```
!pip install transformers
```

```
Requirement already satisfied: transformers in /usr/local/lib/python3.11/dist-packages (4.52.4)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from transformers) (3.18.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in /usr/local/lib/python3.11/dist-packages (from transform
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-packages (from transformers) (2.0.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11/dist-packages (from transformers) (2024
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from transformers) (2.32.3)
Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python3.11/dist-packages (from transformers) (
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.5
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.11/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.11/dist-packages (from huggingface
Requirement already satisfied: hf-xet<2.0.0,>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->tra
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->transform
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->transform
```

Zero-Shot learning

```
from transformers import pipeline
import torch
pipeline = pipeline(task="text-generation", model="mistralai/Mistral-7B-Instruct-v0.1", torch_dtype=torch.bfloat16, device
prompt = """Classify the text into neutral, negative or positive.
Text: This movie is definitely one of my favorite movies of its kind. The interaction between respectable and morally stro
Sentiment:
"""

outputs = pipeline(prompt, max_new_tokens=10)
for output in outputs:
    print(f"Result: {output['generated_text']}")
```

```
config.json: 100%  571/571 [00:00<00:00, 19.2kB/s]
model.safetensors.index.json: 100%  25.1k/25.1k [00:00<00:00, 695kB/s]
Fetching 2 files: 100%  2/2 [01:43<00:00, 103.84s/it]
model-00002-of-00002.safetensors: 100%  4.54G/4.54G [00:56<00:00, 101MB/s]
model-00001-of-00002.safetensors: 100%  9.94G/9.94G [01:43<00:00, 195MB/s]
Loading checkpoint shards: 100%  2/2 [00:59<00:00, 27.18s/it]
generation_config.json: 100%  116/116 [00:00<00:00, 6.72kB/s]
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
  warnings.warn(
WARNING:accelerate.big_modeling:Some parameters are on the meta device because they were offloaded to the cpu.
tokenizer_config.json: 100%  2.10k/2.10k [00:00<00:00, 105kB/s]
tokenizer.model: 100%  493k/493k [00:00<00:00, 11.9MB/s]
tokenizer.json: 100%  1.80M/1.80M [00:00<00:00, 7.64MB/s]
special_tokens_map.json: 100%  414/414 [00:00<00:00, 39.5kB/s]
Device set to use cuda:0
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Result: Classify the text into neutral, negative or positive.
Text: This movie is definitely one of my favorite movies of its kind. The interaction between respectable and morally
Sentiment:
Text: I hate this movie. The acting was
```

One-Shot learning

```
from transformers import pipeline
import torch

pipeline = pipeline(task="text-generation", model="mistralai/Mistral-7B-Instruct-v0.1", torch_dtype=torch.bfloat16, device
prompt = """Text: The first human went into space and orbited the Earth on April 12, 1961.
Date: 04/12/1961
Text: The first-ever televised presidential debate in the United States took place on September 28, 1960, between presider
Date:"""

outputs = pipeline(prompt, max_new_tokens=12, do_sample=True, top_k=10)
for output in outputs:
    print(f"Result: {output['generated_text']}")
```

```
Loading checkpoint shards: 100%                                    2/2 [00:07<00:00,  7.86s/it]
WARNING:accelerate.big_modeling:Some parameters are on the meta device because they were offloaded to the disk and cpu
Device set to use cuda:0
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Result: Text: The first human went into space and orbited the Earth on April 12, 1961.
Date: 04/12/1961
Text: The first-ever televised presidential debate in the United States took place on September 28, 1960, between pres
Date: 09/28/1960
```

Few-Shot learning

```
from transformers import pipeline
import torch

pipeline = pipeline(task="text-generation", model="mistralai/Mistral-7B-Instruct-v0.1", torch_dtype=torch.bfloat16, device
prompt = """Text: The first human went into space and orbited the Earth on April 12, 1961.
Date: 04/12/1961
Text: The first IPL was at Kolakatha on April 18, 2008
Date: 18/04/2008:
Text: The kerala state formed on 1 November 1956
Date:
"""

outputs = pipeline(prompt, max_new_tokens=12, do_sample=True, top_k=10)
for output in outputs:
    print(f"Result: {output['generated_text']}")
```

```
Loading checkpoint shards: 100%                                    2/2 [00:01<00:00,  1.62s/it]
WARNING:accelerate.big_modeling:Some parameters are on the meta device because they were offloaded to the disk and cpu
Device set to use cuda:0
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Result: Text: The first human went into space and orbited the Earth on April 12, 1961.
Date: 04/12/1961
Text: The first IPL was at Kolakatha on April 18, 2008
Date: 18/04/2008:
Text: The kerala state formed on 1 November 1956
Date:
Text: The first human went into space and orbited the
```

Chat Based Prompting

```
from transformers import pipeline
import torch

pipeline = pipeline(task="text-generation", model="mistralai/Mistral-7B-Instruct-v0.1", torch_dtype=torch.bfloat16, device
```

```
messages = [
    {"role": "user", "content": "Text: The first human went into space and orbited the Earth on April 12, 1961."},
    {"role": "assistant", "content": "Date: 04/12/1961"},
    {"role": "user", "content": "Text: The first-ever televised presidential debate in the United States took place on Sep
]

prompt = pipeline.tokenizer.apply_chat_template(messages, tokenize=False, add_generation_prompt=True)

outputs = pipeline(prompt, max_new_tokens=12, do_sample=True, top_k=10)

for output in outputs:
    print(f"Result: {output['generated_text']}")
```

```
Loading checkpoint shards: 100%                                    2/2 [00:00<00:00,  3.72it/s]
WARNING:accelerate.big_modeling:Some parameters are on the meta device because they were offloaded to the disk and cpu
Device set to use cpu
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Result: <s> [INST] Text: The first human went into space and orbited the Earth on April 12, 1961. [/INST] Date: 04/12/
```

Chain-of-thought

```
from transformers import pipeline
import torch

pipeline = pipeline(task="text-generation", model="mistralai/Mistral-7B-Instruct-v0.1", torch_dtype=torch.bfloat16, device
prompt = """Let's go through this step-by-step:
1. You start with 15 muffins.
2. You eat 2 muffins, leaving you with 13 muffins.
3. You give 5 muffins to your neighbor, leaving you with 8 muffins.
4. Your partner buys 6 more muffins, bringing the total number of muffins to 14.
5. Your partner eats 2 muffins, leaving you with 12 muffins.
If you eat 6 muffins, how many are left?"""

outputs = pipeline(prompt, max_new_tokens=20, do_sample=True, top_k=10)
for output in outputs:
    print(f"Result: {output['generated_text']}")
```

```
Loading checkpoint shards: 100%                                    2/2 [00:00<00:00,  4.00it/s]
WARNING:accelerate.big_modeling:Some parameters are on the meta device because they were offloaded to the disk and cpu
Device set to use cpu
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Result: Let's go through this step-by-step:
1. You start with 15 muffins.
2. You eat 2 muffins, leaving you with 13 muffins.
3. You give 5 muffins to your neighbor, leaving you with 8 muffins.
4. Your partner buys 6 more muffins, bringing the total number of muffins to 14.
5. Your partner eats 2 muffins, leaving you with 12 muffins.
If you eat 6 muffins, how many are left?
There are 6 muffins left.
```