```
from transformers import MBartForConditionalGeneration, MBart50TokenizerFast

model_checkpoint = "aryaumesh/english-to-telugu"

tokenizer = MBart50TokenizerFast.from_pretrained(model_checkpoint)

model = MBartForConditionalGeneration.from_pretrained(model_checkpoint)

text = "Hello, How are you?" # Sentence to translate

inputs = tokenizer(text, return_tensors="pt")

outputs = model.generate(**inputs)

print(tokenizer.decode(outputs[0], skip_special_tokens=True))
```

```
హలో, ఎలా ఉన్నావు?
```

```
from transformers import pipeline

# Load a translation pipeline for English to French (en→fr)

translator = pipeline("translation_en_to_fr")

# Input English text

text = "Hugging Face makes working with AI easy and fun."

# Translate text

result = translator(text)

print("Translated Text:", result[0]['translation_text'])
```

```
No model was supplied, defaulted to google-t5/t5-base and revision a9723ea (https://huggingface.co/google-t5/t5-base).
Using a pipeline without specifying a model name and revision in production is not recommended.
Device set to use cpu
Translated Text: Hugging Face rend le travail avec AI facile et amusant.
```

```
!pip install datasets
```

```
Collecting datasets
  Downloading datasets-3.3.2-py3-none-any.whl.metadata (19 kB)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from datasets) (3.17.0)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-packages (from datasets) (1.26.4)
Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.11/dist-packages (from datasets) (18.1.0)
Collecting dill<0.3.9,>=0.3.0 (from datasets)
  Downloading dill-0.3.8-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (from datasets) (2.2.2)
Requirement already satisfied: requests>=2.32.2 in /usr/local/lib/python3.11/dist-packages (from datasets) (2.32.3)
Requirement already satisfied: tqdm>=4.66.3 in /usr/local/lib/python3.11/dist-packages (from datasets) (4.67.1)
Collecting xxhash (from datasets)
  Downloading xxhash-3.5.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Collecting multiprocess<0.70.17 (from datasets)
  Downloading multiprocess-0.70.16-py311-none-any.whl.metadata (7.2 kB)
Requirement already satisfied: fsspec<=2024.12.0,>=2023.1.0 in /usr/local/lib/python3.11/dist-packages (from fsspec[http]<=2024.12.0,>=2
Requirement already satisfied: aiohttp in /usr/local/lib/python3.11/dist-packages (from datasets) (3.11.13)
Requirement already satisfied: huggingface-hub>=0.24.0 in /usr/local/lib/python3.11/dist-packages (from datasets) (0.28.1)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-packages (from datasets) (24.2)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from datasets) (6.0.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (2.4.6)
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (25.1.0)
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (1.5.0)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (6.1.0)
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (0.3.0)
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (1.18.3)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.24.0->data
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests>=2.32.2->datasets) (3
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests>=2.32.2->datasets) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests>=2.32.2->datasets) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests>=2.32.2->datasets) (2025.1.3
```

```
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2025.1)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2025.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas->datasets) (1.17
Downloading datasets-3.3.2-py3-none-any.whl (485 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 485.4/485.4 kB 15.8 MB/s eta 0:00:00
Downloading dill-0.3.8-py3-none-any.whl (116 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 116.3/116.3 kB 8.5 MB/s eta 0:00:00
Downloading multiprocess-0.70.16-py311-none-any.whl (143 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 143.5/143.5 kB 5.9 MB/s eta 0:00:00
Downloading xxhash-3.5.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 194.8/194.8 kB 10.4 MB/s eta 0:00:00
Installing collected packages: xxhash, dill, multiprocess, datasets
Successfully installed datasets-3.3.2 dill-0.3.8 multiprocess-0.70.16 xxhash-3.5.0
```

```
import re

from datasets import Dataset

# Sample raw text with noise

raw_text = """

Hugging Face!!! specializes in NLP... It's a great platform for working with pre-trained models.

Visit: https://huggingface.co for more info.

Contact us at support@huggingface.co.

"""
# Function to clean the text

def clean_text(text):

    text = text.lower()  # Convert to lowercase

    text = re.sub(r'https?://\S+', '', text)  # Remove URLs

    text = re.sub(r'\S+@\S+', '', text)  # Remove emails

    text = re.sub(r'[^a-zA-Z0-9\s]', '', text)  # Remove special characters

    text = re.sub(r'\s+', ' ', text).strip()  # Remove extra whitespaces

    return text

# Create a dataset

dataset = Dataset.from_dict({"text": [raw_text]})

cleaned_dataset = dataset.map(lambda x: {"text": clean_text(x["text"])})
# Display the cleaned text

print("Cleaned Text:", cleaned_dataset[0]['text'])
```

```
Map: 100% 1/1 [00:00<00:00, 57.12 examples/s]
```

```
from transformers import pipeline
sentiment_analyzer = pipeline("sentiment-analysis")
text = "I love working with Hugging Face! It's amazing"
result = sentiment_analyzer(text)
print("Sentiment Analysis Result:",result)
```

```
No model was supplied, defaulted to distilbert/distilbert-base-uncased-finetuned-sst-2-english and revision 714eb0f (https://huggingfac
Using a pipeline without specifying a model name and revision in production is not recommended.
config.json: 100%  629/629 [00:00<00:00, 27.0kB/s]
model.safetensors: 100%  268M/268M [00:02<00:00, 130MB/s]
tokenizer_config.json: 100%  48.0/48.0 [00:00<00:00, 965B/s]
vocab.txt: 100%  232k/232k [00:00<00:00, 2.53MB/s]
Device set to use cpu
Sentiment Analysis Result: [{'label': 'POSITIVE', 'score': 0.9998880624771118}]
```

```
from datasets import load_dataset
from transformers import pipeline
from huggingface_hub import login

# Log in to Hugging Face with your access token
login(token="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")

# Load Amazon Food Reviews dataset (small subset for efficiency)
dataset = load_dataset("amazon_polarity", split="test[ :100]")

# Preview dataset
print(dataset)
print(dataset[0])
```

```
README.md: 100%  6.81k/6.81k [00:00<00:00, 95.9kB/s]
train-00000-of-00004.parquet: 100%  260M/260M [00:02<00:00, 81.0MB/s]
train-00001-of-00004.parquet: 100%  258M/258M [00:03<00:00, 141MB/s]
train-00002-of-00004.parquet: 100%  255M/255M [00:02<00:00, 77.8MB/s]
train-00003-of-00004.parquet: 100%  254M/254M [00:01<00:00, 194MB/s]
test-00000-of-00001.parquet: 100%  117M/117M [00:00<00:00, 197MB/s]
Generating train split: 100%  3600000/3600000 [00:19<00:00, 430447.44 examples/s]
Generating test split: 100%  400000/400000 [00:01<00:00, 290064.90 examples/s]
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-11-3df3cd3738a1> in <cell line: 0>()
      7
      8 # Load Amazon Food Reviews dataset (small subset for efficiency)
----> 9 dataset = load_dataset("amazon_polarity", split="test[ :100]")
     10
     11 # Preview dataset

 9 frames
/usr/local/lib/python3.11/dist-packages/datasets/arrow_reader.py in _str_to_read_instruction(spec)
    399     res = _SUB_SPEC_RE.match(spec)
    400     if not res:
--> 401         raise ValueError(f"Unrecognized instruction format: {spec}")
    402     unit = "%" if res.group("from_pct") or res.group("to_pct") else "abs"
    403     return ReadInstruction(

ValueError: Unrecognized instruction format: test[ :100]
```

Next steps: Explain error

```
from datasets import load_dataset
from transformers import pipeline
from huggingface_hub import login

# Log in to Hugging Face with your access token
login(token="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")

# Load the full "test" split of the Amazon Food Reviews dataset
dataset = load_dataset("amazon_polarity", split="test")

# Select the first 100 examples using slicing
dataset = dataset.select(range(100))

# Preview dataset
```

```
print(dataset)
print(dataset[0])
```

```
Dataset({
    features: ['label', 'title', 'content'],
    num_rows: 100
})
{'label': 1, 'title': 'Great CD', 'content': 'My lovely Pat has one of the GREAT voices of her generation. I have listened to this CD fo
```

Start coding or generate with AI.