

Maven_TextStreamer_Mistral_Nemo_Instruct_2407

August 15, 2025

```
[1]: %%capture
      #!pip install transformers
      #!pip install bitsandbytes

[2]: # from huggingface_hub import notebook_login
      # notebook_login()
      import os
      os.getpid()
```

[2]: 76980

0.1 Using TextStreamer

```
[3]: from dotenv import load_dotenv
      import os

      load_dotenv() # Make sure .env is in the same folder as this notebook
      token = os.getenv("HUGGINGFACE_HUB_TOKEN")
      assert token, "No HUGGINGFACE_HUB_TOKEN found in .env"

      # ---- 2) Log in to Hugging Face ----
      from huggingface_hub import login, whoami

      login(token=token) # Auth for this Python process
      #print("HF user info:", whoami()) # Should print your username/orgs

      from transformers import AutoTokenizer, AutoModelForCausalLM, pipeline, \
          TextStreamer
      import torch
      from transformers import BitsAndBytesConfig

      model_id = "unsloth/Mistral-Nemo-Instruct-2407" # Replace with your model

      # 4-bit quantization configuration
      quantization_config = BitsAndBytesConfig(
          load_in_4bit=True,
          bnb_4bit_compute_dtype=torch.float16,
          bnb_4bit_use_double_quant=True,
```

```

        bnb_4bit_quant_type="nf4"
    )

    # Load tokenizer and model in 4-bit
    tokenizer = AutoTokenizer.from_pretrained(model_id)
    model = AutoModelForCausalLM.from_pretrained(
        model_id,
        quantization_config=quantization_config,
        device_map="auto"
    )

```

```

Loading checkpoint shards: 0%|          | 0/5 [00:00<?, ?it/s]
generation_config.json: 0%|          | 0.00/162 [00:00<?, ?B/s]

```

```

[4]: from transformers import TextStreamer

# Define Alpaca-style prompt format
alpaca_prompt = """Below is an instruction that describes a task. Write a
↳response that appropriately completes the request.

### Instruction:
{}

### Response:
"""

# Prepare input text
prompt_text = alpaca_prompt.format("What is the importance of using renewable
↳energy?") # instruction

inputs = tokenizer([prompt_text], return_tensors="pt").to(model.device) # Move
↳inputs to model's device

# Initialize text streamer
text_streamer = TextStreamer(tokenizer, skip_prompt=True,
↳skip_special_tokens=False)

# Generate response with streamer
_ = model.generate(**inputs, streamer=text_streamer, max_new_tokens=300)

```

Using renewable energy is important for several reasons:

1. **Environmental Benefits**: Renewable energy sources like solar, wind, and hydropower produce little to no greenhouse gas emissions or air pollution, helping to mitigate climate change and improve overall air quality.
2. **Energy Independence**: By harnessing our own natural resources, countries

can reduce their dependence on imported fossil fuels, improving energy security and potentially lowering energy costs.

3. ****Sustainability****: Renewable energy is virtually inexhaustible, ensuring a long-term, stable energy supply. This is in contrast to finite fossil fuels, which will eventually run out.

4. ****Economic Opportunities****: The renewable energy sector creates new jobs and stimulates economic growth. According to the International Renewable Energy Agency (IRENA), the renewable energy sector already employed 11.5 million people worldwide in 2019.

5. ****Health Benefits****: By reducing air pollution, renewable energy can lead to significant health benefits, including fewer cases of respiratory diseases and related health issues.

6. ****Meeting Climate Goals****: Transitioning to renewable energy is crucial for meeting global climate goals, such as those outlined in the Paris Agreement, which aim to limit global temperature rise to well below 2°C.

[]: