


```
from transformers import GPT2LMHeadModel, GPT2Tokenizer
import torch
```


```
model_name = "gpt2"
tokenizer = GPT2Tokenizer.from_pretrained(model_name)
model = GPT2LMHeadModel.from_pretrained(model_name)
```

 /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 :
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% 26.0/26.0 [00:00<00:00, 809B/s]
vocab.json: 100% 1.04M/1.04M [00:00<00:00, 15.5MB/s]
merges.txt: 100% 456k/456k [00:00<00:00, 17.8MB/s]
tokenizer.json: 100% 1.36M/1.36M [00:00<00:00, 62.0MB/s]
config.json: 100% 665/665 [00:00<00:00, 62.5kB/s]
Xet Storage is enabled for this repo, but the 'hf_xet' package is not installed. Falling back to regular HTTP download. For better performance, install the 'hf_xet' package.
WARNING:huggingface_hub.file_download:Xet Storage is enabled for this repo, but the 'hf_xet' package is not installed. Falling back to regular HTTP download.
model.safetensors: 100% 548M/548M [00:01<00:00, 306MB/s]
generation_config.json: 100% 124/124 [00:00<00:00, 9.63kB/s]

```
prompt_text = "Once upon a time in the world of Artificial Intelligence,"
input_ids = tokenizer.encode(prompt_text, return_tensors='pt')
```

```
output = model.generate(  
    input_ids,  
    max_length=100,  
    num_return_sequences=1,  
    no_repeat_ngram_size=2,  
    early_stopping=True  
)
```

 /usr/local/lib/python3.11/dist-packages/transformers/generation/configuration_utils.py:679: UserWarning: `num_beams` is set to 1. However, the attention mask and the pad token id were not set. As a consequence, you may observe unexpected behavior. Please pass your input's `attention_mask` and `pad_token_id` to `generate`. Setting `pad_token_id` to `eos_token_id`:50256 for open-end generation.
The attention mask is not set and cannot be inferred from input because pad token is same as eos token. As a consequence, you may observe unexpected behavior. Please pass your input's `attention_mask` and `pad_token_id` to `generate`.

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

 Once upon a time in the world of Artificial Intelligence, the human mind was a machine. It was the only thing that could be controlled by the human brain. The human brain was an artificial intelligence. The human body was not. But it was still a human being. And it had the ability to control the machine.