🤗 **Hugging Face**   🔍 Search models, datasets, users...   ☰

Transformers documentation

FLAN-T5 ⌄   🔍

Join the Hugging Face community

and get access to the augmented documentation experience

📦 Collaborate on models, datasets and Spaces

⚡ Faster examples with accelerated inference

🌙 Switch between documentation themes

[**Sign Up**]   to get started

FLAN-T5

Overview

FLAN-T5 was released in the paper Scaling Instruction-Finetuned Language Models - 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:

```python
>>> 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 here 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.

The original checkpoints can be found here.

> Refer to T5's documentation page 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 →