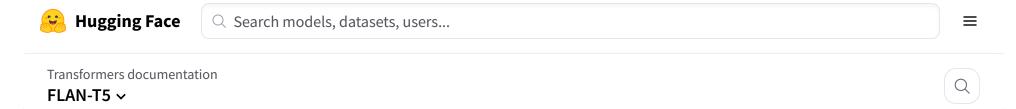
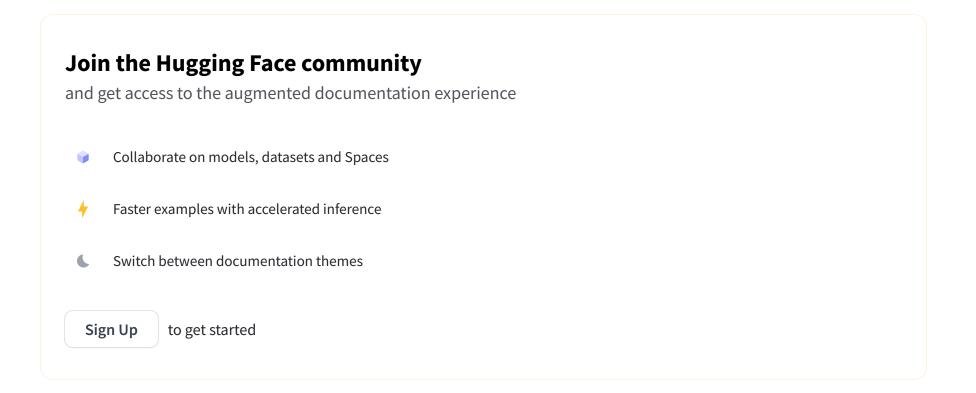
5/17/24, 3:15 PM FLAN-T5





## FLAN-T5

## Overview

5/17/24, 3:15 PM FLAN-T5

FLAN-T5 was released in the paper <u>Scaling Instruction-Finetuned Language Models</u> - it is an enhanced version of T5 that has been finetuned in a mixture of tasks.

One can directly use FLAN-T5 weights without finetuning the model:

```
>>> from transformers import AutoModelForSeq2SeqLM, AutoTokenizer
>>> model = AutoModelForSeq2SeqLM.from_pretrained("google/flan-t5-small")
>>> tokenizer = AutoTokenizer.from_pretrained("google/flan-t5-small")
>>> inputs = tokenizer("A step by step recipe to make bolognese pasta:", return_tensors="pt")
>>> outputs = model.generate(**inputs)
>>> print(tokenizer.batch_decode(outputs, skip_special_tokens=True))
['Pour a cup of bolognese into a large bowl and add the pasta']
```

FLAN-T5 includes the same improvements as T5 version 1.1 (see <u>here</u> for the full details of the model's improvements.)

Google has released the following variants:

- google/flan-t5-small
- google/flan-t5-base
- google/flan-t5-large
- google/flan-t5-xl
- google/flan-t5-xxl.

5/17/24, 3:15 PM FLAN-T5

The original checkpoints can be found here.

Refer to <u>T5's documentation page</u> for all API reference, code examples and notebooks. For more details regarding training and evaluation of the FLAN-T5, refer to the model card.

Update on GitHub

← FastSpeech2Conformer FLAN-UL2 →