

# Using the Transformer

## Add-On: Which model to use?















### Learning goals




- Huggingface model hub
- Grasp the difficulty of model choice

# THE HUGGINGFACE HUB

## Tasks

-  Image Classification
-  Image Segmentation
-  Automatic Speech Recognition
-  Sentence Similarity
-  Question Answering
-  Zero-Shot Classification
-  Translation
-  Fill-Mask
-  Token Classification
-  Audio Classification
-  Summarization
-  + 23 Tasks









## Libraries

-  PyTorch
-  TensorFlow
-  JAX
- + 36




## Datasets

-  mozilla-foundation/common\_voice\_7\_0
-  squad
-  wikipedia
-  common\_voice
-  glue
-  emotion
-  xtreme
-  imdb
- + 328

## Languages

-  English
-  French
-  Spanish
-  German
-  Chinese
-  Portuguese
-  Japanese
-  Russian
- + 198

## Licenses

-  apache-2.0
-  mit
-  afl-3.0
- + 56

## Other

-  AutoTrain Compatible
-  Eval Results
-  Has a Space
-  Carbon Emissions

## Models 100,642

Filter by name

11 Sort: Most Downloads

### bert-base-uncased

Updated 28 days ago • 21.4M • 373

### xln-roberta-base

Updated 28 days ago • 13M • 123

### distilbert-base-uncased-finetuned-sst-2-english

Updated 9 days ago • 10.1M • 121

### distilbert-base-uncased

Updated 28 days ago • 8.22M • 105

### bert-base-multilingual-cased

Updated 27 days ago • 6.95M • 76

### Jean-Baptiste/camembert-ner

Updated Oct 13 • 5.84M • 48

### sentence-transformers/paraphrase-multilingual-Mi...

Updated Jun 15 • 3.7M • 75

### roberta-large

Updated Sep 29 • 2.57M • 69

### bert-base-chinese

Updated 27 days ago • 2.39M • 175

### cardiffnlp/twitter-roberta-base-sentiment

Updated Apr 6 • 2.17M • 106

### prajjwalibert-tiny

Updated Oct 27, 2021 • 15.8M • 19

### gpt2

Updated 21 days ago • 10.9M • 360

### openai/clip-vit-large-patch14

Updated Oct 4 • 9.71M • 107

### roberta-base

Updated Sep 29 • 6.99M • 92

### bert-base-cased

Updated 28 days ago • 6.91M • 63

### google/vit-base-patch32-224-in21k

Updated 6 days ago • 3.9M • 4

### sentence-transformers/all-MiniLM-L6-v2

Updated Nov 7 • 2.77M • 145

### albert-base-v2

Updated Aug 30, 2021 • 2.4M • 29

### microsoft/beit-base-patch16-224-pt22k-ft22k


Updated Jan 28 • 2.37M • 16

### ProsusAI/finbert


Updated Oct 2 • 2.14M • 121















# A SINGLE MODEL

Example: <https://huggingface.co/bert-base-uncased>

 **Hugging Face**

Models Datasets Spaces Docs Solutions Pricing Log In Sign Up

**bert-base-uncased**  Like 373

 Fill-Mask  PyTorch  TensorFlow  JAX  Rust  Safetensors  Transformers  bookcorpus  wikipedia  English  arxiv:1810.04805  bert  exbert  AutoTrain Compatible

License: apache-2.0

Model card Files and versions Community 11

[Edit model card](#)

### BERT base model (uncased)

Pretrained model on English language using a masked language modeling (MLM) objective. It was introduced in [this paper](#) and first released in [this repository](#). This model is uncased: it does not make a difference between english and English.


Disclaimer: The team releasing BERT did not write a model card for this model so this model card has been written by the Hugging Face team.

### Model description


BERT is a transformers model pretrained on a large corpus of English data in a self-supervised fashion. This means it was pretrained on the raw texts only, with no humans labeling them in any way (which is why it can use lots of publicly available data) with an automatic process to generate inputs and labels from those texts. More precisely, it was pretrained with two objectives:

- Masked language modeling (MLM): taking a sentence, the model randomly masks 15% of the words in the input then run the entire masked sentence through the model and has to predict the masked words. This is different from traditional recurrent neural networks (RNNs) that usually see the words one after the other, or from autoregressive models like GPT which internally masks the future tokens. It allows the model to learn a bidirectional representation of the sentence.

Downloads last month  
**21,362,332**



### Hosted inference API

 Fill-Mask Examples ▾

Mask token: [MASK]


Paris is the [MASK] of France.


[Compute](#)

The model is loaded and running on Intel Xeon 3rd Gen Scalable CPU

[JSON Output](#) [Maximize](#)

### Datasets used to train bert-base-uncased

 **wikipedia**  
Updated Nov 3 • 23.6k • 72

 **bookcorpus**  
[Preview](#) • Updated Nov 3 • 8.69k • 29

# MODEL RECYCLING

## A variety of models:

- Everyone can push his/her fine-tuned model to the hub
- Result: A multitude of different models available for each architecture

→ *Burning Question*: Which one should I use??

- The "original" pre-trained version
- A fine-tuned model on a similar task
- An ensemble?
- An averaged model?
- Something else?

# MODEL RECYCLING



🔍 Search Model Recycling

[Model Recycling on GitHub](#)

Home

Rankings



FAQ

Citation

Contact us

## Welcome to model-recycling page

Hardly anyone trains from scratch anymore, we all finetune over a pretrained model.

**Research** slowly reaches consensus that some finetuned models are better base models than the pretrained models themselves.

This site presents a dynamic view of the best models to choose for a given model size and architecture. We follow the findings and methodology from our **paper**: We download finetuned models found in HuggingFace per architecture and efficiently rank them over a representative task. We then evaluate the top ranked models by finetuning over a large set of 36 target tasks, and report the average performance of each base model.

Tested so far: 1986 (and counting)

Project website: <https://ibm.github.io/model-recycling/>

Paper: <https://arxiv.org/pdf/2211.00107.pdf>

# MODEL RECYCLING

## Best models per architectures

Pretrained	Best model	Avg.	Pretrained Avg.	Ranking
<a href="#">roberta-base</a>	<a href="#">ibm/CoLD-Fusion-itr13-seed2</a>	78.72	76.22	<a href="#">link</a>
<a href="#">bert-base-uncased</a>	<a href="#">ffgcc/InfoCSE-bert-base</a>	74.28	72.20	<a href="#">link</a>
<a href="#">bert-base-cased</a>	<a href="#">Dylan1999/bert-finetuned-squad-accelerate</a>	74.07	72.43	<a href="#">link</a>
<a href="#">t5-base</a>	<a href="#">adit94/nlpcharade</a>	78.23	75.45	<a href="#">link</a>
<a href="#">google/t5-v1_1-base</a>	<a href="#">anshoomehra/t5-v1-base-s2-auto-qgen</a>	74.27	68.82	<a href="#">link</a>