Copy of Untitled5.ipynb ☆

File Edit View Insert Runtime Tools Help   All changes saved

+ Code   + Text

Connect T4   ◆ Gemini

+ Code   + Text

```
[ ] print("Hello, Google Colab!")
```

Hello, Google Colab!

∨ MY FIRST COLAB NOTEBOOK

This is a markdown cdll where you can add text, links, images, and more.

```
from huggingface_hub import login
# Replace 'your-access-token' with your actual Hugging Face access token
login(token="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")
```

The token has not been saved to the git credentials helper. Pass `add_to_git_credential=True` in this function directly or `--add-to-git-credential` if using vi
Token is valid (permission: read).
Your token has been saved to /root/.cache/huggingface/token
Login successful

```
[ ] from transformers import AutoModelForMaskedLM, AutoTokenizer, pipeline

    # Load the pre-trained model and tokenizer from Hugging Face
    model_name = "bert-base-uncased"
    tokenizer = AutoTokenizer.from_pretrained(model_name)
```

```
pip install huggingface-hub
```

```
Requirement already satisfied: huggingface-hub in /usr/local/lib/python3.10/dist-packages (0.24.7)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (3.16.1)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (2024.6.1)
Requirement already satisfied: packaging>=20.9 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (24.1)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (6.0.2)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (2.32.3)
Requirement already satisfied: tqdm>=4.42.1 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (4.66.5)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub) (2024.8.30)
```

Copy of Untitled6.ipynb ☆

File Edit View Insert Runtime Tools Help   Last saved at 11:29 AM

Code + Text   Connect ▼   ✦ Gemini ∧

                        + Code   + Text

```
pip install git+https://github.com/huggingface/transformers
```

Collecting git+https://github.com/huggingface/transformers
  Cloning https://github.com/huggingface/transformers to /tmp/pip-req-build-2nojo21o
  Running command git clone --filter=blob:none --quiet https://github.com/huggingface/transformers /tmp/pip-req-build-2nojo21o
  Resolved https://github.com/huggingface/transformers to commit 78b2929c0554b79e0489b451ce4ece14d265ead2
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers==4.45.0.dev0) (3.16.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.23.2 in /usr/local/lib/python3.10/dist-packages (from transformers==4.45.0.dev0) (0.24.7)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from transformers==4.45.0.dev0) (1.26.4)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers==4.45.0.dev0) (24.1)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers==4.45.0.dev0) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers==4.45.0.dev0) (2024.9.11)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from transformers==4.45.0.dev0) (2.32.3)
Requirement already satisfied: tokenizers<0.20,>=0.19 in /usr/local/lib/python3.10/dist-packages (from transformers==4.45.0.dev0) (0.19.1)
Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from transformers==4.45.0.dev0) (0.4.5)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.10/dist-packages (from transformers==4.45.0.dev0) (2024.6)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub<1.0,>=0.23.2->transformers==4.45.0.dev0) (4.66.5)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub<1.0,>=0.23.2->transformers==4.45.0.dev
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->transformers==4.45.0.dev0) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->transformers==4.45.0.dev0) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->transformers==4.45.0.dev0) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->transformers==4.45.0.dev0) (2024.8.30)
```

```
pip install accelerate protobuf sentencepiece torch
```

```
Requirement already satisfied: accelerate in /usr/local/lib/python3.10/dist-packages (0.34.2)
Requirement already satisfied: protobuf in /usr/local/lib/python3.10/dist-packages (3.20.3)
Requirement already satisfied: sentencepiece in /usr/local/lib/python3.10/dist-packages (0.1.99)
Requirement already satisfied: torch in /usr/local/lib/python3.10/dist-packages (2.4.1+cu121)
Requirement already satisfied: numpy<3.0.0,>=1.17 in /usr/local/lib/python3.10/dist-packages (from accelerate) (1.26.4)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from accelerate) (24.1)
Requirement already satisfied: psutil in /usr/local/lib/python3.10/dist-packages (from accelerate) (5.9.5)
Requirement already satisfied: pyyaml in /usr/local/lib/python3.10/dist-packages (from accelerate) (6.0.2)
Requirement already satisfied: huggingface-hub>=0.21.0 in /usr/local/lib/python3.10/dist-packages (from accelerate) (0.24.7)
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.10/dist-packages (from accelerate) (0.4.5)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from torch) (3.16.1)
Requirement already satisfied: typing-extensions>=4.8.0 in /usr/local/lib/python3.10/dist-packages (from torch) (4.12.2)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch) (1.13.2)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch) (3.3)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from torch) (3.1.4)
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (from torch) (2024.6.1)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.21.0->accelerate) (2.32.3)
Requirement already satisfied: tqdm>=4.42.1 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.21.0->accelerate) (4.66.5)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2->torch) (2.1.5)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.10/dist-packages (from sympy->torch) (1.3.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub>=0.21.0->accelerate) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub>=0.21.0->accelerate) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub>=0.21.0->accelerate) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub>=0.21.0->accelerate) (2024.8.30)
```

CO ▲ Copy of Untitled5.ipynb ☆   💬 Comment   👥 Share ⚙ 👤

File Edit View Insert Runtime Tools Help   All changes saved

+ Code + Text   Connect T4 ▼ ✦ Gemini ⌃

```
model_name = bert-base-uncased
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForMaskedLM.from_pretrained(model_name)

# Create a pipeline for masked language modeling
nlp_pipeline = pipeline("fill-mask", model=model, tokenizer=tokenizer)

# Test the pipeline with a simple input
test_sentence = "The quick brown fox jumps over the [MASK] dog."
result = nlp_pipeline(test_sentence)

print(result)
```

```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_token.py:89: 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 secret in your Google Colab a
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(
tokenizer_config.json: 100%  ████████████████  48.0/48.0 [00:00<00:00, 2.32kB/s]
config.json: 100%             ████████████████  570/570 [00:00<00:00, 28.3kB/s]
vocab.txt: 100%               ████████████████  232k/232k [00:00<00:00, 10.2MB/s]
tokenizer.json: 100%          ████████████████  466k/466k [00:00<00:00, 941kB/s]
/usr/local/lib/python3.10/dist-packages/transformers/tokenization_utils_base.py:1601: FutureWarning: `clean_up_tokenization_spaces` was not set. It will be set
  warnings.warn(
model.safetensors: 100%       ████████████████  440M/440M [00:03<00:00, 101MB/s]
```

+ Code + Text

```python
from transformers import AutoModelForCausalLM
```

```python
from transformers import AutoModelForCausalLM, AutoTokenizer
```

```python
# Replace 'your-access-token' with you
from huggingface_hub import login

# Assign your token to a variable
token = "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"   # Replace with your actual token

# Log in using the token
login(token=token)

login(token="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")
model_id = "NousResearch/Llama-2-7b-chat-hf"
model = AutoModelForCausalLM.from_pretrained(model_id, torch_dtype=torch.float16, device_map="auto")
tokenizer = AutoTokenizer.from_pretrained(model_id)
tokenizer.use_default_system_prompt = False
```

```
LlamaForCausalLM(
    (model): LlamaModel(
        (embed_tokens): Embedding(32000, 4096, padding_idx=0)
        (layers): ModuleList(
            (0-31): 32 x LlamaDecoderLayer(
                (self_attn): LlamaSdpaAttention(
                    (q_proj): Linear(in_features=4096, out_features=4096, bias=False)
                    (k_proj): Linear(in_features=4096, out_features=4096, bias=False)
                    (v_proj): Linear(in_features=4096, out_features=4096, bias=False)
                    (o_proj): Linear(in_features=4096, out_features=4096, bias=False)
                    (rotary_emb): LlamaRotaryEmbedding()
                )
                (mlp): LlamaMLP(
                    (gate_proj): Linear(in_features=4096, out_features=11008, bias=False)
                    (up_proj): Linear(in_features=4096, out_features=11008, bias=False)
                    (down_proj): Linear(in_features=11008, out_features=4096, bias=False)
                    (act_fn): SiLU()
                )
                (input_layernorm): LlamaRMSNorm((4096,), eps=1e-05)
                (post_attention_layernorm): LlamaRMSNorm((4096,), eps=1e-05)
            )
        )
        (norm): LlamaRMSNorm((4096,), eps=1e-05)
        (rotary_emb): LlamaRotaryEmbedding()
    )
    (lm_head): Linear(in_features=4096, out_features=32000, bias=False)
)
```

CO △ Copy of Untitled6.ipynb ☆

+ Code + Text   Connect ▼   ◆ Gemini   ⌃

```python
# Replace 'your-access-token' with you
from huggingface_hub import login

# Assign your token to a variable
token = "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"   # Replace with your actual token

# Log in using the token
login(token=token)
import torch
login(token="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")
model_id = "NousResearch/Llama-2-7b-chat-hf"
model = AutoModelForCausalLM.from_pretrained(model_id, torch_dtype=torch.float16, device_map="auto")
tokenizer = AutoTokenizer.from_pretrained(model_id)
tokenizer.use_default_system_prompt = False
```

```
The token has not been saved to the git credentials helper. Pass `add_to_git_credential=True` in this function directly or `--add-to-git-credential` if using vi.
Token is valid (permission: read).
Your token has been saved to /root/.cache/huggingface/token
Login successful
The token has not been saved to the git credentials helper. Pass `add_to_git_credential=True` in this function directly or `--add-to-git-credential` if using vi.
Token is valid (permission: read).
Your token has been saved to /root/.cache/huggingface/token
Login successful
Loading checkpoint shards: 100% ████████████████████████  2/2 [00:00<00:00,  3.97it/s]
WARNING:accelerate.big_modeling:Some parameters are on the meta device because they were offloaded to the disk and cpu.
```

```python
def display_masked_sentence(sentence, ner_results):
    masked_sentence = sentence
    for entity in ner_results:
        entity_word = entity['word']
        entity_label = entity['entity']
        masked_sentence = masked_sentence.replace(entity_word, f"[{entity_label}]")
     return masked_sentence

# Test the function with the NER results
masked_sentence = display_masked_sentence(test_sentence, result)
print(masked_sentence)
```

```
[I-ORG] is planning to build a new campus in [I-LOC].
```

```python
# Load the summarization pipeline
summarizer = pipeline("summarization", model="facebook/bart-large-cnn")

# Test the summarizer with a longer input text
long_text = (
    "The quick brown fox jumps over the lazy dog. The lazy dog, however, was not really lazy. "
    "It was simply tired from chasing after the quick brown fox all day. The two animals had a "
    "long history of playful rivalry, with the fox always outwitting the dog. Despite their differences, "
    "they shared a bond of mutual respect and friendship."
)
```

```python
# Load a pre-trained NER model
model_name = "dbmdz/bert-large-cased-finetuned-conll03-english"
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForTokenClassification.from_pretrained(model_name)

# Create an NER pipeline
ner_pipeline = pipeline("ner", model=model, tokenizer=tokenizer)

# Test the NER pipeline with a sample sentence
test_sentence = "Apple is planning to build a new campus in Austin."
result = ner_pipeline(test_sentence)
print(result)
```

tokenizer_config.json: 100%   60.0/60.0 [00:00<00:00, 3.16kB/s]

config.json: 100%   998/998 [00:00<00:00, 56.2kB/s]

vocab.txt: 100%   213k/213k [00:00<00:00, 4.06MB/s]

/usr/local/lib/python3.10/dist-packages/transformers/tokenization_utils_base.py:1617: FutureWarning: `clean_up_tokenization_spaces` was not set. It will be set
    warnings.warn(

model.safetensors: 100%   1.33G/1.33G [00:14<00:00, 149MB/s]

Some weights of the model checkpoint at dbmdz/bert-large-cased-finetuned-conll03-english were not used when initializing BertForTokenClassification: ['bert.pool
- This IS expected if you are initializing BertForTokenClassification from the checkpoint of a model trained on another task or with another architecture (e.g. :
- This IS NOT expected if you are initializing BertForTokenClassification from the checkpoint of a model that you expect to be exactly identical (initializing a