```
!pip install spacy transformers torch pandas datasets seqeval
!python -m spacy download en_core_web_sm
```

```
Requirement already satisfied: spacy in /usr/local/lib/python3.12/dist-packages (3.8.11)
Requirement already satisfied: transformers in /usr/local/lib/python3.12/dist-packages (5.0.0)
Requirement already satisfied: torch in /usr/local/lib/python3.12/dist-packages (2.10.0+cpu)
Requirement already satisfied: pandas in /usr/local/lib/python3.12/dist-packages (2.2.2)
Requirement already satisfied: datasets in /usr/local/lib/python3.12/dist-packages (4.0.0)
Collecting seqeval
  Downloading seqeval-1.2.2.tar.gz (43 kB)
                                            ━━━━━━━━━━━━ 43.6/43.6 kB 3.3 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (1.0.15)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.0.13)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.0.12)
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /usr/local/lib/python3.12/dist-packages (from spacy) (8.3.10)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.12/dist-packages (from spacy) (1.1.3)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.5.2)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from spacy) (0.4.3)
Requirement already satisfied: typer-slim<1.0.0,>=0.3.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (0.24.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (4.67.3)
Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.0.2)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (2.32.4)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.12/dist-packages (from spacy) (2
Requirement already satisfied: jinja2 in /usr/local/lib/python3.12/dist-packages (from spacy) (3.1.6)
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-packages (from spacy) (75.2.0)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from spacy) (26.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.12/dist-packages (from transformers) (3.24.2)
Requirement already satisfied: huggingface-hub<2.0,>=1.3.0 in /usr/local/lib/python3.12/dist-packages (from transformers) (1.4
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.12/dist-packages (from transformers) (6.0.3)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.12/dist-packages (from transformers) (2025.11.3)
Requirement already satisfied: tokenizers<=0.23.0,>=0.22.0 in /usr/local/lib/python3.12/dist-packages (from transformers) (0.22
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.12/dist-packages (from transformers) (0.7.0)
Requirement already satisfied: typing-extensions>=4.10.0 in /usr/local/lib/python3.12/dist-packages (from torch) (4.15.0)
Requirement already satisfied: sympy>=1.13.3 in /usr/local/lib/python3.12/dist-packages (from torch) (1.14.0)
Requirement already satisfied: networkx>=2.5.1 in /usr/local/lib/python3.12/dist-packages (from torch) (3.6.1)
Requirement already satisfied: fsspec>=0.8.5 in /usr/local/lib/python3.12/dist-packages (from torch) (2025.3.0)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.3)
Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.12/dist-packages (from datasets) (18.1.0)
Requirement already satisfied: dill<0.3.9,>=0.3.0 in /usr/local/lib/python3.12/dist-packages (from datasets) (0.3.8)
Requirement already satisfied: xxhash in /usr/local/lib/python3.12/dist-packages (from datasets) (3.6.0)
Requirement already satisfied: multiprocess<0.70.17 in /usr/local/lib/python3.12/dist-packages (from datasets) (0.70.16)
Requirement already satisfied: scikit-learn>=0.21.3 in /usr/local/lib/python3.12/dist-packages (from seqeval) (1.6.1)
Requirement already satisfied: aiohttp!=4.0.0a0,!=4.0.0a1 in /usr/local/lib/python3.12/dist-packages (from fsspec[http]<=2025.3
Requirement already satisfied: hf-xet<2.0.0,>=1.2.0 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub<2.0,>=1.3
Requirement already satisfied: httpx<1,>=0.23.0 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub<2.0,>=1.3.0->t
Requirement already satisfied: shellingham in /usr/local/lib/python3.12/dist-packages (from huggingface-hub<2.0,>=1.3.0->transt
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8.1,<3
Requirement already satisfied: pydantic-core==2.41.4 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8.1,<3
Requirement already satisfied: typing-inspection>=0.4.2 in /usr/local/lib/python3.12/dist-packages (from pydantic!=1.8,!=1.8.1
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spacy) (3
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spa
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests<3.0.0,>=2.13.0->spa
```

```python
import spacy                        # For NLP processing and pretrained NER
from spacy import displacy          # For visualizing named entities
from transformers import pipeline, AutoTokenizer, AutoModelForTokenClassification
                                    # For Transformer-based NER
import torch                        # Backend deep learning framework
import pandas as pd                 # For tabular data handling
```

```python
nlp = spacy.load("en_core_web_sm")
```

```python
sentences = [
    "Apple CEO Tim Cook announced the new iPhone in California.",
    "Lionel Messi scored a goal for Inter Miami in the MLS final.",
    "The United Nations held a climate summit in New York.",
    "Google is investing $2 billion in AI research in London.",
    "President Joe Biden met Prime Minister Narendra Modi in Washington."
]
```

```
for sent in sentences:
    doc = nlp(sent)
    print(f"\nSentence: {sent}")
    for ent in doc.ents:
        print(f"Entity: {ent.text} | Label: {ent.label_}")
```

```
Sentence: Apple CEO Tim Cook announced the new iPhone in California.
Entity: Apple | Label: ORG
Entity: Tim Cook | Label: PERSON
Entity: California | Label: GPE

Sentence: Lionel Messi scored a goal for Inter Miami in the MLS final.
Entity: Lionel Messi | Label: PERSON
Entity: MLS | Label: ORG

Sentence: The United Nations held a climate summit in New York.
Entity: The United Nations | Label: ORG
Entity: New York | Label: GPE

Sentence: Google is investing $2 billion in AI research in London.
Entity: Google | Label: ORG
Entity: $2 billion | Label: MONEY
Entity: AI | Label: GPE
Entity: London | Label: GPE

Sentence: President Joe Biden met Prime Minister Narendra Modi in Washington.
Entity: Joe Biden | Label: PERSON
Entity: Narendra Modi | Label: PERSON
Entity: Washington | Label: GPE
```

```
data = []

for sent in sentences:
    doc = nlp(sent)
    for ent in doc.ents:
        data.append([sent, ent.text, ent.label_])

spacy_df = pd.DataFrame(data, columns=["Sentence", "Entity", "Label"])
spacy_df
```

	Sentence	Entity	Label
0	Apple CEO Tim Cook announced the new iPhone in...	Apple	ORG
1	Apple CEO Tim Cook announced the new iPhone in...	Tim Cook	PERSON
2	Apple CEO Tim Cook announced the new iPhone in...	California	GPE
3	Lionel Messi scored a goal for Inter Miami in ...	Lionel Messi	PERSON
4	Lionel Messi scored a goal for Inter Miami in ...	MLS	ORG
5	The United Nations held a climate summit in Ne...	The United Nations	ORG
6	The United Nations held a climate summit in Ne...	New York	GPE
7	Google is investing $2 billion in AI research ...	Google	ORG
8	Google is investing $2 billion in AI research ...	$2 billion	MONEY
9	Google is investing $2 billion in AI research ...	AI	GPE
10	Google is investing $2 billion in AI research ...	London	GPE
11	President Joe Biden met Prime Minister Narendr...	Joe Biden	PERSON
12	President Joe Biden met Prime Minister Narendr...	Narendra Modi	PERSON
13	President Joe Biden met Prime Minister Narendr...	Washington	GPE

Next steps: [Generate code with `spacy_df`] [New interactive sheet]

```
doc = nlp(sentences[0])
displacy.render(doc, style="ent", jupyter=True)
```

Apple **ORG** CEO Tim Cook **PERSON** announced the new iPhone in California **GPE** .

```python
model_name = "dbmdz/bert-large-cased-finetuned-conll03-english"

tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForTokenClassification.from_pretrained(model_name)

ner_pipeline = pipeline("ner", model=model, tokenizer=tokenizer)
```

```
/usr/local/lib/python3.12/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
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%                                     998/998 [00:00<00:00, 12.7kB/s]

tokenizer_config.json: 100%                           60.0/60.0 [00:00<00:00, 929B/s]

vocab.txt:       213k/? [00:00<00:00, 1.81MB/s]

Warning: You are sending unauthenticated requests to the HF Hub. Please set a HF_TOKEN to enable higher rate limits and faster d
WARNING:huggingface_hub.utils._http:Warning: You are sending unauthenticated requests to the HF Hub. Please set a HF_TOKEN to en

model.safetensors: 100%                               1.33G/1.33G [00:20<00:00, 160MB/s]

Loading weights: 100%                                 391/391 [00:00<00:00, 378.46it/s, Materializing param=classifier.weight]

BertForTokenClassification LOAD REPORT from: dbmdz/bert-large-cased-finetuned-conll03-english
Key                      | Status      | |
-------------------------+------------+--+-
bert.pooler.dense.bias   | UNEXPECTED |  |
bert.pooler.dense.weight | UNEXPECTED |  |

Notes:
- UNEXPECTED    :can be ignored when loading from different task/architecture; not ok if you expect identical arch.
```

```python
for sent in sentences:
    print(f"\nSentence: {sent}")
    results = ner_pipeline(sent)
    for r in results:
        print(f"Word: {r['word']} | Label: {r['entity']} | Score: {r['score']:.4f}")
```

```
Sentence: Apple CEO Tim Cook announced the new iPhone in California.
Word: Apple | Label: I-ORG | Score: 0.9973
Word: Tim | Label: I-PER | Score: 0.9995
Word: Cook | Label: I-PER | Score: 0.9997
Word: iPhone | Label: I-MISC | Score: 0.9956
Word: California | Label: I-LOC | Score: 0.9997

Sentence: Lionel Messi scored a goal for Inter Miami in the MLS final.
Word: Lionel | Label: I-PER | Score: 0.9991
Word: Me | Label: I-PER | Score: 0.9992
Word: ##ssi | Label: I-PER | Score: 0.9982
Word: Inter | Label: I-ORG | Score: 0.9989
Word: Miami | Label: I-ORG | Score: 0.9645
Word: MLS | Label: I-ORG | Score: 0.9215

Sentence: The United Nations held a climate summit in New York.
Word: United | Label: I-ORG | Score: 0.9994
Word: Nations | Label: I-ORG | Score: 0.9994
Word: New | Label: I-LOC | Score: 0.9994
Word: York | Label: I-LOC | Score: 0.9994

Sentence: Google is investing $2 billion in AI research in London.
Word: Google | Label: I-ORG | Score: 0.9987
Word: London | Label: I-LOC | Score: 0.9995

Sentence: President Joe Biden met Prime Minister Narendra Modi in Washington.
Word: Joe | Label: I-PER | Score: 0.9991
Word: B | Label: I-PER | Score: 0.9975
Word: ##iden | Label: I-PER | Score: 0.9966
Word: Na | Label: I-PER | Score: 0.9994
Word: ##ren | Label: I-PER | Score: 0.9915
Word: ##dra | Label: I-PER | Score: 0.9984
Word: Mo | Label: I-PER | Score: 0.9987
Word: ##di | Label: I-PER | Score: 0.9848
Word: Washington | Label: I-LOC | Score: 0.9990
```

```python
clean_data = []

for sent in sentences:
    results = ner_pipeline(sent)
```

```
    merged_entities = {}

    for r in results:
        word = r['word'].replace("##", "")
        label = r['entity']
        score = r['score']

        clean_data.append([sent, word, label, score])

hf_df = pd.DataFrame(clean_data,
                columns=["Sentence", "Entity", "Label", "Confidence Score"])

hf_df
```

	Sentence	Entity	Label	Confidence Score
0	Apple CEO Tim Cook announced the new iPhone in...	Apple	I-ORG	0.997256
1	Apple CEO Tim Cook announced the new iPhone in...	Tim	I-PER	0.999539
2	Apple CEO Tim Cook announced the new iPhone in...	Cook	I-PER	0.999692
3	Apple CEO Tim Cook announced the new iPhone in...	iPhone	I-MISC	0.995607
4	Apple CEO Tim Cook announced the new iPhone in...	California	I-LOC	0.999726
5	Lionel Messi scored a goal for Inter Miami in ...	Lionel	I-PER	0.999144
6	Lionel Messi scored a goal for Inter Miami in ...	Me	I-PER	0.999151
7	Lionel Messi scored a goal for Inter Miami in ...	ssi	I-PER	0.998244
8	Lionel Messi scored a goal for Inter Miami in ...	Inter	I-ORG	0.998893
9	Lionel Messi scored a goal for Inter Miami in ...	Miami	I-ORG	0.964528
10	Lionel Messi scored a goal for Inter Miami in ...	MLS	I-ORG	0.921494
11	The United Nations held a climate summit in Ne...	United	I-ORG	0.999358
12	The United Nations held a climate summit in Ne...	Nations	I-ORG	0.999425
13	The United Nations held a climate summit in Ne...	New	I-LOC	0.999352
14	The United Nations held a climate summit in Ne...	York	I-LOC	0.999412
15	Google is investing $2 billion in AI research ...	Google	I-ORG	0.998746
16	Google is investing $2 billion in AI research ...	London	I-LOC	0.999530
17	President Joe Biden met Prime Minister Narendr...	Joe	I-PER	0.999054
18	President Joe Biden met Prime Minister Narendr...	B	I-PER	0.997452
19	President Joe Biden met Prime Minister Narendr...	iden	I-PER	0.996629
20	President Joe Biden met Prime Minister Narendr...	Na	I-PER	0.999361
21	President Joe Biden met Prime Minister Narendr...	ren	I-PER	0.991492
22	President Joe Biden met Prime Minister Narendr...	dra	I-PER	0.998367
23	President Joe Biden met Prime Minister Narendr...	Mo	I-PER	0.998671
24	President Joe Biden met Prime Minister Narendr...	di	I-PER	0.984771
25	President Joe Biden met Prime Minister Narendr...	Washington	I-LOC	0.999043

Next steps: (Generate code with `hf_df`) (New interactive sheet)