

☰

🔍

<>

🔑

📁

▶

```
import gradio as gr
import torch
from transformers import AutoTokenizer, AutoModelForCausalLM

# Load model and tokenizer
model_name = "ibm-granite/granite-70b-instruct"

tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForCausalLM.from_pretrained(model_name,
                                              torch_dtype=torch.float16,
                                              device_map="auto" # Automatic device placement
)

# Function to get response
def ask_question(prompt):
    inputs = tokenizer(prompt, return_tensors="pt").to(model.device)
    outputs = model.generate(**inputs,
                             max_new_tokens=200,
                             temperature=0.7,
                             top_p=0.9,
                             do_sample=True)
    return tokenizer.decode(outputs[0][len(inputs[0]):])

# Gradio UI
demo = gr.Interface(
    fn=ask_question,
    inputs=gr.Textbox(lines=3),
    outputs="text",
    title="Edu Tutor AI",
    description="Personalized AI Tutor"
)

demo.launch()
```

🔄

```
/usr/local/lib/python3.12/dist-packages/transformers/tokenization_utils_base.py:1219: FutureWarning: `tokenizer_config` is deprecated and will be removed in a future version. Please use `tokenizer_kwargs` instead.
warnings.warn(

tokenizer_config.json: 100%
vocab.json: 100%
merges.txt: 100%
tokenizer.json: 100%
added_tokens.json: 100%
special_tokens_map.json: 100%
config.json: 100%
model.safetensors.index.json: 100%
Fetching 2 files: 0%
model-00002-of-00002.safetensors: 100%
model-00001-of-00002.safetensors: 34%
Loading checkpoint shards: 0%
generation_config.json: 0%
It looks like you are running this from a Colab notebook. To share your notebook publicly, please click on the 'Share' button in the top right corner.

Colab notebook detected. To share your notebook publicly, please click on the 'Share' button in the top right corner.
* Running on public URL: https://huggingface.co/spaces/ibm-granite-granite-70b-instruct-gradio

This share link expires in 1 week.
```

No interface