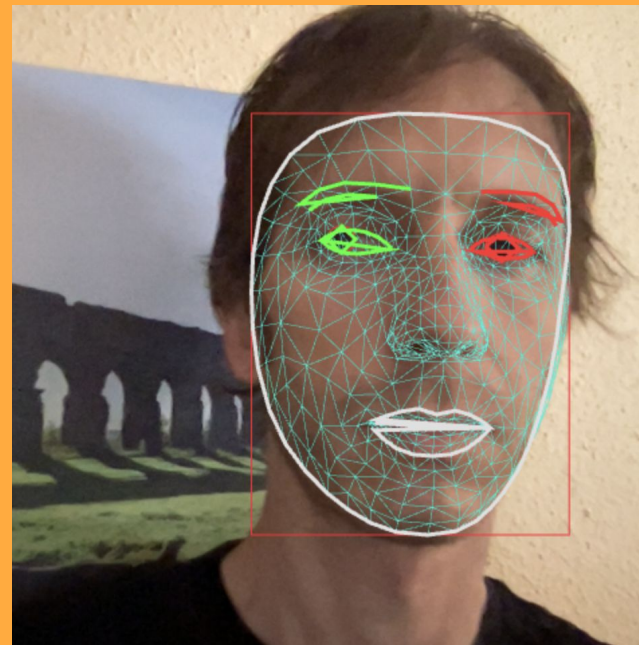


Inteligencia Ambiental

Máster Universitario en Inteligencia Artificial Aplicada

Transformers.js



Contenido

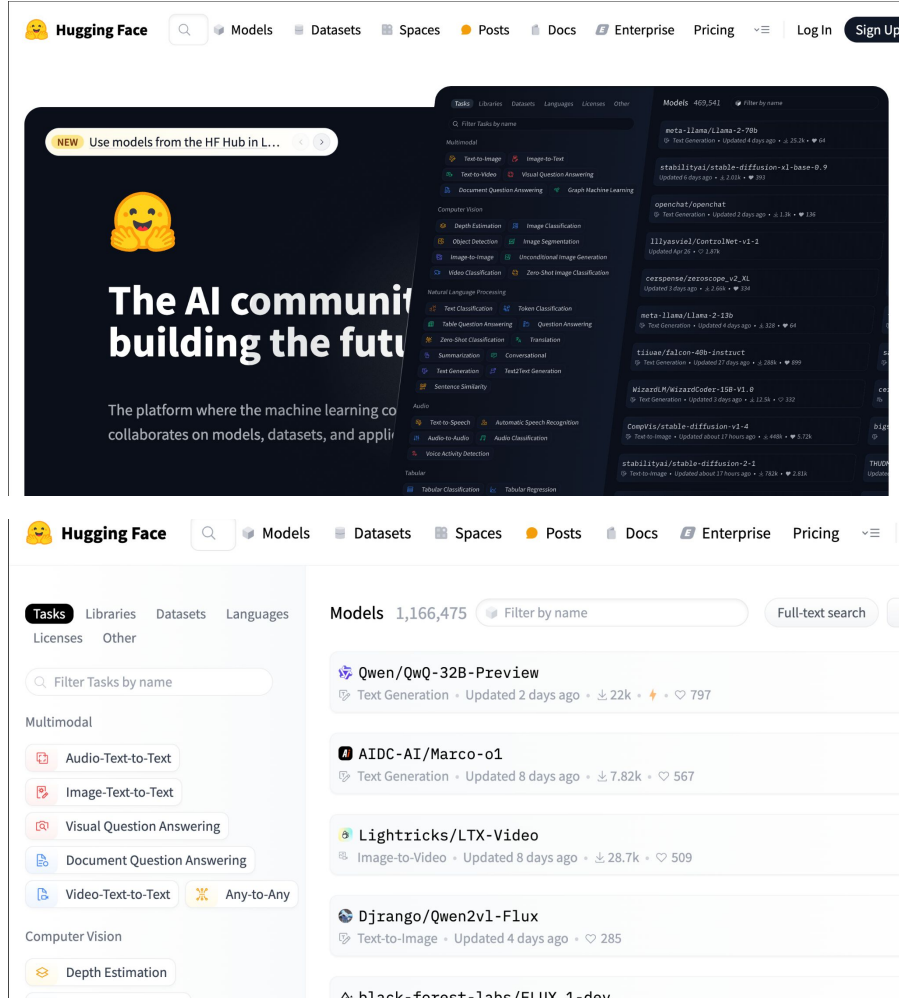
1. Transformers.js
2. Tercera práctica

1. Transformers.js

HuggingFace

<https://huggingface.co/>

- Plataforma de modelos y herramientas de AI especialmente enfocada en NLP
- Ecosistema de código abierto
 - Proporciona librerías populares como Transformers, Datasets y Tokenizers, que permiten a los desarrolladores entrenar, ajustar y usar modelos con facilidad
- Marketplace de modelos y comunidad activa
 - Repositorio en línea donde los usuarios pueden compartir y descargar modelos preentrenados, conjuntos de datos y herramientas personalizadas



Transformers.js

<https://huggingface.co/docs/transformers.js/index>

<https://github.com/huggingface/transformers.js/tree/v3>

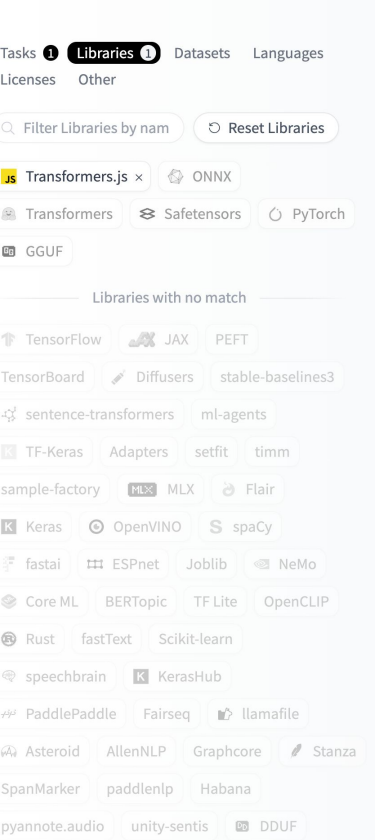
- Librería de código abierto para trabajar con modelos basados en Transformers directamente en el navegador **sin necesidad de servidores**
- Diseñada para ser funcionalmente equivalente a la biblioteca original en Python
- Usa el entorno de ejecución ONNX
- Proporciona acceso a modelos preentrenados y herramientas para entrenar, ajustar y desplegar modelos de NLP, visión por computadora y más
- API intuitiva para tareas comunes como clasificación, generación de texto, traducción, procesamiento de imágenes y más
- Permite desarrollar aplicaciones web interactivas

The screenshot shows the Hugging Face Transformers.js library interface. On the left, there's a sidebar with filters for Tasks, Libraries (selected), Datasets, Languages, Licenses, and Other. The main area displays a list of models under the 'Models' tab. The models listed are:

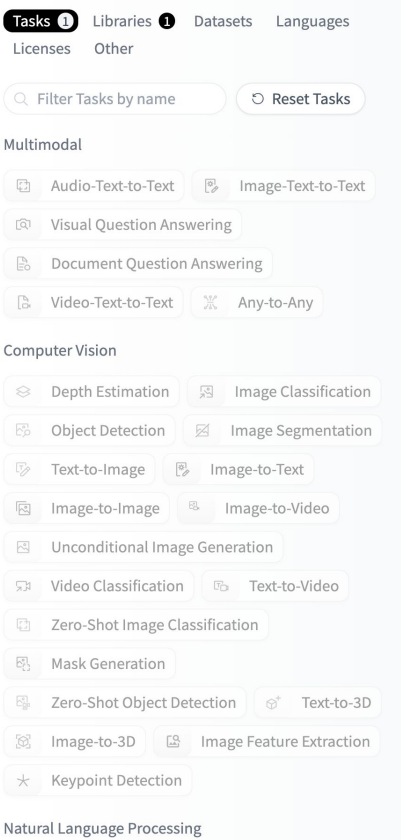
- jinaai/jina-clip-v2**: Feature Extraction, Updated 2 days ago, 6.27k, 91
- HuggingFaceTB/SmolLM2-1.7B-Instruct**: Text Generation, Updated 5 days ago, 92.5k, 394
- nomic-ai/nomic-embed-text-v1.5**: Sentence Similarity, Updated 13 days ago, 1.55M, 438
- HuggingFaceTB/SmolLM2-135M-Instruct**: Text Generation, Updated 6 days ago, 26k, 65
- onnx-community/Llama-3.2-1B-Instruct**: Text Generation, Updated Oct 8, 808, 20
- HuggingFaceTB/SmolLM2-360M-Instruct**: Text Generation, Updated 6 days ago, 39.7k, 58

Modelos

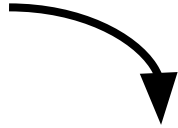
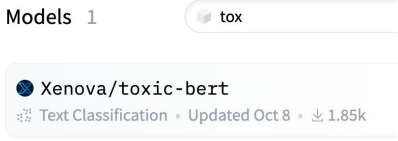
Filtramos por librería



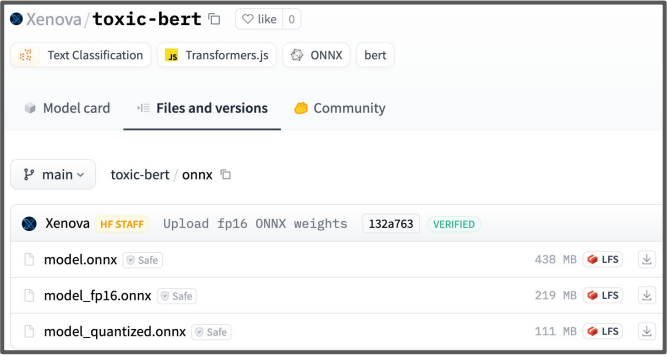
Filtramos por tarea



Filtrado adicional: p ej., nombre modelo



Comprobamos que exista **model.onnx**



Problemas

- Algunos de los modelos no se han actualizado a la nueva versión de la librería 3.1.0
- Para estos modelos es necesario utilizar la versión anterior del API 2.x.0

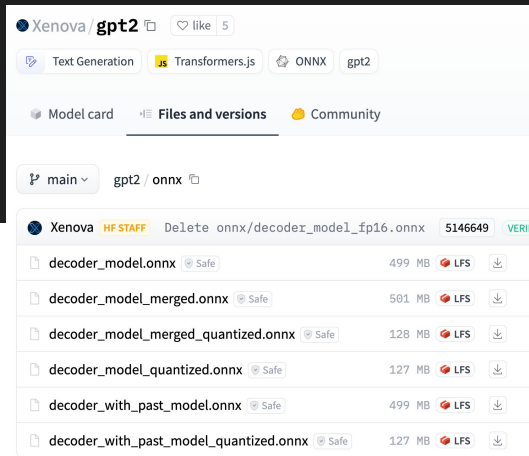
```
import {pipeline} from "https://cdn.jsdelivr.net/npm/@xenova/transformers";

// Create a text generation pipeline
const generator = await pipeline( "text-generation", "Xenova/gpt2");

// Define text
const message = "Once upon a time, "

// Generate a response
let output = await generator(message, {
  max_new_tokens: 128,
  temperature: 0.9,
  repetition_penalty: 2.0,
  no_repeat_ngram_size: 3
});

console.log(output[0].generated_text);
```



Xenova / gpt2			
Text Generation Transformers.js ONNX gpt2			
Model card Files and versions Community			
main gpt2 onnx			
Xenova	HF STAFF	Delete onnx/decoder_model_fp16.onnx	5146649
decoder_model.onnx	Safe	499 MB	LFS
decoder_model_merged.onnx	Safe	501 MB	LFS
decoder_model_merged_quantized.onnx	Safe	128 MB	LFS
decoder_model_quantized.onnx	Safe	127 MB	LFS
decoder_with_past_model.onnx	Safe	499 MB	LFS
decoder_with_past_model_quantized.onnx	Safe	127 MB	LFS

Tutorial 6: Generación de texto con transformers.js

<https://docs.google.com/document/d/13xR4MHBI3Q94TjsVSjXAJXtC6vzNSw24c35KKe-4XvM>

Aprender a:

- Cargar y utilizar un modelo de lenguaje como Qwen2.5-0.5B desde Transformers.js
- Gestionar la generación progresiva de texto en tiempo real
- Utilizar hilos de trabajo (workers) para procesamiento asíncronico
- Diseñar prompts efectivos para obtener respuestas contextuales

2. Tercera práctica

Tercera práctica: Asistente inteligente para el transporte público

<https://docs.google.com/document/d/1Kf4GIKQiFpzXZAKrks82M0vXG-HGL9z7cfVxraBe1Lc>

- Prototipo de asistente inteligente para el transporte público utilizando detección de eventos basada en audio para mejorar el entorno de los pasajeros
- Reconocimiento de tres sonidos utilizando aprendizaje por transferencia con YamNet
 - Sonido 1: Golpes de Asiento al Levantarse
 - Sonido 2: Sonidos de alarma
 - Sonido 3: Equipaje rodante en zonas no permitidas
- Utiliza transformers.js directamente en el navegador para interactuar con modelos generativos de lenguaje y generar sugerencias y respuestas contextualmente relevantes