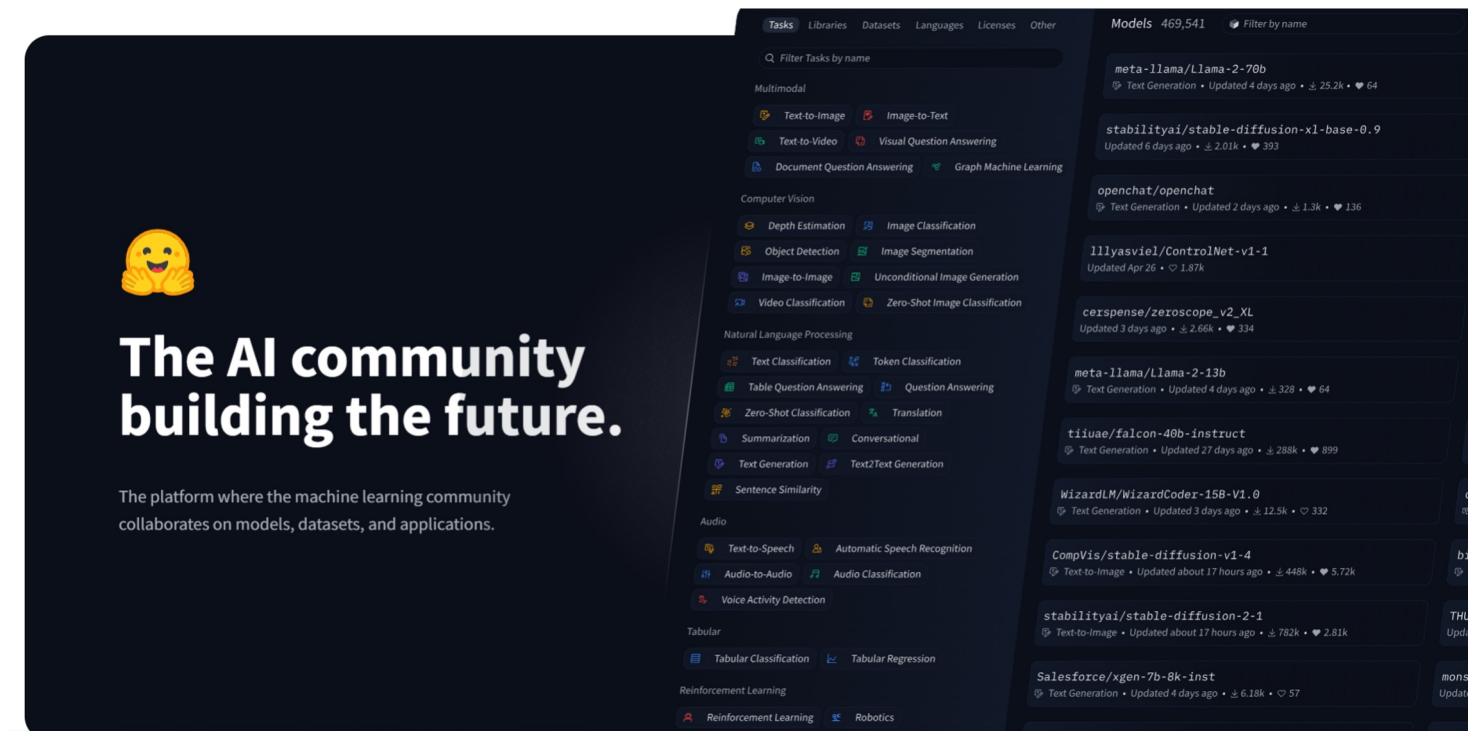


Huggingface

Credit to TA.Cheetah

What is Huggingface?

- HuggingFace is an AI community that promotes open source contributions. It is a hub of open source models for Natural Language Processing, computer vision, and other fields where AI plays its role. Even the tech giants like Google, Facebook, AWS, Microsoft, and others use the models, datasets, and libraries.



Datasets in Huggingface

The screenshot displays the Huggingface Datasets interface. On the left, there are navigation tabs for 'Tasks', 'Sizes', 'Sub-tasks', 'Languages', 'Licenses', and 'Other'. Below these, a search bar 'Filter Tasks by name' is present. The main content area is titled 'Datasets 56,783' and includes a 'Filter by name' input field. A 'new' button and a 'Full-text search' input are also visible. A 'Sort: Trending' dropdown is located at the top right. The datasets are listed in a grid, each with a repository name, a 'Preview' or 'Viewer' link, and statistics for updates, downloads, and likes. The datasets shown are:

- allenai/dolma (Preview • Updated 4 days ago • 49 • 179)
- garage-bAInd/Open-Platypus (Viewer • Updated 8 days ago • 4.09k • 125)
- BAAI/COIG-PC (Preview • Updated 12 days ago • 373 • 166)
- Open-Orca/OpenOrca (Viewer • Updated 3 days ago • 20.7k • 601)
- timdettmers/openassistant-guanaco (Viewer • Updated May 28 • 37.1k • 184)
- OpenAssistant/oasst1 (Viewer • Updated May 2 • 20.9k • 978)
- ehartford/dolphin (Preview • Updated 23 days ago • 2.27k • 152)
- fka/awesome-chatgpt-prompts (Viewer • Updated Mar 7 • 3.96k • 3.09k)
- PygmalionAI/PIPPA (Preview • Updated 9 days ago • 271 • 69)
- google/dreambooth (Viewer • Updated 8 days ago • 8 • 26)
- b-mc2/sql-create-context (Viewer • Updated Apr 21 • 1.38k • 68)
- databricks/databricks-dolly-15k (Viewer • Updated Jul 1 • 42.8k • 325)
- togethercomputer/RedPajama-Data-1T (Viewer • Updated Jul 1 • 17k • 828)
- mlabonne/guanaco-llama2-1k (Viewer • Updated 28 days ago • 11.3k • 34)

Ref: <https://huggingface.co/datasets?sort=trending>

Models in Huggingface















The screenshot displays the Huggingface Models interface. On the left, a sidebar lists tasks: Multimodal (Feature Extraction, Text-to-Image, Image-to-Text, Text-to-Video, Visual Question Answering, Document Question Answering, Graph Machine Learning) and Computer Vision (Depth Estimation, Image Classification, Object Detection, Image Segmentation, Image-to-Image, Unconditional Image Generation, Video Classification, Zero-Shot Image Classification). Below these are Natural Language Processing tasks (Text Classification, Token Classification, Table Question Answering, Question Answering, Zero-Shot Classification, Translation). The main area shows a list of models under the 'Models' tab, which is highlighted with a red box and shows 301,886 models. A search bar and a 'Filter by name' button are at the top. The models are sorted by 'Trending'. The first few models listed are: stabilityai/control-lora (Text-to-Image, Updated 4 days ago, 289 likes), meta-llama/Llama-2-7b (Text Generation, Updated Jul 20, 2k likes), Open-Orca/OpenOrca-Platypus2-13B (Text Generation, Updated 2 days ago, 15.3k likes, 137 likes), diffusers/controlnet-canny-sdxl-1.0 (Text-to-Image, Updated 10 days ago, 12.1k likes, 298 likes), facebook/seamless-m4t-large (Updated about 17 hours ago, 75 likes), stabilityai/stablecode-instruct-alpha-3b (Text Generation, Updated 15 days ago, 7.02k likes, 229 likes), TheBloke/Llama-2-7B-Chat-GGML (Text Generation, Updated 28 days ago, 8.42k likes, 369 likes), stabilityai/stable-diffusion-xl-base-1.0 (Text-to-Image, Updated 19 days ago, 856k likes, 2.11k likes), Deci/DeciCoder-1b (Text Generation, Updated about 17 hours ago, 2.59k likes, 141 likes), meta-llama/Llama-2-7b-chat-hf (Text Generation, Updated 14 days ago, 474k likes, 865 likes), meta-llama/Llama-2-70b-chat-hf (Text Generation, Updated 14 days ago, 174k likes, 1.1k likes), garage-bAInd/Platypus2-70B-instruct (Text Generation, Updated 3 days ago, 3.77k likes, 107 likes), runwayml/stable-diffusion-v1-5 (Text-to-Image, Updated Jul 5, 7.77M likes, 9.02k likes), and defog/sqlcoder (Text Generation, Updated 1 day ago, 430 likes, 56 likes).

Ref: <https://huggingface.co/models>

















Examples notebook

PyTorch Examples

Natural Language Processing

Notebook	Description	
Train your tokenizer	How to train and use your very own tokenizer	 Open in Colab  Open in Studio Lab
Train your language model	How to easily start using transformers	 Open in Colab  Open in Studio Lab
How to fine-tune a model on text classification	Show how to preprocess the data and fine-tune a pretrained model on any GLUE task.	 Open in Colab  Open in Studio Lab
How to fine-tune a model on language modeling	Show how to preprocess the data and fine-tune a pretrained model on a causal or masked LM task.	 Open in Colab  Open in Studio Lab
How to fine-tune a model on token classification	Show how to preprocess the data and fine-tune a pretrained model on a token classification task (NER, PoS).	 Open in Colab  Open in Studio Lab
How to fine-tune a model on question answering	Show how to preprocess the data and fine-tune a pretrained model on SQUAD.	 Open in Colab  Open in Studio Lab
How to fine-tune a model on multiple choice	Show how to preprocess the data and fine-tune a pretrained model on SWAG.	 Open in Colab  Open in Studio Lab

Computer Vision

Notebook	Description	
How to fine-tune a model on image classification (Torchvision)	Show how to preprocess the data using Torchvision and fine-tune any pretrained Vision model on Image Classification	 Open in Colab  Open in Studio Lab
How to fine-tune a model on image classification (Albumentations)	Show how to preprocess the data using Albumentations and fine-tune any pretrained Vision model on Image Classification	 Open in Colab  Open in Studio Lab
How to fine-tune a model on image classification (Kornia)	Show how to preprocess the data using Kornia and fine-tune any pretrained Vision model on Image Classification	 Open in Colab  Open in Studio Lab
How to perform zero-shot object detection with OWL-ViT	Show how to perform zero-shot object detection on images with text queries	 Open in Colab  Open in Studio Lab
How to fine-tune an image captioning model	Show how to fine-tune BLIP for image captioning on a custom dataset	 Open in Colab  Open in Studio Lab
How to build an image similarity system with Transformers	Show how to build an image similarity system	 Open in Colab  Open in Studio Lab
How to fine-tune a SegFormer model on semantic segmentation	Show how to preprocess the data and fine-tune a pretrained SegFormer model on Semantic Segmentation	 Open in Colab  Open in Studio Lab
How to fine-tune a VideoMAE model on video classification	Show how to preprocess the data and fine-tune a pretrained VideoMAE model on Video Classification	 Open in Colab  Open in Studio Lab

Fine-tune a pretrained model: <https://huggingface.co/docs/transformers/training>

Ref: <https://huggingface.co/docs/transformers/notebooks>