```
!pip install datasets evaluate transformers[sentencepiece]
```

```
Requirement already satisfied: datasets in /usr/local/lib/python3.11/dist-packages (4.0.0)
Collecting evaluate
  Downloading evaluate-0.4.5-py3-none-any.whl.metadata (9.5 kB)
Requirement already satisfied: transformers[sentencepiece] in /usr/local/lib/python3.11/dist-packages (4.55.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from datasets) (3.18.0)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-packages (from datasets) (2.0.2)
Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.11/dist-packages (from datasets) (18.1.0)
Requirement already satisfied: dill<0.3.9,>=0.3.0 in /usr/local/lib/python3.11/dist-packages (from datasets) (0.3.8)
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)
Requirement already satisfied: xxhash in /usr/local/lib/python3.11/dist-packages (from datasets) (3.5.0)
Requirement already satisfied: multiprocess<0.70.17 in /usr/local/lib/python3.11/dist-packages (from datasets) (0.70.16)
Requirement already satisfied: fsspec<=2025.3.0,>=2023.1.0 in /usr/local/lib/python3.11/dist-packages (from fsspec[http]<=2025.3.0,>=202
Requirement already satisfied: huggingface-hub>=0.24.0 in /usr/local/lib/python3.11/dist-packages (from datasets) (0.34.3)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-packages (from datasets) (25.0)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from datasets) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11/dist-packages (from transformers[sentencepiece]) (2024.11.
Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python3.11/dist-packages (from transformers[sentencepiece]) (0.2
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.11/dist-packages (from transformers[sentencepiece]) (0.6.1)
Requirement already satisfied: sentencepiece!=0.1.92,>=0.1.91 in /usr/local/lib/python3.11/dist-packages (from transformers[sentencepiec
Requirement already satisfied: protobuf in /usr/local/lib/python3.11/dist-packages (from transformers[sentencepiece]) (5.29.5)
Requirement already satisfied: aiohttp!=4.0.0a0,!=4.0.0a1 in /usr/local/lib/python3.11/dist-packages (from fsspec[http]<=2025.3.0,>=2023
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: hf-xet<2.0.0,>=1.1.3 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.24.0->datasets)
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) (2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests>=2.32.2->datasets) (2025.8.3
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2025.2)
Requirement already satisfied: aiohappyeyeballs>=2.5.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fssp
Requirement already satisfied: aiosignal>=1.4.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]<=
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[htt
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[h
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[htt
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 evaluate-0.4.5-py3-none-any.whl (84 kB)
                                                    84.1/84.1 kB 4.0 MB/s eta 0:00:00
Installing collected packages: evaluate
Successfully installed evaluate-0.4.5
```

```
from transformers import pipeline
classifier=pipeline("sentiment-analysis")
classifier("I've been adoring you man")
```

```
No model was supplied, defaulted to distilbert/distilbert-base-uncased-finetuned-sst-2-english and revision 714eb0f (https://huggingface
Using a pipeline without specifying a model name and revision in production is not recommended.
/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), set it as secre
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(
config.json: 100%                                          629/629 [00:00<00:00, 41.9kB/s]
model.safetensors: 100%                                    268M/268M [00:03<00:00, 107MB/s]
tokenizer_config.json: 100%                                48.0/48.0 [00:00<00:00, 4.36kB/s]
vocab.txt:       232k/? [00:00<00:00, 9.26MB/s]
Device set to use cuda:0
[{'label': 'POSITIVE', 'score': 0.9997172951698303}]
```

```
classifier("just look at you, this dressing makes you look like bad")
```

```
[{'label': 'NEGATIVE', 'score': 0.9996805191040039}]
```

```
classifier=pipeline("zero-shot-classification")
classifier(
    "This is a course about the Transformers library",
    candidate_labels=["education", "politics", "business"],
```

```
)
```

No model was supplied, defaulted to facebook/bart-large-mnli and revision d7645e1 (https://huggingface.co/facebook/bart-large-mnli).
Using a pipeline without specifying a model name and revision in production is not recommended.

config.json: 1.15k/? [00:00<00:00, 95.3kB/s]

model.safetensors: 100%   1.63G/1.63G [00:23<00:00, 146MB/s]

tokenizer_config.json: 100%   26.0/26.0 [00:00<00:00, 950B/s]

vocab.json: 899k/? [00:00<00:00, 20.8MB/s]

merges.txt: 456k/? [00:00<00:00, 15.0MB/s]

tokenizer.json: 1.36M/? [00:00<00:00, 20.5MB/s]

```
Device set to use cuda:0
{'sequence': 'This is a course about the Transformers library',
 'labels': ['education', 'business', 'politics'],
 'scores': [0.844595193862915, 0.11197695881128311, 0.04342786595225334]}
```

```
pipeline.__doc__
```

'\n   Utility factory method to build a [`Pipeline`].\n\n   A pipeline consists of:\n\n   - One or more components for pre-proce
ssing model inputs, such as a tokenizer,\n   image_processor, [feature_extractor](feature_extracto
r), or [processor](processors).\n   - A model that generates predictions from the inputs.\n   - Optional post-proces
sing steps to refine the model\'s output, which can also be handled by processors.\n   <Tip>\n   While there are such optional argu
ments as `tokenizer`, `feature_extractor`, `image_processor`, and `processor`,\n   they shouldn\'t be specified all at once. If these
components are not provided, `pipeline` will try to load\n   required ones automatically. In case you want to provide these components

```
from transformers import pipeline
generator=pipeline("text-generation")
generator("In this course, we will teach you how to")
```

No model was supplied, defaulted to openai-community/gpt2 and revision 607a30d (https://huggingface.co/openai-community/gpt2).
Using a pipeline without specifying a model name and revision in production is not recommended.

config.json: 100%   665/665 [00:00<00:00, 37.1kB/s]

model.safetensors: 100%   548M/548M [00:10<00:00, 69.1MB/s]

generation_config.json: 100%   124/124 [00:00<00:00, 2.99kB/s]

tokenizer_config.json: 100%   26.0/26.0 [00:00<00:00, 874B/s]

vocab.json: 1.04M/? [00:00<00:00, 34.7MB/s]

merges.txt: 456k/? [00:00<00:00, 13.0MB/s]

tokenizer.json: 1.36M/? [00:00<00:00, 18.0MB/s]

```
Device set to use cuda:0
Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.
[{'generated_text': "In this course, we will teach you how to handle and use common web frameworks in a user-friendly way. We will also
explain the basic concepts of a web framework and how to set up an application in a way that will be easy to use and maintain. We will
also show you how to use a web application's built-in database to help you manage your data and your data's data integrity.\n\nCourse
Overview:\n\nThis course will be taught as part of a comprehensive training in the most important topics of web engineering. Alongside
the web frameworks we will be covering:\n\nA web design framework\n\nA data flow model\n\nA data structure model\n\nAn object-oriented
programming language\n\nA database\n\nA data structure framework\n\nA data-based programming language\n\nA data model\n\nA data-driven
web framework\n\nA data integrity framework\n\nCourse Requirements:\n\nA bachelor's degree in web engineering/web design, computer
science, or equivalent, or equivalent in a language other than English or Spanish, and an understanding of web development and data
systems.\n\nThis course requires the following courses:\n\nA bachelor's degree in web design, data science/data analysis, or
```

```
from transformers import pipeline
generator=pipeline("text-generation",model="distilgpt2")
generator("In this course, we will teach you how to",max_length=30,num_return_sequences=2)
```

config.json: 100%   762/762 [00:00<00:00, 85.2kB/s]

model.safetensors: 100%   353M/353M [00:12<00:00, 40.0MB/s]

generation_config.json: 100%   124/124 [00:00<00:00, 11.5kB/s]

tokenizer_config.json: 100%   26.0/26.0 [00:00<00:00, 2.96kB/s]

vocab.json: 100%   1.04M/1.04M [00:00<00:00, 40.3MB/s]

merges.txt: 100%   456k/456k [00:00<00:00, 34.8MB/s]

tokenizer.json: 100%   1.36M/1.36M [00:00<00:00, 36.0MB/s]

```
Device set to use cuda:0
Truncation was not explicitly activated but `max_length` is provided a specific value, please use `truncation=True` to explicitly trunca
Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.
Both `max_new_tokens` (=256) and `max_length`(=30) seem to have been set. `max_new_tokens` will take precedence. Please refer to the doc
[{'generated_text': 'In this course, we will teach you how to write a good and consistent writing experience in your project.'},
 {'generated_text': 'In this course, we will teach you how to read the entire book.\n\n\nStep 1.\n1. Learn how to read the entire
book.\nStep 2.\nStep 3.\nStep 4.\nStep 5.\nStep 6.\nStep 7.\nStep 8.\nStep 9.\nStep 10.\nStep 11.\nStep 12.\nStep 13.\nStep 14.\nStep
15.\nStep 16.\nStep 17.\nStep 18.\nStep 19.\nStep 20.\nStep 21.\nStep 22.\nStep 23.\nStep 24.\nStep 25.\nStep 26.\nStep 27.\nStep
28.\nStep 29.\nStep 30.\nStep
31.\n1.\n1.\n2.\n3.\n4.\n5.\n6.\n7.\n8.\n9.\n10.\n11.\n12.\n13.\n14.\n15.\n16.\n17.\n18.\n19.\n20.\n21.\n22.\n23.\n24.\n25.\n26.\n27.\n2
```

Start coding or generate with AI.

```python
from transformers import pipeline
```

```python
unmasker=pipeline("fill-mask",model="bert-large-uncased")
```

config.json: 100%   571/571 [00:00<00:00, 60.1kB/s]

model.safetensors: 100%   1.34G/1.34G [01:50<00:00, 48.8MB/s]

```
Some weights of the model checkpoint at bert-large-uncased were not used when initializing BertForMaskedLM: ['bert.pooler.dense.bias', '
- This IS expected if you are initializing BertForMaskedLM from the checkpoint of a model trained on another task or with another archit
- This IS NOT expected if you are initializing BertForMaskedLM from the checkpoint of a model that you expect to be exactly identical (i
```

tokenizer_config.json: 100%   48.0/48.0 [00:00<00:00, 2.08kB/s]

vocab.txt: 100%   232k/232k [00:00<00:00, 6.82MB/s]

tokenizer.json: 100%   466k/466k [00:00<00:00, 12.2MB/s]

```
Device set to use cuda:0
```

```python
unmasker("i am trying to [MASK] python language to get a job as a data scientist",top_k=2)
```

```
[{'score': 0.4161553680896759,
  'token': 4553,
  'token_str': 'learn',
  'sequence': 'i am trying to learn python language to get a job as a data scientist'},
 {'score': 0.3323000371456146,
  'token': 2224,
  'token_str': 'use',
  'sequence': 'i am trying to use python language to get a job as a data scientist'}]
```

```python
new=pipeline("ner",grouped_entities=True)
new("my name is divya and i study at nit in srinagar")
```

No model was supplied, defaulted to dbmdz/bert-large-cased-finetuned-conll03-english and revision 4c53496 (https://huggingface.co/dbmdz/)
Using a pipeline without specifying a model name and revision in production is not recommended.

config.json: 100%   998/998 [00:00<00:00, 30.7kB/s]

model.safetensors: 100%   1.33G/1.33G [00:51<00:00, 20.0MB/s]

Some weights of the model checkpoint at dbmdz/bert-large-cased-finetuned-conll03-english were not used when initializing BertForTokenCla
- This IS expected if you are initializing BertForTokenClassification from the checkpoint of a model trained on another task or with and
- This IS NOT expected if you are initializing BertForTokenClassification from the checkpoint of a model that you expect to be exactly i

tokenizer_config.json: 100%   60.0/60.0 [00:00<00:00, 6.22kB/s]

vocab.txt: 213k/? [00:00<00:00, 10.9MB/s]

Device set to use cuda:0
/usr/local/lib/python3.11/dist-packages/transformers/pipelines/token_classification.py:186: UserWarning: `grouped_entities` is deprecate
  warnings.warn(
[{'entity_group': 'PER',
  'score': np.float32(0.73323137),
  'word': 'divya',
  'start': 11,
  'end': 16},
 {'entity_group': 'LOC',
  'score': np.float32(0.5105514),
  'word': 'srinagar',
  'start': 39,
  'end': 47}]

```python
que=pipeline("question-answering")
que(
    question="Where do I work?",
    context="My name is Divya and I work at NIT SRINAGAR",
)
```

No model was supplied, defaulted to distilbert/distilbert-base-cased-distilled-squad and revision 564e9b5 (https://huggingface.co/disti)
Using a pipeline without specifying a model name and revision in production is not recommended.

config.json: 100%   473/473 [00:00<00:00, 47.2kB/s]

model.safetensors: 100%   261M/261M [00:03<00:00, 107MB/s]

tokenizer_config.json: 100%   49.0/49.0 [00:00<00:00, 3.45kB/s]

vocab.txt: 213k/? [00:00<00:00, 11.2MB/s]

tokenizer.json: 436k/? [00:00<00:00, 32.6MB/s]

Device set to use cuda:0
{'score': 0.9773680487269303, 'start': 31, 'end': 43, 'answer': 'NIT SRINAGAR'}

```python
summ=pipeline("summarization")
summ(
    """
    Wikipedia is a free online encyclopedia that anyone can edit, and millions already have.

Wikipedia's purpose is to benefit readers by presenting information on all branches of knowledge. Hosted by the Wikimedia Foundation, Wikipe

Written collaboratively by volunteers known as Wikipedians, Wikipedia articles can be edited by anyone with Internet access, except in limit
    """
)
```

No model was supplied, defaulted to sshleifer/distilbart-cnn-12-6 and revision a4f8f3e (https://huggingface.co/sshleifer/distilbart-cnn-)
Using a pipeline without specifying a model name and revision in production is not recommended.

config.json: 1.80k/? [00:00<00:00, 110kB/s]

pytorch_model.bin: 100%   1.22G/1.22G [00:13<00:00, 57.4MB/s]

model.safetensors: 100%   1.22G/1.22G [00:28<00:00, 160MB/s]

tokenizer_config.json: 100%   26.0/26.0 [00:00<00:00, 1.00kB/s]

vocab.json: 899k/? [00:00<00:00, 20.2MB/s]

merges.txt: 456k/? [00:00<00:00, 22.4MB/s]

Device set to use cuda:0
[{'summary_text': " Wikipedia is a free online encyclopedia that can be edited by anyone with Internet access . Since its creation on
January 15, 2001, it has grown into the world's largest reference website . Wikipedia currently has more than sixty-four million

```python
tra=pipeline("translation",model="Helsinki-NLP/opus-mt-fr-en")
tra("Ce cours est produit par Hugging Face.")
```

config.json: 1.42k/? [00:00<00:00, 99.5kB/s]

pytorch_model.bin: 100%   301M/301M [00:16<00:00, 10.8MB/s]

generation_config.json: 100%   293/293 [00:00<00:00, 8.29kB/s]

config.json: 1.42k/? [00:00<00:00, 99.5kB/s]

pytorch_model.bin: 100%   301M/301M [00:16<00:00, 10.8MB/s]