

Falcon 40B

```
pip install torch --no-cache-dir
pip install accelerate
pip install transformers
```

```
from transformers import AutoTokenizer, AutoModelForCausalLM
import transformers
import torch

model = "tiiuae/falcon-40b"

tokenizer = AutoTokenizer.from_pretrained(model)
pipeline = transformers.pipeline(
    "text-generation",
    model=model,
    tokenizer=tokenizer,
    torch_dtype=torch.bfloat16,
    device_map="auto",
)
sequences = pipeline(
    "Inventate una historia de un asesinato en una casa unifamiliar y explica 10 pistas encontradas:",
    max_length=200,
    do_sample=True,
    top_k=10,
    num_return_sequences=1,
    eos_token_id=tokenizer.eos_token_id,
)
for seq in sequences:
    print(f"Result: {seq['generated_text']}")
```

Falcon 180B

<https://huggingface.co/tiiuae/falcon-180B>

Hay que logarse para poder usarlo. L< b del model tiene que ser mayúscula.

From:
<http://wiki.legido.com/> - **Legido Wiki**

Permanent link:
http://wiki.legido.com/doku.php?id=informatica:inteligencia_artificial:falcon

Last update: **2023/10/23 20:11**



