

# Transformers.js

Run 🤖 Transformers in your browser!

<https://github.com/xenova/transformers.js>

npm v2.1.1

downloads 2.6k/week

license Apache-2.0



# Introduction

The what, how, and why of Transformers.js



# Library Overview



## ML + JS

Run ML models directly in the browser with JavaScript!



## Open source

Community-driven development on GitHub. New features added daily!

★ Stars 2.7k

🔗 Forks 126



## Easy to use

Add state-of-the-art ML to your web-app in just a few lines of code!

```
npm i @xenova/transformers
```



# What can it do?

## Text



Translation, summarization, text-generation, classification, NER, and much more!

## Vision



Classification, captioning, segmentation, and object detection.

## Audio



Automatic Speech Recognition (ASR) and audio classification.

## Multimodal



Zero-shot image classification

Transformers.js supports over 20 popular model architectures, including:

BERT, T5, GPT-2, BART

ViT, Vision Encoder Decoder, DETR

Whisper

CLIP

We have **over 100** *ready-to-use* models available on the Hugging Face Hub!



# How does it work?

1

**Convert your model to ONNX with 🤗 Optimum**

Supports PyTorch, TensorFlow, and JAX models.

2

**Write JavaScript code**

Get started with just a few lines!

```
import { pipeline } from '@xenova/transformers';  
  
let detector = await pipeline('object-detection');  
  
let predictions = await detector('cats.png');
```

3

**Run in the browser**

It's really that simple!



```
{  
  "boxes": [  
    [30.09, 68.47, 187.94, 118.53],  
    [329.51, 66.39, 370.07, 192.87],  
    [-2.80, 0.13, 636.90, 472.21],  
    [6.57, 53.89, 321.98, 468.69],  
    [331.87, 21.79, 648.55, 369.94]  
  ],  
  "classes": [75, 75, 63, 17, 17],  
  "scores": [0.998, 0.996, 0.995, 0.998, 0.999],  
  "labels": ["remote", "remote", "couch", "cat", "cat"]  
}
```

# Why was it created?



## Original reason

Browser extension for removing spam YouTube comments



## Current plan

To support all 🤖 Transformers models, tokenizers, processors, pipelines, and tasks.



## Ultimate goal

Help bridge the gap between web development and machine learning



2

# Applications

What are the use-cases?

# WebML environments

## Websites and PWAs

Reach + scale of the web.  
Use browser APIs.  
Zero install.

## Browser extensions

Enhancing the browsing experience with ML-powered extensions

## Server-side / Electron apps

JS/TS server-side  
Or desktop apps using Electron



**Whisper Web**  
ML-powered speech recognition directly in your browser

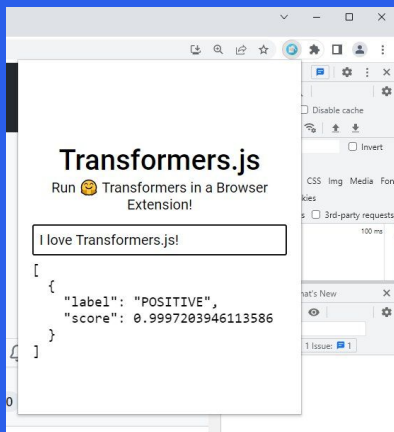
From URL From file

0:00 / 1:00

Transcribe Audio

- 00:50 So I planned things out and I decided.
- 00:52 I had to go something like this.
- 00:55 This is how the year we got.
- 00:57 So I'd start off light and I'd bump it up.

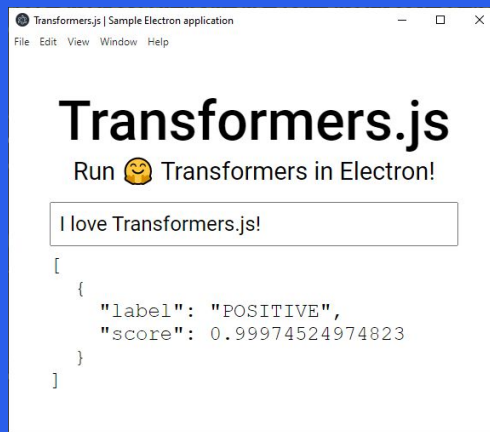
Export JSON



**Transformers.js**  
Run 🤖 Transformers in a Browser Extension!

I love Transformers.js!

```
{  
  "label": "POSITIVE",  
  "score": 0.9997203946113586  
}
```



**Transformers.js**  
Run 🤖 Transformers in Electron!

I love Transformers.js!

```
{  
  "label": "POSITIVE",  
  "score": 0.99974524974823  
}
```



# Feasible tasks

Text



Vision



Audio



Multimodal



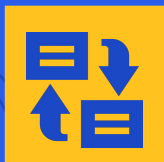
## Text Classification

Sentiment analysis, NER, etc.



## Code Completion

Constrained text-generation problems

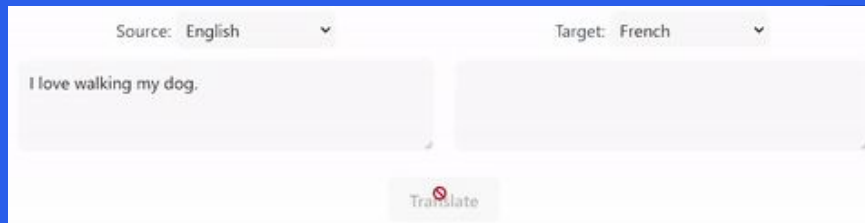
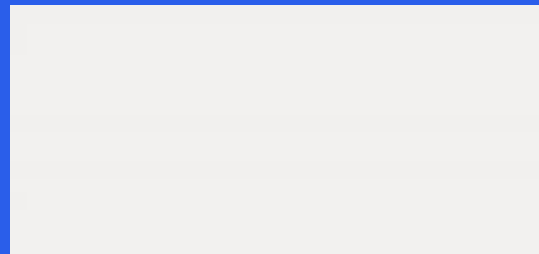


## Text-to-text

Translation, summarization, etc.

Hugging Face is a technology company that was founded in 2016 by Clément Delangue, Julien Chaumond, and Thomas Wolf. The company is headquartered in New York City, and is focused on developing natural language processing software and tools.

Hugging Face<sup>ORG</sup> is a technology company that was founded in 2016 by Clément Delangue<sup>PER</sup>, Julien Chaumond<sup>PER</sup>, and Thomas Wolf<sup>PER</sup>. The company is headquartered in New York City<sup>LOC</sup>, and is focused on developing natural language processing software and tools.



# Feasible tasks

Text



Vision



Audio



Multimodal



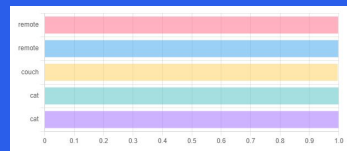
## Image Classification

Label images according to predefined classes



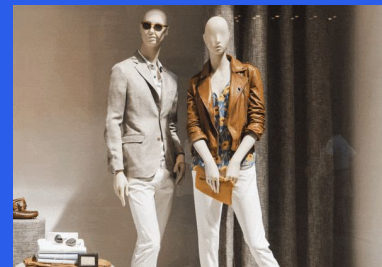
## Object Detection

Compute bounding boxes for objects



## Segmentation

Divide an image into meaningful parts



Segment Anything Model (Meta)

# Feasible **tasks**

Text



Vision



Audio



Multimodal



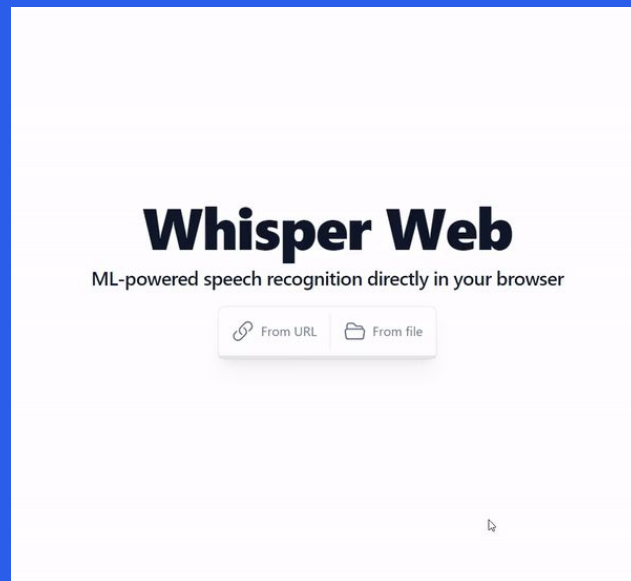
## Speech-to-Text

Automatic speech recognition



## Text-to-Speech

*\*Coming soon\**



<https://hf.co/spaces/Xenova/whisper-web>

# Feasible tasks

Text



Vision



Audio

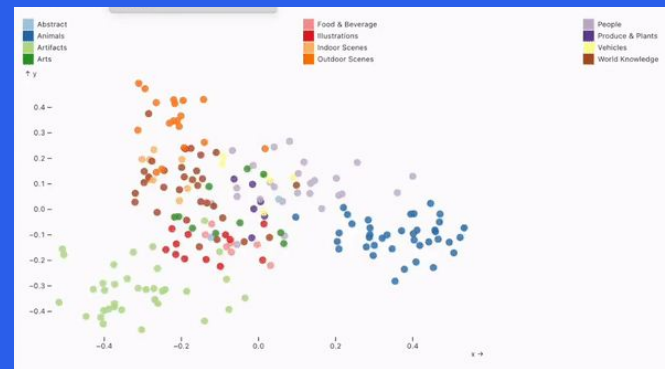


Multimodal



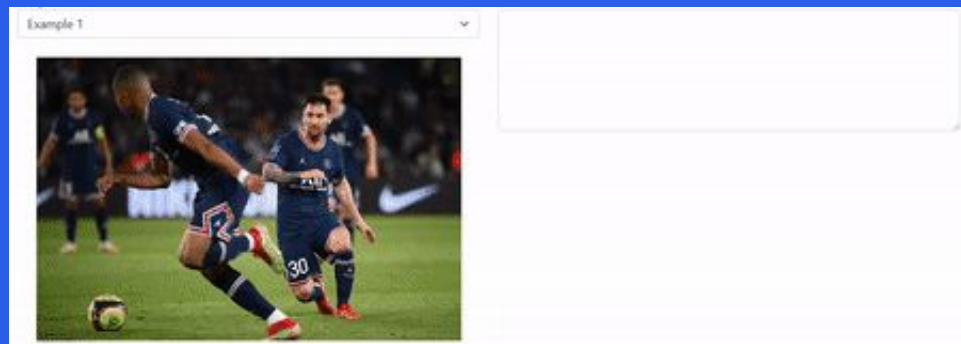
## Embeddings

Semantic search, clustering, data analysis



## Image-to-text

Adding captions to images





3

# Limitations

The good, the bad, and the ugly.

# What do we wish were **better**?



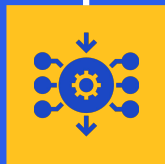
## Speed

Currently CPU-only  
(WebGPU on the way!)



## Memory

WASM models can't  
exceed 4GB in size.



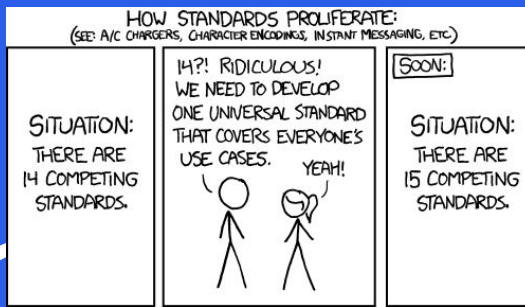
## Models

Standards, distribution  
and interoperability



## Browsers

Unified model  
caching, Tensor API





# Thanks!

Any questions?

[joshua@huggingface.co](mailto:joshua@huggingface.co)  
Hugging Face



CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Storyset** and **Chunte Lee**