024/04/17/elections-europeenne...

Le Monde

Élections européennes 2024 | Résultats des européennes | Carte des résultats | Les 81 députés français | S'abonner

ÉLECTIONS EUROPÉENNES • EMMANUEL MACRON

Elections européennes : Emmanuel Macron dénonce « l'hypocrisie » du Rassemblement national

Le chef de l'Etat a déploré, mercredi depuis Bruxelles, que les « peurs profitent aux réponses les plus simplistes » dans la campagne pour le scrutin du 9 juin, fustigeant ceux qui dénoncent l'Union européenne tout en engrangeant les « dividendes silencieux ».

Le Monde avec AFP

Publié le 17 avril 2024 à 20h55, modifié le 18 avril 2024 à 03h01 • Lecture 2 min.

Ajouter à vos sélections



Les plus lus

- En direct, la situation politique en France : Elisabeth Borne siégera au sein du groupe Renaissance, son président n'est toujours pas élu
- L'avertissement de Gérard Larcher à Emmanuel Macron : « On ne joue pas avec la France »

Interrogé en marge de son déplacement à Bruxelles sur la campagne des élections européennes, Emmanuel Macron a ciblé, sans le nommer, le Rassemblement national (RN) dont la liste conduite par Jordan Bardella caracole en tête des sondages.

« Je pense qu'il y a beaucoup de peurs, d'inquiétudes dans le moment que nous vivons et que ces colères profitent toujours aux réponses les plus simplistes », a déclaré le chef de l'Etat français lors de son arrivée, mercredi 17 avril, dans la capitale belge avant la tenue d'un sommet européen, au cours d'un bref échange avec la presse aux côtés de la tête de liste de son camp aux européennes, Valérie Heyer. « Il y a aujourd'hui une espèce d'hypocrisie du débat et j'espère qu'en rentrant dans les prochaines semaines dans celui-ci, cette hypocrisie sera levée », a ajouté M. Macron, visant en creux le RN.

Il s'agissait de la première apparition publique du chef de l'Etat et de M^{me} Hayer depuis le début de la campagne, avant qu'ils n'assistent ensemble à une réunion du groupe Renew Europe.

Lire aussi | [Elections européennes 2024 : le calendrier du scrutin](#)

La candidate du camp présidentiel peine pour l'instant à imprimer sa marque dans la campagne. « On s'organise, on a la bonne candidate (...) On a surtout les bonnes idées, celles qui se sont mises en place en Europe ces dernières années », a lancé M. Macron, interpellé par un journaliste.

- =CLAUDEEXTRACT("people mentioned in the article, with titles if available. One per line.", B5)
- =CLAUDEEXTRACT("brief one-sentence summary", B5)
- =CLAUDEEXTRACT("English translation", E5)
- =CLAUDEEXTRACT("article category, one of politics, sports, international, other", B5)

B	C	E	F	G	H	I
content	sources	summary	language code	is it English?	English translation	article category
Ante inversores en Washington, Caputo defendió la estrategia fiscal y reafirmó el compromiso para levantar el cepo El jefe del Palacio de Hacienda participó de un seminario organizado por el banco J.P. Morgan WASHINGTON.- El ministro de Economía, Luis Caputo, llegó al seminario del banco de inversión J.P. Morgan en el Park Hyatt hotel poco después de las tres de la tarde, <small>Luego llegó un nota desmentida de Bloomberg</small>	Luis Caputo, Ministro de Economía Pablo Quirno, Secretario de Finanzas Javier Milei, gobierno Mauricio Macri, gobierno Nicolás Posse, Jefe de Gabinete Alberto Fernández, gobierno Pierre-Olivier Gourinchas, Economista Jefe del FMI	The economy minister defended the government's fiscal adjustment strategy and reiterated the intention to lift currency controls, but said they cannot do so with the current level of reserves <small>unless they receive a fresh injection of money.</small>	en	yes	The economy minister defended the government's fiscal adjustment strategy and reiterated the intention to lift currency controls, but said they cannot do so with the current level of reserves unless they receive a fresh injection of money.	international
Elections européennes : Emmanuel Macron dénonce « l'hypocrisie » du RN Le chef de l'Etat a déploré mercredi depuis Bruxelles que les « peurs profitent aux réponses les plus simplistes » dans la campagne pour le scrutin du 9 juin, fustigeant ceux qui dénoncent l'Union européenne tout en engrangeant les « dividendes silencieux ». Vous pouvez partager un article en cliquant sur les icônes <small>de partage en haut à droite de cette page</small>	Emmanuel Macron, President of France Valérie Heyer, Head of list for Macron's camp in European elections Jordan Bardella, Head of list for National Rally (Rassemblement National, RN) Raphaël Glucksmann, Head of list for Socialist Party (Parti Socialiste) and Place Publique Gabriel Attal, Prime Minister of France	Emmanuel Macron a déploré que les "peurs profitent aux réponses les plus simplistes", mercredi 17 avril.	fr	no	Emmanuel Macron lamented that "fears profit from the simplest answers", Wednesday April 17.	politics
岸田首相、韓国・尹大統領に日米首脳会談の成果説明　日韓連携も確認 岸田文雄首相は17日、韓国の尹錫悦（ウンソンニヨル）大統領と電話で協議した。首相によると、10日にワシントンで行われた日米首脳会談での合意事項などについて説明したという。 協議後、首相は首相官邸で記者団に「日韓、日米韓の連携をより深化させ、日韓首脳間を含め緊密に意思疎通を図つ	Fumio Kishida, Prime Minister; Yoon Suk-yeol, President of South Korea	岸田首相は17日、韓国の尹大統領と電話で協議し、日米首脳会談の成果を説明、日韓連携も確認した。	ja	no	Prime Minister Kishida held talks by phone with South Korean President Yoon on the 17th, explaining the results of the Japan-U.S. summit and confirming cooperation between Japan and South Korea.	other

Welcome To Instructor - Instr +

python.useinstructor.com

For updates follow @jxnlico on Twitter and star us on GitHub. If you don't like python, check out the TS/JS and Elixir ports. X

Instructor

Introduction Cookbook Blog Concepts Hub CLI Reference Tutorials Jobs Board (External) Prompting Tips

Search

instructor 1.3.4 6.6k 524

Introduction

Why use Instructor?
Help with Instructor
Installation
Contributing
Philosophy

Instructor, Generating Structure from LLMs 

Structured outputs powered by llms. Designed for simplicity, transparency, and control.

 Follow @jxnlico discord 97 online downloads 296k/month docs InstructorGPT

Instructor makes it easy to get structured data like JSON from LLMs like GPT-3.5, GPT-4, GPT-4-Vision, and open-source models including [Mistral/Mixtral](#), [Anyscale](#), [Ollama](#), and [llama-cpp-python](#).

It stands out for its simplicity, transparency, and user-centric design, built on top of Pydantic. Instructor helps you manage [validation context](#), retries with [Tenacity](#), and streaming [Lists](#) and [Partial](#) responses.

- Instructor provides support for a wide range of programming languages, including:
- Python
- TypeScript
- Ruby
- Go
- Elixir

Table of contents

Want your logo on our website?
Why use Instructor?
Getting Started
Using OpenAI
Using Anthropic
Using Gemini
Google AI
Vertex AI
Using Litellm
Correct Typing
Calling create
Handling async: await create
Returning the original completion:
create_with_completion
Streaming Partial Objects:
create_partial
Streaming Iterables:
create_iterable
Validation
More Examples
Contributing



```
class Person(BaseModel):
    name: str = Field(description="Person's name")
    title: Optional[str] = Field(description="Person's title, if available")

class Article(BaseModel):
    summary: str = Field(description="Brief, one-sentence summary (in English)")
    sources: List[Person] = Field(description="People mentioned in the article")
    language_code: str = Field(description="Two-letter language code of original text")
    category: Literal["politics", "sports", "international", "other"]
```



```
# Extract structured data from natural language
result = client.chat.completions.create(
    model="gpt-4",
    response_model=Article,
    messages=[{"role": "user", "content": article},
)
```

```
● ○ ●

{

  "category": "politics",
  "language_code": "fr",
  "sources": [
    { "name": "Emmanuel Macron", "title": "President of France" },
    { "name": "Jordan Bardella", "title": "Head of the National Rally" },
    { "name": "Valérie Heyer", "title": "Candidate for the European Elections" },
    { "name": "Gabriel Attal", "title": "Prime Minister of France" },
    { "name": "Raphaël Glucksmann", "title": "Candidate for the European Elections" }
  ],
  "summary": "Emmanuel Macron criticizes the National Rally (RN) for hypocrisy in the run-up to the European elections, arguing that they profit from EU membership while denouncing it."
}
```

FWS-R5-ES-2014-0050-0003_attachment_1....
1 page

Attn: FWS-R5-ES-2014-0050
Public Comments Processing
Division of Policy, Performance, and Management Programs
US Fish and Wildlife Service
5275 Leesburg Pike
Falls Church VA 22041

Re: Copenhagen Wind Farm

There are a couple of things that I would like to see addressed in the application for ITP and HCP.

First is my interest in the detection of morbidity and mortality at the wind turbines. Daily surveillance, preferably at sunrise, is, in my opinion, the only way to get reliable numbers. What has been proposed by the applicant and FWS?

Second is my interest in avoidance of bird electrocutions and collisions at collector lines and transmission lines. What is the proposal for avoidance of electrocution along these lines? What is the proposal for avoidance of collisions where these lines cross streams and wetlands?

Please keep me abreast of this application progress, either by sending me copies of pertinent documents, or by informing me when they are available on regulations.gov.

Thank you.



C G Spies

860-583-5739

cgspies47@gmail.com

6/9/11
Sandra Neily
POB 102
Greenville Maine 04441
neily.sandra26@gmail.com

Comments on proposal: Docket No. [FWS-R3-ES-2011-0029]

I attended the informational session and public hearing offered by USFW in Augusta, Maine (6/8/11) and appreciated the service's outreach work to bring this issue to us.

Several unresolved issues arise.

1. The current proposal to delist gray wolves in the eastern region (Great Lakes and remaining eastern states) without an official designation or classification that would protect dispersing wolves from Canada to Maine or other eastern states seems is a serious omission. Given the current anti-coyote climate in Maine where coyotes are blamed for serious declines in the deer herd, any dispersing wolf could easily be killed 365 days a year as wolves fall within the "similar appearance" category in relation to coyotes. (State and NGO biologists are clear: the accumulated loss of winter shelter from years of unrestrained harvesting on industrial lands has left deer without winter shelter and protected corridors making the herd more vulnerable in some areas to predation and severe winters. Anti-coyote attitudes threaten wolf dispersal.)
2. If this same situation (unprotected status during a study period) had been present in adjoining states/regions, it's less likely wolves would have successfully dispersed from Minnesota to Wisconsin and Michigan or dispersed from Idaho to Washington. Dispersing populations must be protected. With seven wolves killed in northeastern states in the past decade, it is clear the region has the

Comment on FR Doc # 2015-26874

Posted by the Federal Aviation Administration on Dec 8, 2015

Docket / Document (FAA-2015-4378-0022) / Comment



Comment ID

FAA-2015-4378-2792

Comment

There should be no regulations



Tracking Number

1jz-8m35-3ip9

box3616#gmail.com

[Comment Details](#)

[Submitter Info](#)

Extract information from the following public comment on regulation.

Please list:

- The submitter's name, if available
- Whether the submitter is FOR or AGAINST regulation
- The submitter's email address

COMMENT TEXT:

There should be no regulations

box3616#gmail.com

The email address listed is not valid,
please fix it



I apologize for any confusion. Given that the email address provided ("box3616#gmail.com") appears incorrect, it's likely intended to be "box3616@gmail.com".





```
class Comment(BaseModel):
    name: Optional[str] = Field(description="Submitter's name")
    email: str = Field(description="Submitter's email address")
    position: Literal["for", "against", "neutral"]

    @field_validator("email")
    @classmethod
    def validate_email(cls, value):
        if value and '@' not in value:
            raise ValueError("Email missing @")
        return value
```



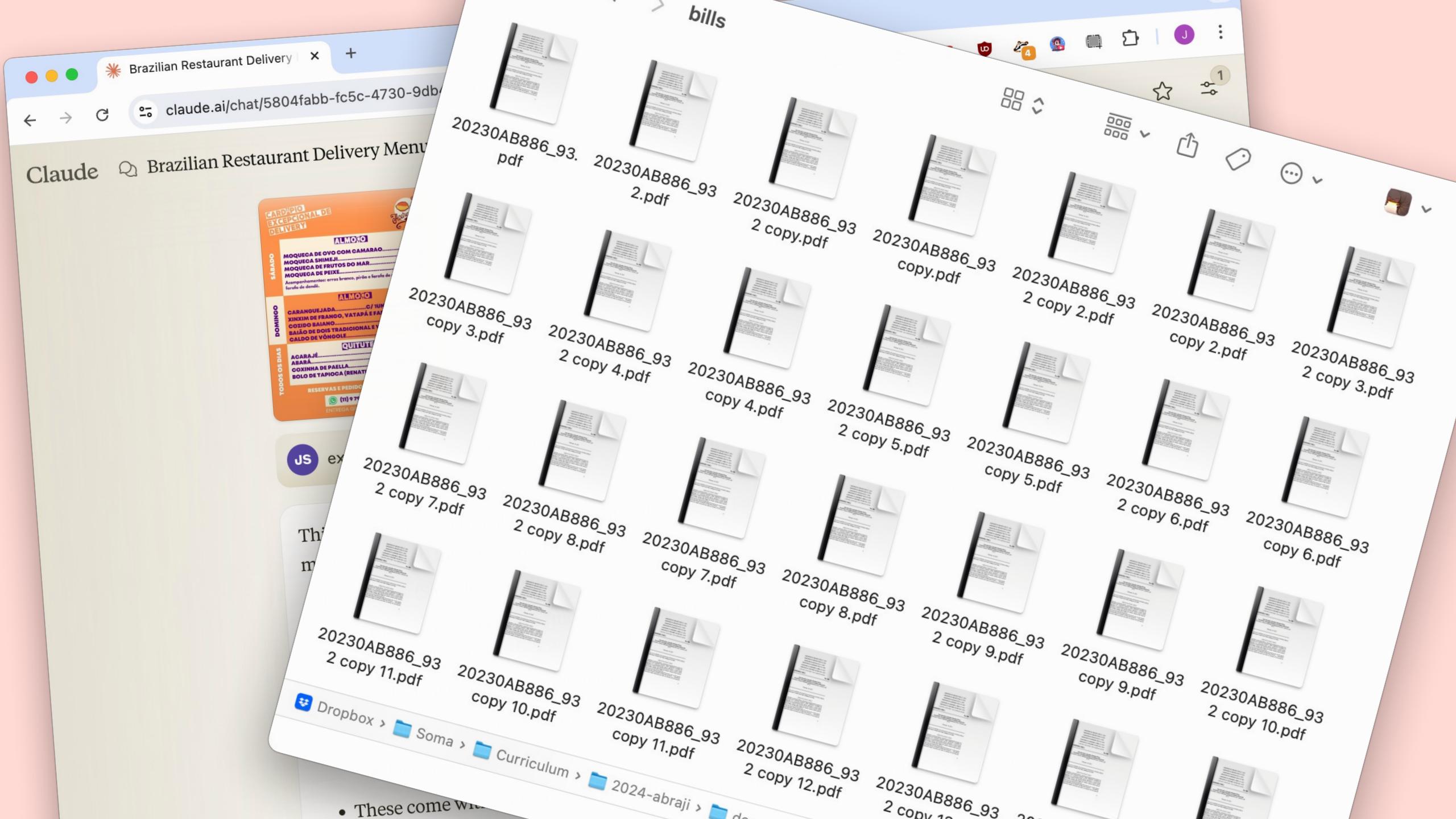
```
result = client.chat.completions.create(
    model="gpt-3.5-turbo",
    response_model=Comment,
    messages=[{"role": "user", "content": comment}],
    max_retries=3
```

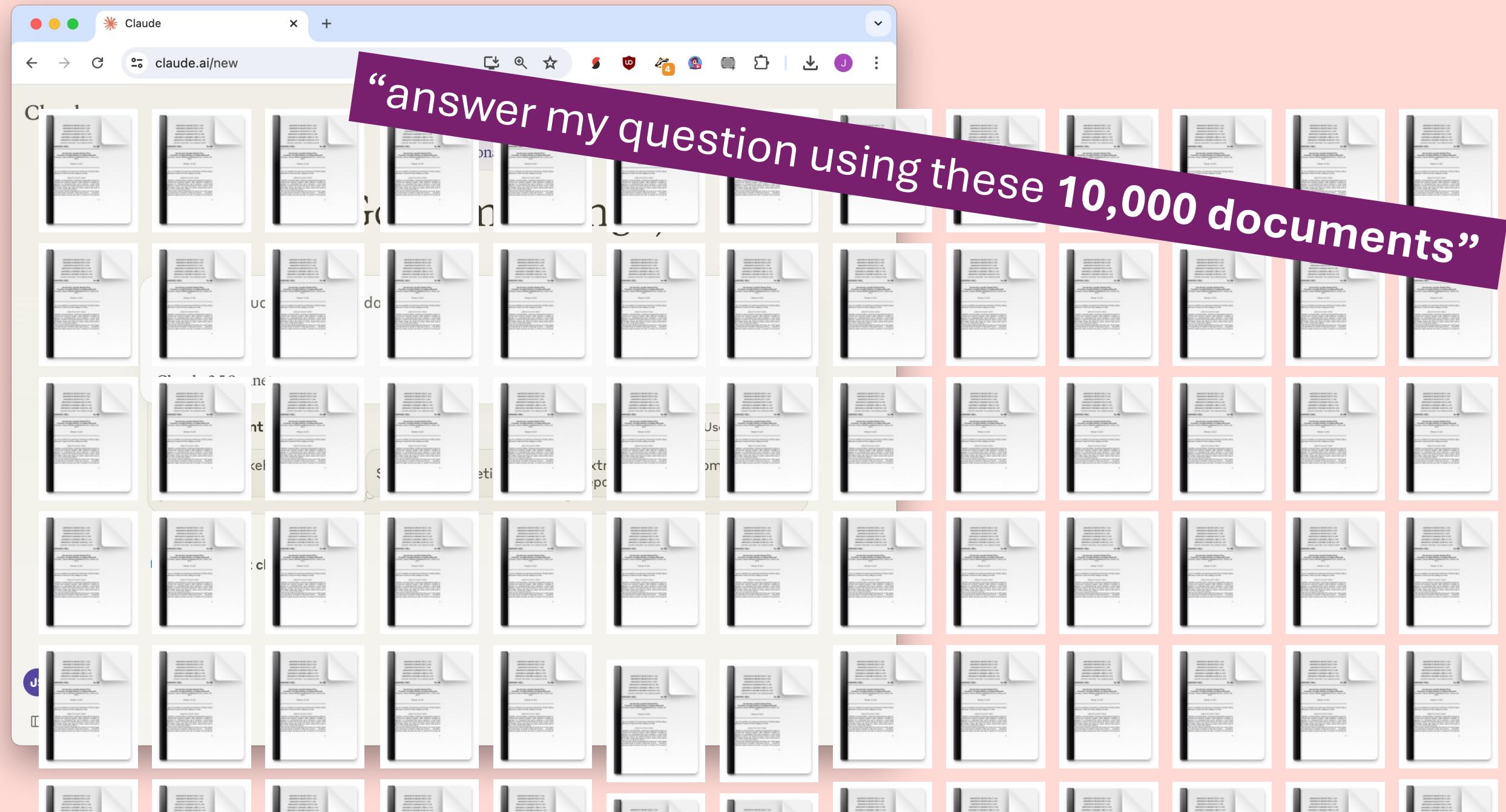


<https://bit.ly/abradi24-audio-video>

Activity:
Maybe we have time
to try it????

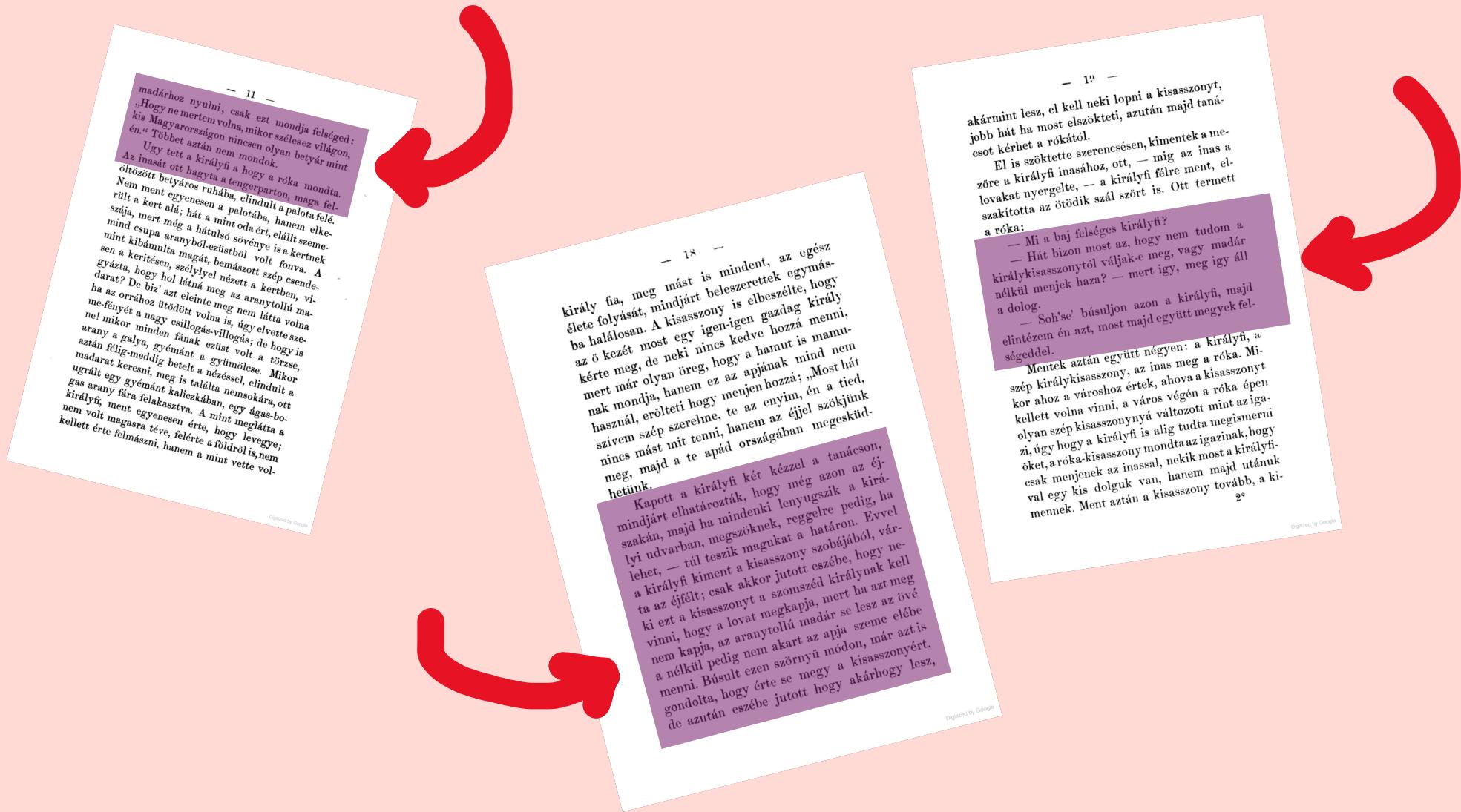
**But sometimes
you just want
one answer**





**Retrieval augmented
generation, also
known as “chatting
with your PDFs”**

Slice up your document, send top relevant sections with your question



But I'm a little worried the answer is going to be wrong!!!!

Did the AI actually find relevant documents?

Did we send too few sections, so it doesn't have the answer?

Did we send too many sections, confusing the AI?

Did it hallucinate even if it had the right content?

***Chatbots May ‘Hallucinate’
More Often Than Many Realize***

When summarizing facts, ChatGPT technology makes things up about 3 percent of the time, according to research from a new start-up. A Google

MyCity Chatbot x +

chat.nyc.gov

NYC MyCity Official website of the City of New York

ChatbotBeta

Select Language

FEEDBACK

MyCity Chatbot Beta

We are continuously working to improve the MyCity Chatbot, which uses business-related NYC agency data and AI to answer your questions. Your feedback is invaluable for refinement.

Example Questions

"How do I avoid noise violations and complaints for my construction company?"

"What information can you help me with?"

"How do I apply for the MWBE program?"

Capabilities

Trained to provide you with official NYC Business information.

Will not use the contents of your chat history to learn new information.

Responds to languages required by Local Law 30.

Limitations

May occasionally produce incorrect, harmful or biased content.

Limited knowledge of the world beyond NYC Business topics.

Trained to decline inappropriate requests.

© 2024 City of New York. All Rights Reserved.

Terms of Use Privacy Policy

Ask MyCity Chatbot business-related questions only

IMPORTANT: Responses may occasionally produce inaccurate or incomplete content. Validate answers on NYC.gov and MyCity Business.



Menu

The Mar

Challenging technology to ser

News

NYC's AI Chatbot Tells Bus the Law

The Microsoft-powered bot says bosses can tak
can discriminate based on source of income

By Colin Lecher

March 29, 2024 06:00 ET



? Question Submitted

Are buildings required to accept section 8 vouchers?

NYC Chatbot Answer (Emphasis by The Markup)

"No, buildings are not required to accept Section 8 vouchers."

Reality

Landlords **cannot discriminate by source of income**, with a minor exception for small buildings where the landlord or their family lives.

Do landlords have to accept tenants on rental assistance?

"No, landlords are not required to accept tenants on rental assistance."

Landlords **cannot discriminate by source of income**, with a minor exception for small buildings where the landlord or their family lives.

Can I keep my funeral home pricing confidential?

"Yes, you can keep your funeral home pricing confidential. There are no specific regulations that mandate the disclosure of funeral home pricing in New York City."

The Federal Trade Commission has outlawed concealing funeral prices.

Can I make my store cashless?

"Yes, you can make your store cashless in New York City. There are no specific requirements that prohibit businesses from operating as cashless establishments in the city."

Stores have been **required to accept cash** as payment since 2020.

**But R.A.G. is 100%
fun and easy, so let's
be irresponsible and
see how to do it.**

Llamaindex - Llamaindex

docs.llamaindex.ai/en/stable/#getting-started

Llamaindex

Getting Started

Llamaindex is available in Python (these docs) and [TypeScript](#). If you're not sure where to start, we recommend reading [how to read these docs](#) which will point you to the right place based on your experience level.

30 second quickstart

Set an environment variable called `OPENAI_API_KEY` with an [OpenAI API key](#). Install the Python library:

```
pip install llama-index
```

Put some documents in a folder called `data`, then ask questions about them with our famous 5-line starter:

```
from llama_index.core import VectorStoreIndex, SimpleDirectoryReader
documents = SimpleDirectoryReader("data").load_data()
index = VectorStoreIndex.from_documents(documents)
query_engine = index.as_query_engine()
response = query_engine.query("Some question about the data should go here")
print(response)
```

If any part of this trips you up, don't worry! Check out our more comprehensive starter tutorials using [remote APIs like OpenAI](#) or [any model that runs on your laptop](#).

LlamaCloud

Table of contents

- Introduction
 - What is context augmentation?
- Llamaindex is the Data Framework for Context-Augmented LLM Apps
- Use cases
 - Who is Llamaindex for?
- Getting Started
- 30 second quickstart
- LlamaCloud
- LlamaParse
- Community
- Getting the library
- Contributing
- Related projects

AnythingLLM | The ultimate AI business intelligence tool.

useanything.com

AnythingLLM

AnythingLLM Cloud Docs Github Download AnythingLLM

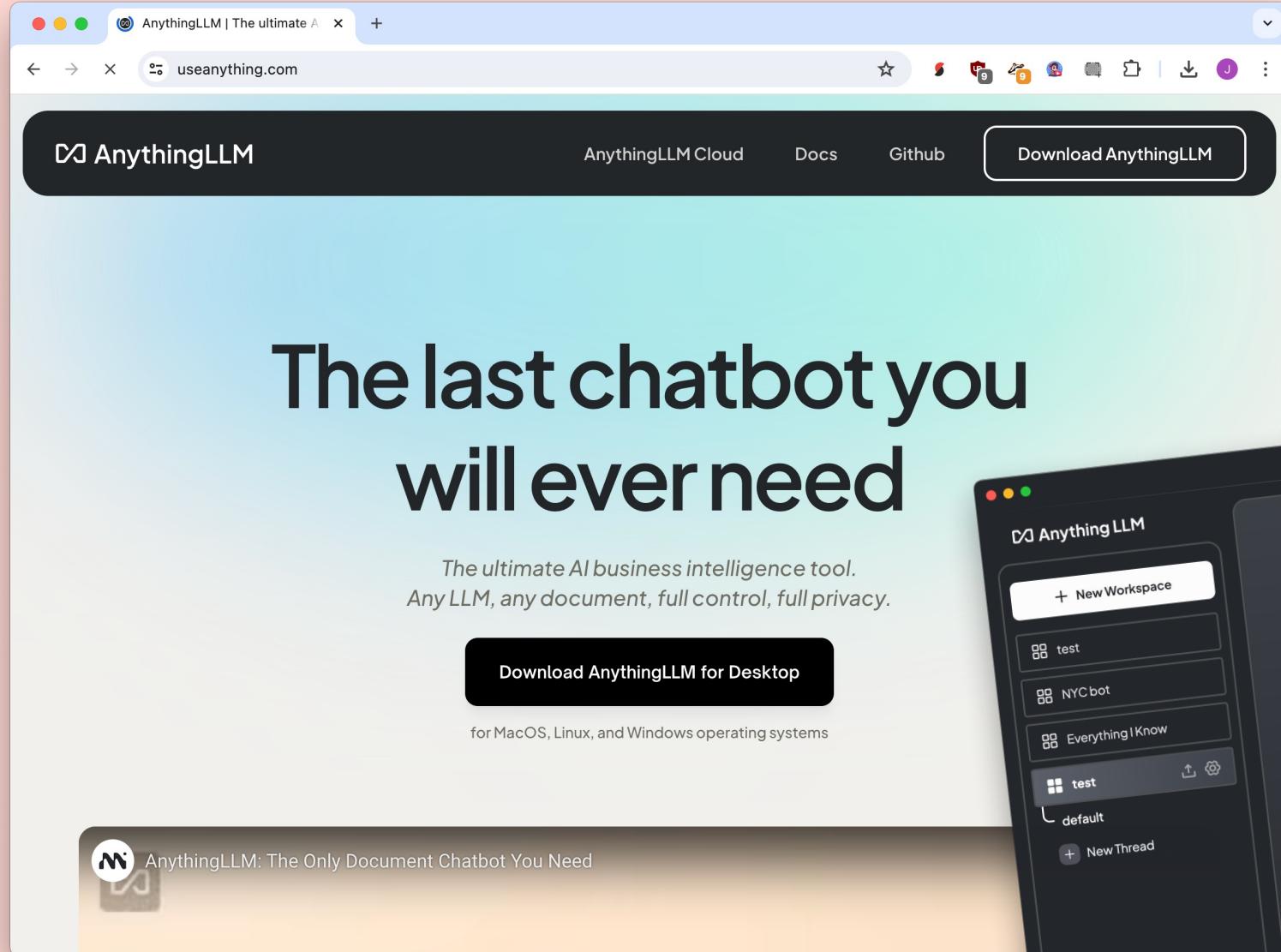
The last chatbot you will ever need

*The ultimate AI business intelligence tool.
Any LLM, any document, full control, full privacy.*

Download AnythingLLM for Desktop

for MacOS, Linux, and Windows operating systems

AnythingLLM: The Only Document Chatbot You Need



[+ New Workspace](#)[test](#)

upload documents!

Name

 custom-documents everything adding-zip-codes-to-lat-long.md

Cached

 annotations

Cached

 annotations.md

Cached

 arcgis-rest.md

Cached

 auto-updating-scaper-viz.md

Cached



Click to upload or drag and drop

supports text files, csv's, spreadsheets, audio files, and more!

or submit a link

Fetch website

These files will be uploaded to the document processor running on this AnythingLLM instance.
These files are not sent or shared with a third party.



GitHub Repo

Import an entire public or private Github repository in a single click.



YouTube Transcript

Import the transcription of an entire YouTube video from a link.



Bulk Link Scraper

Scrape a website and its sub-links up to a certain depth.



Confluence

Import an entire Confluence page in a single click.

Website URL

URL of the website you want to scrape.

https://example.com

Depth

This is the number of child-links that the worker should follow from the original page.

1

Max Links

Maximum number of links to scrape.

20

Submit

scrape websites!

do a better job
than an NYC
agency that
was paid a lot
of money!

Anything LLM

+ New Workspace

Can I deny renters based on Section 8 vouchers

No, you cannot deny renters based on Section 8 vouchers. It is against the law to discriminate based on a current or prospective tenant's source of income, including housing vouchers.

Show Citations >

but can i consider credit scores?

If the voucher/subsidy covers 100% of the rent, you are prohibited from considering credit. However, if the tenant pays a portion of the rent, you may consider credit. Each application should be considered on a case-by-case basis.

Show Citations >

Send a message

/ @ T





Can I deny renters based on Section 8 vouchers



No, you cannot deny renters based on Section 8 vouchers. It is against the law to discriminate based on a tenant's source of income, including Section 8.



Hide Citations ▾



SourceOfIncomeFact...eet.pdf



FairH...



Section 8 Guide fo...nts.pdf

FairHouse_FAQBroker-2021Eng.pdf

Referenced 2 times.

March 2023
NYC.gov/HumanRights
@NYCCHR

1 BEST PRACTICES FOR LICENSED SALESPEOPLE AND BROKERS TO AVOID SOURCE OF INCOME DISCRIMINATION

Yes. You are responsible for the actions of anyone who works under your broker's license or holds themselves out as an agent of your brokerage company. What should I do if an owner refuses to accept housing vouchers/subsidies? You should not work with owners who discriminate based on a current or prospective tenant's source of income. Source of income discrimination must be treated just like any other kind of discrimination. It is against the law. Can I keep a separate list ("waitlist") of apartments for voucher holders? No. Maintaining separate lists of apartments for voucher/subsidy recipients is a discriminatory practice. Steering applicants to specific buildings not only constitutes discrimination, but may also be considered aiding and abetting the discrimination of owners.

60% match

Citations aren't reliable, so be sure to fact-check!

Evaluating Verifiability in Generative Search Engines

Nelson F. Liu* Tianyi Zhang Percy Liang

Department of Computer Science
Stanford University

nfliu@cs.stanford.edu



Nelson Liu
@nelsonfliu

...

Generative search engines are transforming how we find info, but are they trustworthy?

We evaluate Bing Chat, NeevaAI, [perplexity.ai](#) & YouChat—only 52% of statements are supported by cites and 75% of cites actually support their statements.

arxiv.org/abs/2304.09848



What are the latest discoveries from the James Webb Space Telescope?



The James Webb Space Telescope is designed to peer into the dusty clouds of gas where stars and planetary systems are born. Webb has captured the first direct image of an exoplanet, and the Pillars of Creation in the Eagle Nebula [1][2]. Additionally, the telescope will be used to study the next interstellar interloper [3].

(*Some generated statements may *not* be fully supported by citations, while others are fully supported.)

Cited Webpages

[1]: nasa.gov (✖citation does not support its associated statement)
[NASA's Webb Confirms Its First Exoplanet](#)
... Researchers confirmed an exoplanet, a planet that orbits another star, using NASA's James Webb Space Telescope for the first time. ...

[2]: cnn.com (⚠citation partially supports its associated statement)
[Pillars of Creation: James Webb Space Telescope ...](#)
... The Pillars of Creation, in the Eagle Nebula, is a star-forming region captured in a new image (right) by the James Webb Space Telescope that reveals more detail than a 2014 image (left) by Hubble ...

[3]: nasa.gov (✓citation fully supports its associated statement)
[Studying the Next Interstellar Interloper with Webb](#)
... Scientists have had only limited ability to study these objects once discovered, but all of that is about to change with NASA's James Webb Space Telescope...The team will use Webb's spectroscopic capabilities in both the near-infrared and mid-infrared bands to study two different aspects of the interstellar object.

JS

You

where did he get his masters?



ChatGPT

Jonathan Soma earned his Master's degree from the University of Denver.



JS

You

Where did he get his master's degree?



ChatGPT

Jonathan Soma earned his Master's degree from the Columbia University Graduate School of Journalism [1].

where did Jonathan Soma get his masters degree? you can search the internet for it



Searched 4 sites

Jonathan Soma does not hold a master's degree from any institution. He himself clarified that responses generated about his educational background by AI tools such as ChatGPT often inaccurately list institutions like the University of Denver or Columbia Graduate School of Journalism, but he does not actually have a master's degree from any of these places ([LaptopMag](#)) ([Columbia Journalism School](#)).



youtube.com/playlist?list=PLewNEVDy7gq1_GPUaL0OQ31QsiHP5ncAQ

YouTube BE

Search

Home

Shorts

Subscriptions

You >

Your channel

History

Playlists

Your videos

Watch later

Liked videos

Your clips

Dearest friend, I daresay I have not partaken of food in ages. I'm positively famished

Practical AI for Investigative Journalism

Jonathan Soma

Playlist • Public • 6 videos • 669 views

A six session series held in April 2024 about real-life use cases for journalism in (mostly investigative) journa...more

Play all

Sort

Dearest friend, I daresay I have not partaken of food in ages. I'm positively famished 2:09:33

Sorting documents (Practical AI for Investigative Journalism, Session 1)

Jonathan Soma • 1.1K views • Streamed 2 months ago

Large language models don't understand facts or concepts, they only know statistical probability 2:14:32

Structured, validated data from LLMs (Practical AI for Investigative Journalism, Session 2)

Jonathan Soma • 743 views • Streamed 2 months ago

Evaluating Verifiability in Generative Search Engines

Jonathan Soma • 581 views • Streamed 2 months ago

LLM citations are misinformation 1:45:26

Why generative AI is a dead end for responsible journalism (Practical AI for Journalism, Session 3)

Jonathan Soma • 581 views • Streamed 2 months ago

AI, Hugging Face and non-chatbot models (Practical AI for Journalism, Session 4)

Jonathan Soma • 836 views • Streamed 2 months ago

Context window

Local models/private AI (Practical AI for Investigative Journalism, Session 5)

Jonathan Soma • 451 views • Streamed 2 months ago

Transcription and audio models (Practical AI for Investigative Journalism, Session 6)

Jonathan Soma • 332 views • Streamed 2 months ago

Downloading and
manipulating videos

yt-dlp/yt-dlp: A feature-rich command-line audio/video downloader

YT-DLP A feature-rich command-line audio/video downloader

DOWNLOAD V2024.04.09 PYPI DONATE 180 USERS
2412 ONLINE SUPPORTED SITES UNLICENSE TESTS PASSING
COMMITS 40/MONTH LAST SUNDAY

yt-dlp is a feature-rich command-line audio/video downloader with support for thousands of sites. The project is a fork of [youtube-dl](#) based on the now inactive [youtube-dlc](#).

- INSTALLATION
 - [Detailed instructions](#)
 - [Release Files](#)
 - [Update](#)
 - [Dependencies](#)
 - [Compile](#)
- USAGE AND OPTIONS
 - [General Options](#)
 - [Network Options](#)
 - [Geo-restriction](#)
 - [Video Selection](#)
 - [Download Options](#)

yt-dlp

You how to use `yt-dlp` to download the audio from <https://www.youtube.com/watch?v=YTTUS8Koi4k>

ChatGPT To download the audio from a YouTube video using `'yt-dlp'`, you first need to have `'yt-dlp'` installed on your computer. If it's not already installed, you can install it via pip if you have Python:

```
bash
pip install -U yt-dlp
```

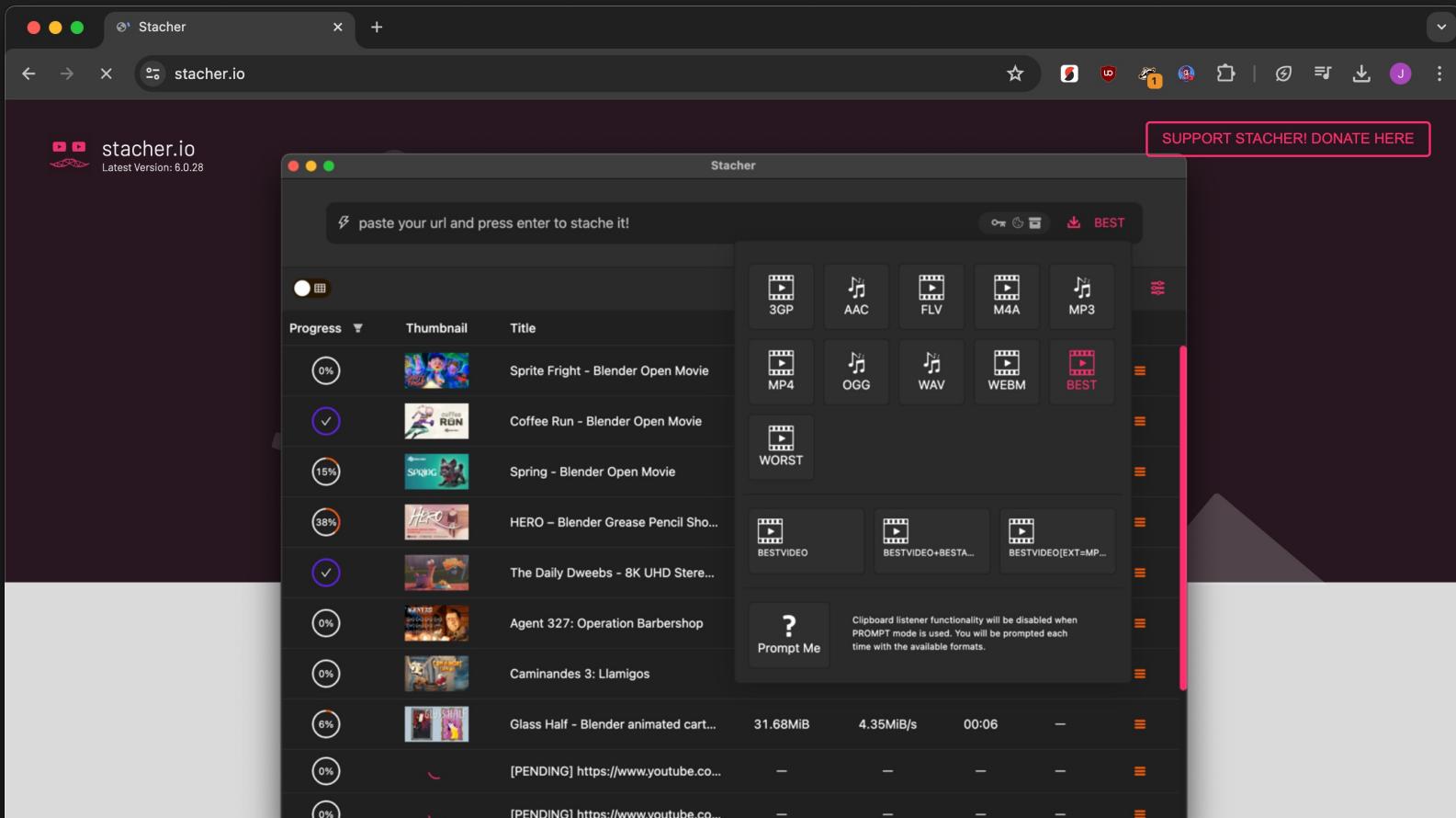
After installing `'yt-dlp'`, you can use the following command to download just the audio from the YouTube video:

```
bash
yt-dlp -f bestaudio https://www.youtube.com/watch?v=YTTUS8Koi4k -x --audio-format mp3
```

Here's what each part of the command does:

- `'-f bestaudio'`: This option tells `'yt-dlp'` to select the best audio quality available.
- `'-x'`: This option extracts the audio from the video.
- `'--audio-format mp3'`: This specifies the desired audio format to convert the audio to (e.g., mp3, aac, etc.).

This command will download and convert the audio to an MP3 file. The resulting file will be saved in the directory from which you ran the command.



A BEAUTIFUL, MODERN GUI FOR YT-DLP

Stacher will automatically install YT-DLP on startup, however you can customize Stacher to use any fork you prefer.

<https://stacher.io/>

JS

You

command to use ffmpeg and extract an image from a video every half-second

GPT

ChatGPT

To extract an image from a video every half-second using FFmpeg, you can use the following command:

bash

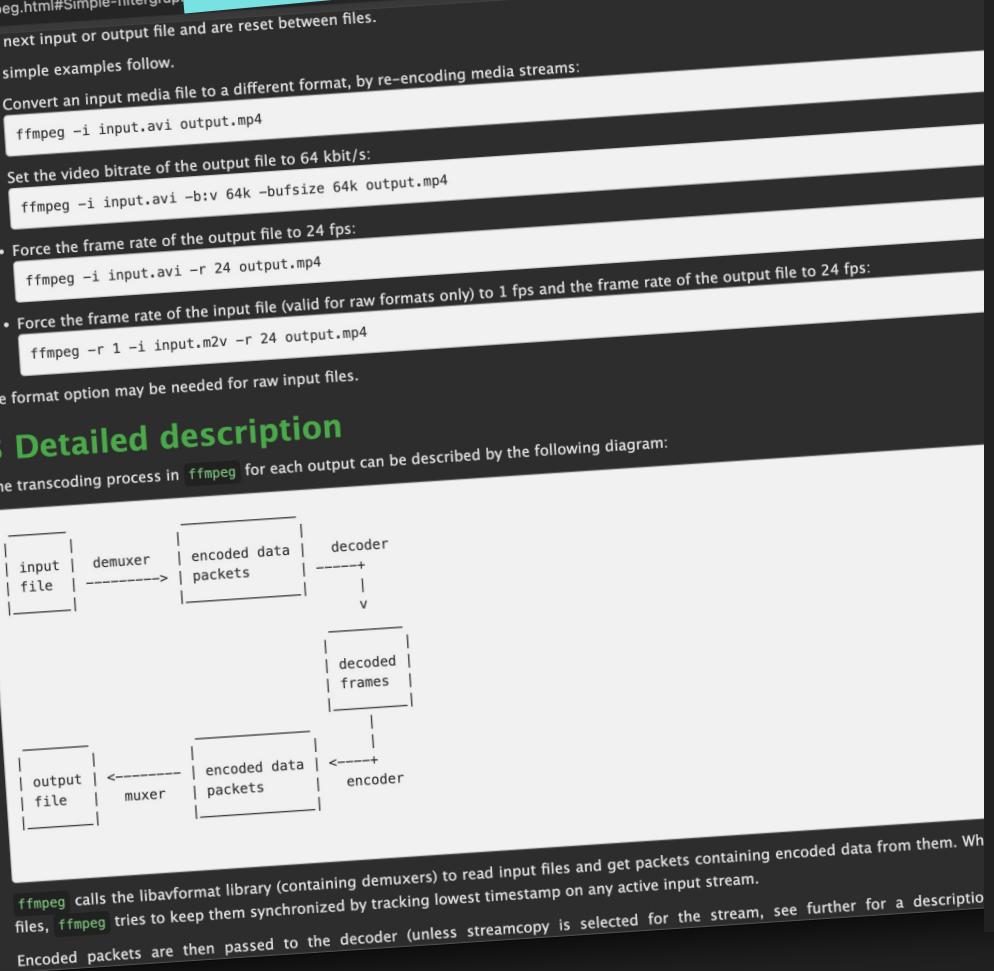
Copy code

```
ffmpeg -i input_video.mp4 -vf "fps=2" output_frame_%04d.jpg
```

Here's a breakdown of the command:

- `ffmpeg`: This is the basic command to run FFmpeg.
- `'-i input_video.mp4'`: This specifies the input video file. Replace `input_video.mp4` with the path to your video file.
- `'-vf "fps=2"'`: This video filter sets the frame rate of the output images. `fps=2` means FFmpeg will extract two frames per second, which is one frame every half-second.
- `output_frame_%04d.jpg`": This is the output file pattern. `%04d` is a placeholder that FFmpeg will replace with the frame number, padded with zeros to four digits (e.g., `output_frame_0001.jpg`, `output_frame_0002.jpg`, etc.).

Make sure to replace `input_video.mp4` with the actual path and filename of your video. This command will create a series of JPEG images in the same directory where you run the command, each one captured every half-second of the video.



shutterencoder.com/en/

A converter designed by video editors.
The professional compression tool accessible to all.

Downloads

More than 844,307 downloads

A Simple and Complete Interface

<https://www.shutterencoder.com>

Shutter Encoder encoding/convert

ffworks.net

Introducing ff·Works 4

A Comprehensive Media Tool for macOS

OSBytes

Try Now

Buy Now

<https://ffworks.net>

yt-dlp + ffmpeg =
everything is possible

Turning videos into data

The CNN Presidential debate

youtube.com/shorts/rDXubdQdJYs

YouTube BR

Search

Home

Shorts

Subscriptions

You

75K

Dislike

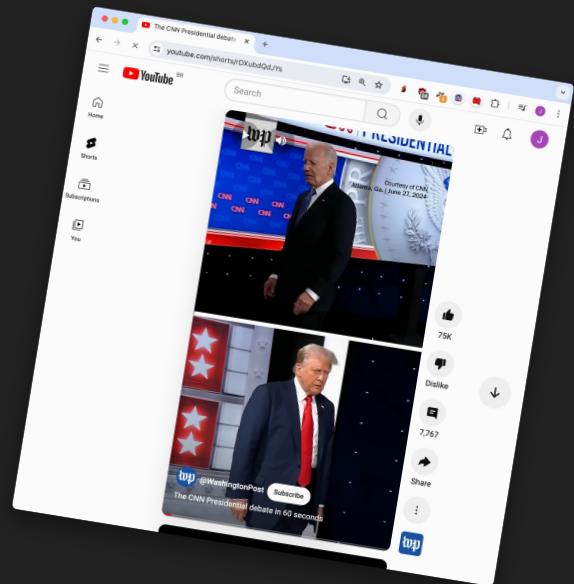
7,767

Share

wp @WashingtonPost

The CNN Presidential debate in 60 seconds

“I want a model to see
if media is biased
towards Joe Biden or
Donald Trump”



“I want to see if Joe Biden or Donald Trump gets more time in the media”

Download the video



Separate into scenes

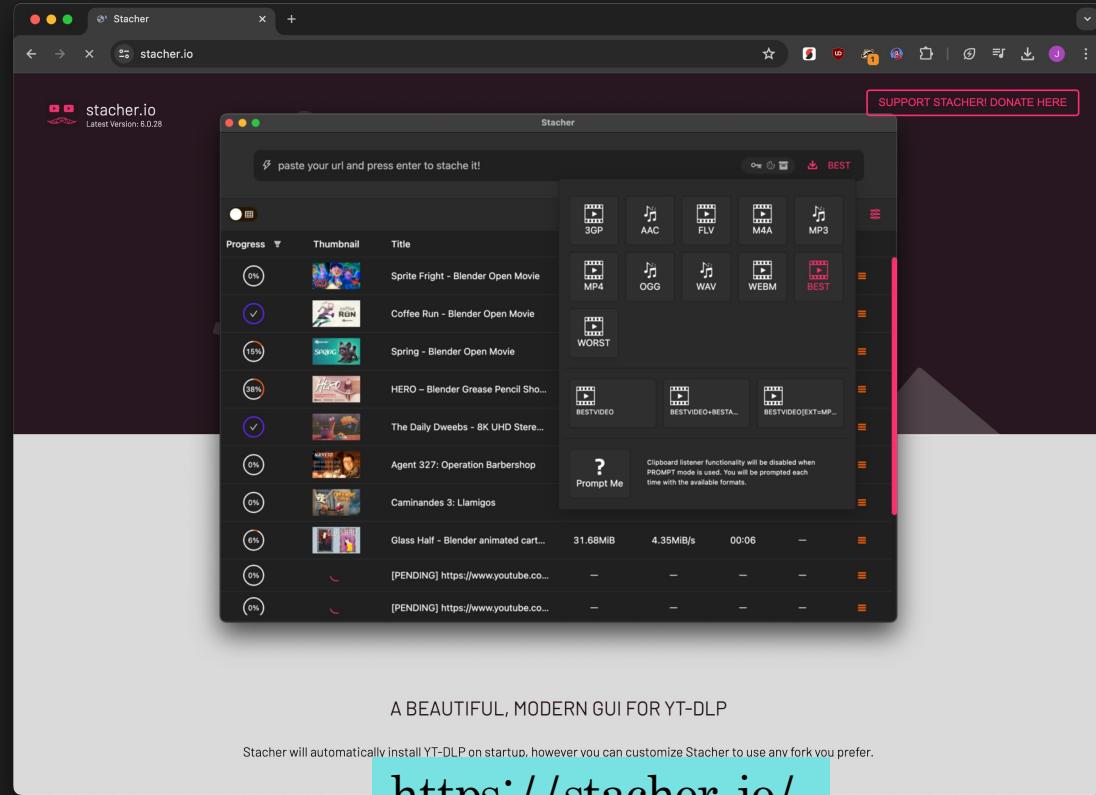
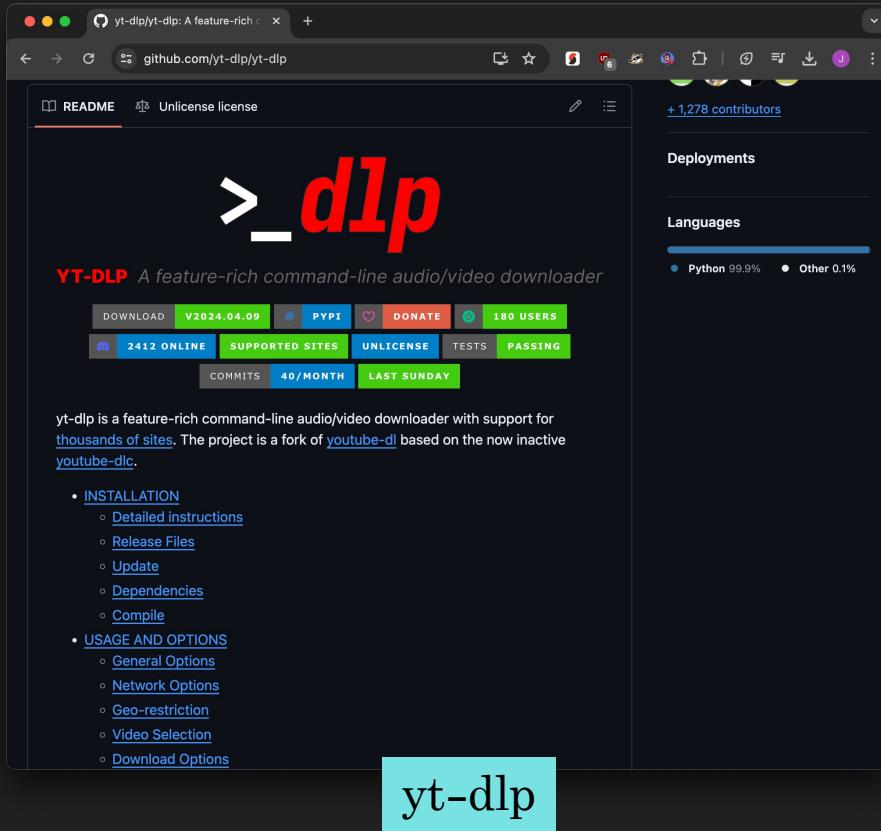


Is the scene Biden or Trump?



Count it up

Download the video



Split into scenes (maybe?)

The screenshot shows a web browser displaying the PySceneDetect documentation site. The page has a light blue header with the PySceneDetect logo and navigation links. The main content area features a large image of a clapperboard, followed by the title "PySceneDetect". Below the title is a green callout box containing release information and download links. The main text explains what PySceneDetect is and how it works. There are sections for "Quickstart" and examples of command-line and Python API usage. A footer at the bottom of the page provides a link to the website.

PySceneDetect

Search docs

PYSCENEDETECT

Home

Features

Download

Changelog

REFERENCE:

Documentation

Command-Line

Python API

SUPPORT:

FAQ

Bugs & Contributing

Development & Support

RESOURCES

Similar Projects

Research & Literature

License & Copyright

Next »

» PySceneDetect » / Home

Edit on PySceneDetect on Github

Latest Release: v0.6.4 (June 10, 2024)

Download Changelog Documentation Getting Started

See the changelog for the latest release notes and known issues.

PySceneDetect is a tool for detecting shot changes in videos ([example](#)), and can automatically split the video into separate clips. PySceneDetect is free and open-source software, and has several [detection methods](#) to find fast-cuts and threshold-based fades.

Quickstart

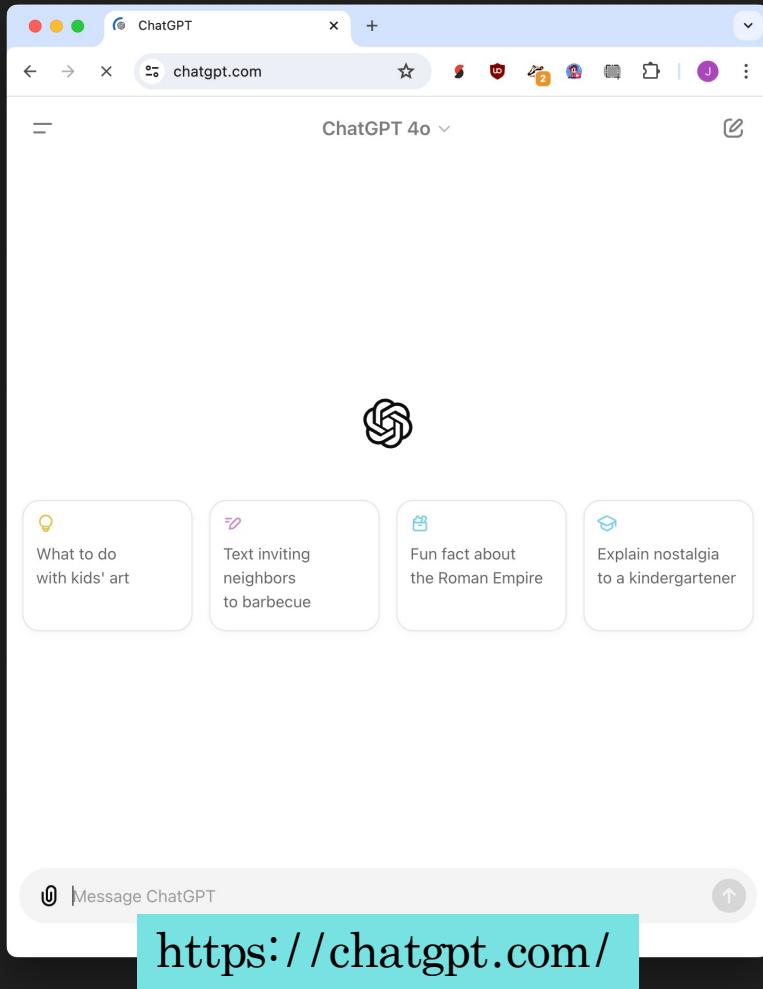
Split video on each fast cut using [command line \(more examples\)](#):

```
scenedetect -i video.mp4 split-video
```

Split video on each fast cut using [Python API \(docs\)](#):

<https://www.scenedetect.com/>

Is it Biden or Trump?



A screenshot of the Hugging Face website showing the "Zero-Shot Image Classification" task page. The URL in the address bar is <https://huggingface.co/tasks/zero-shot-image-classification>. The page has a header with the Hugging Face logo and a search bar. Below the header, there's a section titled "Zero-Shot Image Classification" with a brief description: "Zero-shot image classification is the task of classifying previously unseen classes during training of a model." To the right, there's a diagram illustrating the process: an "Inputs" box contains a photo of a cat, with the text "Classes: cat, dog, bird". An arrow points from this box to a central yellow box labeled "Zero-Shot Image Classification Model". Another arrow points from the model box to an "Output" box on the right, which lists the results:

Class	Score
Cat	0.664
Dog	0.329
Bird	0.008

. Below this diagram, there's a section titled "About Zero-Shot Image Classification" and another titled "About the Task". On the right side of the page, there's a sidebar titled "Compatible libraries" with links to "Transformers" and "Transformers.js". Below that, there's a "Zero-Shot Image Classification demo" section featuring the "openai/clip-vit-large-patch14-336" model, with a note that it's incompatible with the Inference API (serverless). Further down, there's a "Models for Zero-Shot Image Classi..." section with a link to "openai/clip-vit-base-patch16". A teal footer bar at the bottom contains the URL <https://huggingface.co/tasks/zero-shot-image-classification>.

Multimodal

-  Image-Text-to-Text
 -  Visual Question Answering
 -  Document Question Answering
- Computer Vision
-  Depth Estimation
 -  Image Classification
 -  Object Detection
 -  Image Segmentation
 -  Text-to-Image
 -  Image-to-Text
 -  Image-to-Image
 -  Image-to-Video
 -  Unconditional Image Generation
 -  Video Classification
 -  Text-to-Video
 -  Zero-Shot Image Classification
 -  Mask Generation
 -  Zero-Shot Object Detection
 -  Text-to-3D
 -  Image-to-3D
 -  Image Feature Extraction

Natural Language Processing

-  Text Classification
 -  Token Classification
 -  Table Question Answering
 -  Question Answering
 -  Zero-Shot Classification
 -  Translation
 -  Summarization
 -  Feature Extraction
 -  Text Generation
 -  Text2Text Generation
 -  Fill-Mask
 -  Sentence Similarity
- Audio
-  Text-to-Speech
 -  Text-to-Audio
 -  Automatic Speech Recognition
 -  Audio-to-Audio
 -  Audio Classification
 -  Voice Activity Detection
- Tabular
-  Tabular Classification
 -  Tabular Regression

Hugging Face Search models, datasets, users...

Models Datasets Spaces Posts Docs Pricing

openai/clip-vit-large-patch14 like 1.25k

Zero-Shot Image Classification Transformers PyTorch TensorFlow JAX Safetensors clip vision Inference Endpoints arxiv:2103.00020 arxiv:1908.04913

Model card Files and versions Community 26

Edit model card Downloads last month 41,110,144

Safetensors Model size 428M params Tensor type I64 · F32

Inference API Zero-Shot Image Classification Drag image file here or click to browse from your device

Possible class names... Compute

This model is currently loaded and running on Inference API (serverless).

```
# Use a pipeline as a high-level helper
from transformers import pipeline

pipe = pipeline("zero-shot-image-classification", model="openai/clip-vit-large-patch14")
```

Deploy Use this model Libraries Transformers

A large red curved arrow points from the top right towards the 'Use this model' button. Another red arrow points from the bottom right towards the same button, which is highlighted with a blue border.

⚡ Inference API ⓘ

Zero-Shot Image Classification



Possible class names (comma-separated)

donald trump, joe biden

Compute

Computation time: 0.352 s

donald trump

joe biden

</> JSON Output

⚡ Inference API ⓘ

Zero-Shot Image Classification



Possible class names (comma-separated)

donald trump, joe biden

Compute

Computation time: 0.443 s

donald trump

joe biden

</> JSON Output

Maximize

Maximize



Count it up

6 + | + | + 2 + 5



<https://bit.ly/abraj24-audio-video>

Activity:
Let's do this example!



<https://bit.ly/abraj24-audio-video>

Activity:
Try your own?

Custom classification models

The examples here are for images, but they'll work just as well for individual frames of video!

How They Did It: Uncovering a Vast Network of Illegal Mining in Venezuela

gijn.org/stories/how-they-did-it-uncovering-a-vast-network-of-illegal-mining-in-venezuela/

Global Investigative Journalism Network

Search

STORIES

TOPICS

» How They Did It

» Methodology

FURTIVO

How They Did It: Uncovering a Vast Network of Illegal Mining in Venezuela

by Mariel Lozada • June 2, 2022

In the Venezuelan Amazon, many hours away from the country's capital, traces of the devastation caused by illegal mining can be seen from the sky.

Although this area is officially protected, and home to a number of Indigenous communities, organized crime groups have seized control of illicit quarries, destroying the rainforest, and even built runways to take their bounties overseas, according to NGOs like Human Rights Watch. While observers suspected that the country's devastating economic crisis had made the problem of illegal mining worse, it was hard to grasp the real magnitude of the problem.

But a collaboration between the Spanish newspaper El País and the Venezuelan investigative outlet Armando.info — a GIJN member — used satellite data, machine learning, and field reporting to try and paint a definitive picture of the problem and to establish the driving forces behind it. They identified over 3,000 mining locations — most of them illegal — and

SHARE

Tweet Facebook Print LinkedIn WhatsApp Telegram Reddit Republish

RELATED RESOURCES

Investigating the Israel-Hamas Conflict

Investigative Journalism and Digital Threats in 2024 Elections

GIJC23 - FOIA/RTI Investigations

A Times Investigation Tracked Israel's Use of One of Its Most Destructive Bombs in South Gaza

nytimes.com/2023/12/21/world/middleeast/israel-gaza-bomb-investigation.html

The New York Times

Israel-Hamas War | LIVE Updates 4m ago

How Hamas Fights in Gaza Displaced Gazans in Egypt What We Saw in Rafah Israeli Generals Want a Truce

GIVE THE TIMES Account

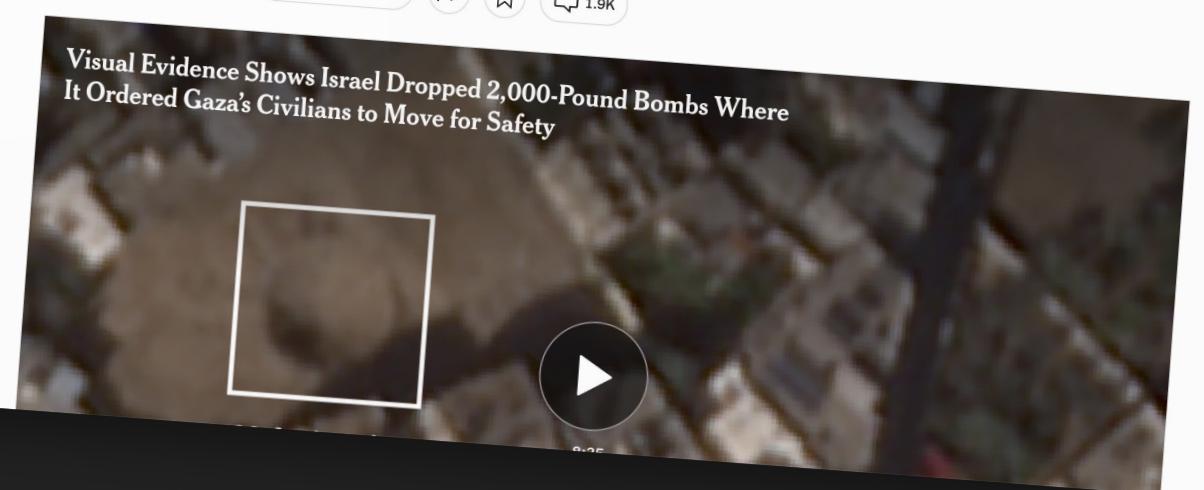
Screenshot

VISUAL INVESTIGATIONS

Share full article

1.9K

Visual Evidence Shows Israel Dropped 2,000-Pound Bombs Where It Ordered Gaza's Civilians to Move for Safety



Leprosy of the land

texty.org.ua/d/2018/amber_eng/

TEXTY.ORG.UA

Українськом

f t

Leprosy of the land



THOUSANDS INTO LUNAR ILLEGAL AME MODEL FOUND ON SATELLIT 70,000 SQUA

Files

- master
- + Q
- Go to file
- > doc
- > examples
- > model
- > datasets
- > onhold
- > versions
- features_backup.csv
- features_backup.csv.zip
- step1.ipynb
- step2.ipynb
- step3.ipynb
- visually_debug_model.ipynb
- xgb_tsne.html
- README.md
- requirements.txt

amber-methodology / model / step2.ipynb

Preview Code Blame 372 lines (372 loc) · 12.5 KB

```
resnet = ResNet50(weights='imagenet', include_top=False) # load ResNet50
```

Feature extraction

In [4]:

```
def extract_features(img_paths, batch_size=64):
    """ This function extracts image features for each image in img_paths using ResNet50.
        Returned features is a numpy array with shape (len(img_paths), 2048).
    """
    global resnet
    n = len(img_paths) # num of images in img_paths
    img_array = np.zeros((n, 224, 224, 3))

    for i, path in enumerate(img_paths):
        img = image.load_img(path, target_size=(224, 224)) # load and scale each image
        img = image.img_to_array(img)
        img = preprocess_input(img)
        img_array[i] = img

    X = resnet.predict(img_array, batch_size=batch_size, verbose=1)
    X = X.reshape((n, 2048))
    return X

# features for our two types of labels
positives_ = extract_features(positive_files)
negatives_ = extract_features(negative_files)
```

250/250 [=====] - 5s 22ms/step
900/900 [=====] - 13s 15ms/step

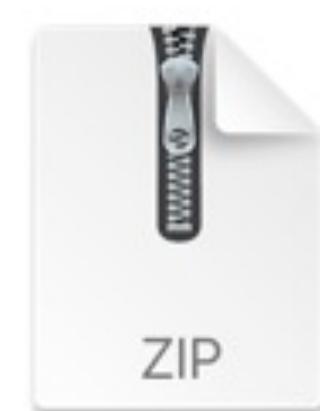
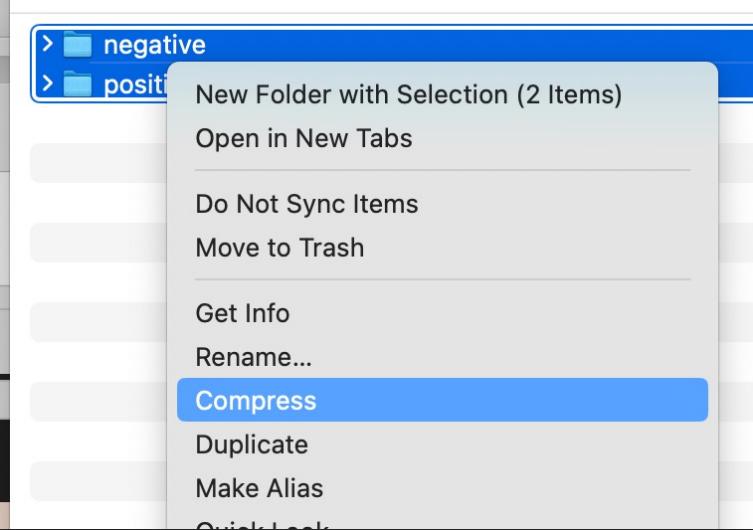
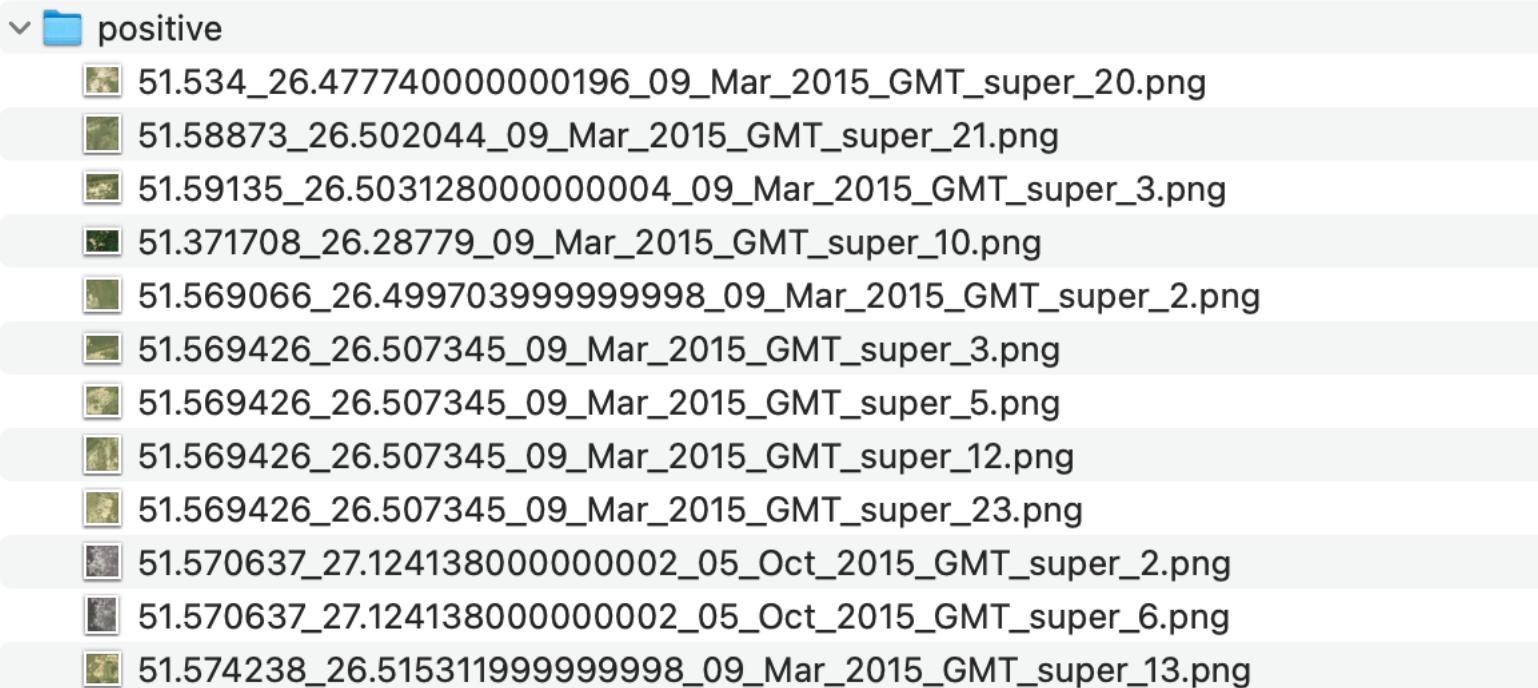
In [7]:

```
# Create dataframe from two types of features, for positive images (with amber min.
```

amber-methodology / model / step2.ipynb

The screenshot shows a web browser window for the Hugging Face platform, specifically the AutoTrain section. The URL in the address bar is <https://huggingface.co/autotrain>. The page features a large, colorful gradient background with a rainbow-like effect. At the top left, there's a yellow emoji icon followed by the text "AutoTrain - Hugging Face". The top navigation bar includes links for "Models", "Datasets", "Spaces", "Posts", "Docs", and "Pricing". A search bar at the top says "Search models, datasets, users...". Below the header, the word "auto TRAIN" is displayed, with "auto" in black and "TRAIN" in red. The main title "Create powerful AI models without code" is prominently shown in bold black text. A subtext below it reads "A new way to automatically train, evaluate and deploy state-of-the-art Machine Learning models." A "Create new project" button is located below this text. Three callout boxes highlight features: "Machine Learning made simple" (Train custom machine learning models by simply uploading data.), "Automatic training" (AutoTrain will find the best models for your data automatically.), and "Fast deployment" (Your models are available on the [Hugging Face Hub](#), and ready to serve.). At the bottom, a list of tasks available in AutoTrain is provided, including LLM Finetuning, Image classification, Text Classification, Token Classification, Question Answering (extractive), Translation, Summarization, Text Regression, Tabular Data Classification, and Tabular Data Regression.

<https://huggingface.co/autotrain>



Archive.zip

The screenshot shows the AutoTrain Advanced web interface for a project named "autotrain-zwu0s-3suak".

Project Configuration:

- Hugging Face User:** wendys-lhc
- Task:** Image Classification
- Hardware:** Local/Space
- Parameter Mode:** Basic

Training Parameters:

- Accelerators:** 0 (No running jobs)
- Project Name:** autotrain-zwu0s-3suak
- Base Model:** google/vit-base-patch16-224
- Dataset Source:** Local
- Parameters:** Mixed precision: none, Optimizer: adamw_torch, Scheduler: linear
- Training Data:** Validation Data (optional)
- Upload Training File(s):** Archive.zip
- Learning rate:** 0.00005

Column Mapping:

- image:** image
- label:** label

A large red arrow points from the left side of the screen towards the "Logs" link in the sidebar. Another large red arrow points from the bottom center towards the "Upload Training File(s)" button.

Multimodal

 Image-Text-to-Text

 Visual Question Answering

 Document Question Answering

Computer Vision

 Depth Estimation

 Image Classification

 Object Detection

 Image Segmentation

 Text-to-Image

 Image-to-Text

 Image-to-Image

 Image-to-Video

 Unconditional Image Generation

 Video Classification

 Text-to-Video

 Zero-Shot Image Classification

 Mask Generation

 Zero-Shot Object Detection

 Text-to-3D

 Image-to-3D

 Image Feature Extraction

Natural Language Processing

 Text Classification

 Token Classification

 Table Question Answering

 Question Answering

 Zero-Shot Classification

 Translation

 Summarization

 Feature Extraction

 Text Generation

 Text2Text Generation

 Fill-Mask

 Sentence Similarity

Audio

 Text-to-Speech

 Text-to-Audio

 Automatic Speech Recognition

 Audio-to-Audio

 Audio Classification

 Voice Activity Detection

Tabular

 Tabular Classification

 Tabular Regression

**What are your
questions and ideas?**



Transforming audio and video into useable data

Jonathan Soma

Knight Chair in Data Journalism, Columbia University

Abraji Congress 2024

Sunday, July 14

js4571@columbia.edu

<https://bit.ly/abraji24-audio-video>

Text-to-speech and deepfakes

TTS Arena - a Hugging Face Space

huggingface.co/spaces/TTS-AGI/TTS-Arena

TTS Arena: Benchmarking TTS Models in the Wild

Vote to help the community find the best available text-to-speech model!

[Vote](#) [Leaderboard](#) [About](#)

Vote

- Input text (English only) to synthesize audio (or press 🗒 for random text).
- Listen to the two audio clips, one after the other.
- Vote on which audio sounds more natural to you.
- Note: Model names are revealed after the vote is cast.*

Note: It may take up to 30 seconds to synthesize audio.

I want to eat barbecue sauce

Synthesize

0:02 0:02

◀️ 1x ▶️ ▶️

Your vote: VoiceCraft 2.0

0:02 0:02

◀️ 1x ▶️ ▶️

GPT-SoVITS

Next round

Citation

<https://huggingface.co/spaces/TTS-AGI/TTS-Arena>

washingtonpost.com/dc-md-va/2024/04/26/baltimore-ai-voice-audio-framing-principal/

The Washington Post
Democracy Dies in Darkness

jonathan.soma

D.C., MD. & VA. The District Maryland Virginia Crime & Public Safety Local Education Obituaries Transportation Capital Weather Gang

Baltimore principal's racist rant was an AI fake. His colleague was arrested.

By Paul Schwartzman and Pranshu Verma
April 26, 2024 at 1:43 p.m. EDT



MOST READ D.C., MD. & VA. >

- Perspective | A teen's body was found in Virginia. His family is devastated and angry.
- He was almost off probation in a murder case. Then a judge saw his videos.
- House Republicans look to

W Adversarial machine learning +

en.wikipedia.org/wiki/Adversarial_machine_learning

WIKIPEDIA The Free Encyclopedia

Adversarial machine learning

Article Talk

From Wikipedia, the free encyclopedia

Not to be confused with [Generative adversarial network](#).

Adversarial machine learning is the study of the attacks on [machine learning](#) algorithms, and of the defenses against such attacks.^[1] A survey from May 2020 exposes the fact that practitioners report a dire need for better protecting machine learning systems in industrial applications.^[2]

Most machine learning techniques are mostly designed to work on specific problem sets, under the assumption that the training and test data are generated from the same statistical distribution ([IID](#)). However, this assumption is often dangerously violated in practical high-stake applications, where users may intentionally supply fabricated data that violates the statistical assumption.

Most common attacks in adversarial machine learning include [evasion attacks](#),^[3] [data poisoning attacks](#),^[4] [Byzantine attacks](#)^[5] and [model extraction](#).^[6]

History [edit]

At the MIT Spam Conference in January 2004, [John Graham-Cumming](#) showed that a machine learning spam filter could be used to defeat another machine learning spam filter by automatically learning which words to add to a spam email to get the email classified as not spam.^[7]

In 2004, Nilesh Dalvi and others noted that [linear classifiers](#) used in [spam filters](#) could be defeated by simple "evasion attacks" as spammers inserted "good words" into their spam emails. (Around 2007, some spammers added random noise to fuzz words within "image spam" in order to defeat [OCR](#)-based filters.) In 2006, Marco Barreno and others published "Can Machine Learning Be Secure?", outlining a broad taxonomy of attacks. As late as 2013 many researchers continued to hope that non-linear classifiers (such as [support vector machines](#) and [neural networks](#)) might be robust to adversaries, until Battista Biggio and others demonstrated the first gradient-based attacks on such machine-learning models (2012^[8]–2013^[9]). In 2012, [deep neural networks](#) began to dominate computer vision problems;

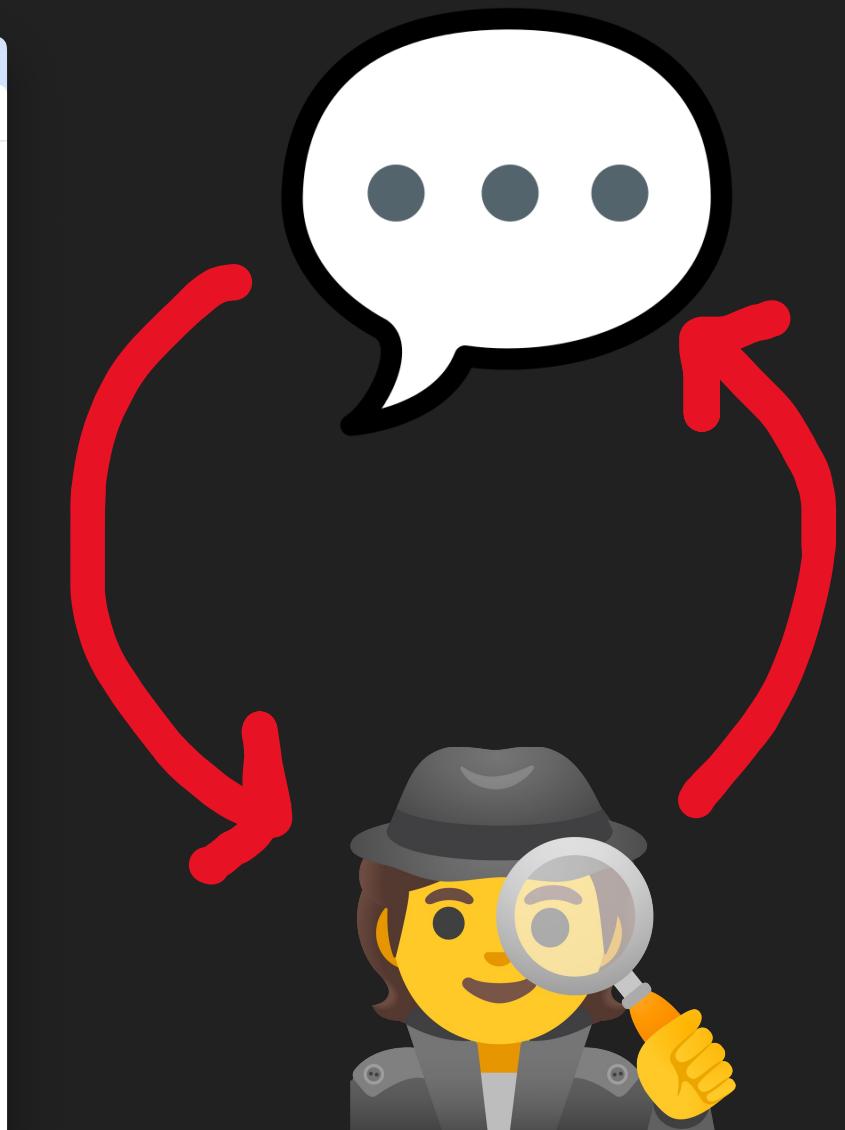
Read Edit View history Tools

9 languages

Part of a series on
Machine learning and data mining

- Paradigms [show]
- Problems [show]
- Supervised learning (classification • regression) [show]
- Clustering [show]
- Dimensionality reduction [show]
- Structured prediction [show]
- Anomaly detection [show]
- Artificial neural network [show]
- Reinforcement learning [show]
- Learning with humans [show]
- Model diagnostics [show]
- Mathematical foundations [show]
- Machine-learning venues [show]
- Related articles [show]

V T E



AI Voice Generator & Text to Speech

elevenlabs.io

Incognito

ElevenLabs

Generative Voice AI

Convert text to speech online for free with our AI voice generator. Create natural AI voices instantly in any language - perfect for video creators, developers, and businesses.

Get Started Free →

Click on a language to convert text to speech: English

Portuguese French German Japanese Arabic
Indonesian Italian Dutch Turkish Polish
Romanian Ukrainian Greek Czech Danish
Slovak Tamil

Eleven lets you voice any length of text automatically matching what is being said. The model works best on longer text sentences.

— Daniel ▾

Alice

Voice settings

102 / 5000

Generate speech

0 / 333

https://elevenlabs.io/

