```
import gradio as gr
import torch
from transformers import AutoT

# Load model and tokenizer
model_name = "ibm-granite/gran

tokenizer = AutoTokenizer.from
model = AutoModelForCausalLM.f
    model_name,
    torch_dtype=torch.float16
    device_map="auto"  # Autom
)

# Function to get response
def ask_question(prompt):
    inputs = tokenizer(prompt,
    outputs = model.generate(
        **inputs,
        max_new_tokens=200,
        temperature=0.7,
        top_p=0.9,
        do_sample=True
    )
    return tokenizer.decode(ou

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

demo.launch()
```

```
/usr/local/lib/python3.12/dist
The secret `HF_TOKEN` does not
To authenticate with the Huggi
You will be able to reuse this
Please note that authenticatio
  warnings.warn(
tokenizer_config.json: 8.88k/? [00:00
vocab.json: 777k/? [00:00<00:00, 5.
merges.txt: 442k/? [00:00<00:00, 2.
tokenizer.json: 3.48M/? [00:00<00:00
added_tokens.json: 100%
special_tokens_map.json: 100%
config.json: 100%
model.safetensors.index.json: 29.8k/
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
Colab notebook detected. To sh
* Running on public URL: https
This share link expires in 1 w
```

No interfac