

```
!pip install gradio torch transformers
```

```
Requirement already satisfied: python-multipart<0.0.12, >=0.0.11 in /usr/local/lib/python3.12/dist-packages (from gradio) (0.0.12)
Requirement already satisfied: pyyaml<7.0, >=5.0 in /usr/local/lib/python3.12/dist-packages (from gradio) (6.0.2)
Requirement already satisfied: ruff>=0.9.3 in /usr/local/lib/python3.12/dist-packages (from gradio) (0.12.11)
Requirement already satisfied: safehttpx<0.2.0, >=0.1.6 in /usr/local/lib/python3.12/dist-packages (from gradio) (0.1.6)
Requirement already satisfied: semantic-version~=2.0 in /usr/local/lib/python3.12/dist-packages (from gradio) (2.10.0)
Requirement already satisfied: starlette<1.0, >=0.40.0 in /usr/local/lib/python3.12/dist-packages (from gradio) (0.47.3)
Requirement already satisfied: tomllkit<0.14.0, >=0.12.0 in /usr/local/lib/python3.12/dist-packages (from gradio) (0.13.3)
Requirement already satisfied: typer<1.0, >=0.12 in /usr/local/lib/python3.12/dist-packages (from gradio) (0.17.3)
Requirement already satisfied: typing-extensions~=4.0 in /usr/local/lib/python3.12/dist-packages (from gradio) (4.15.0)
Requirement already satisfied: uvicorn>=0.14.0 in /usr/local/lib/python3.12/dist-packages (from gradio) (0.35.0)
Requirement already satisfied: fsspec in /usr/local/lib/python3.12/dist-packages (from gradio-client==1.12.1->gradio) (2025.3.0)
Requirement already satisfied: websockets<16.0, >=10.0 in /usr/local/lib/python3.12/dist-packages (from gradio-client==1.12.1->gradio) (13.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.12/dist-packages (from torch) (3.19.1)
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-packages (from torch) (75.2.0)
Requirement already satisfied: sympy>=1.13.3 in /usr/local/lib/python3.12/dist-packages (from torch) (1.13.3)
Requirement already satisfied: networkx in /usr/local/lib/python3.12/dist-packages (from torch) (3.5)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch) (12.6.77)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch) (12.6.77)
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.6.80 in /usr/local/lib/python3.12/dist-packages (from torch) (12.6.80)
Requirement already satisfied: nvidia-cudnn-cu12==9.10.2.21 in /usr/local/lib/python3.12/dist-packages (from torch) (9.10.2.21)
Requirement already satisfied: nvidia-cublas-cu12==12.6.4.1 in /usr/local/lib/python3.12/dist-packages (from torch) (12.6.4.1)
Requirement already satisfied: nvidia-cufft-cu12==11.3.0.4 in /usr/local/lib/python3.12/dist-packages (from torch) (11.3.0.4)
Requirement already satisfied: nvidia-curand-cu12==10.3.7.77 in /usr/local/lib/python3.12/dist-packages (from torch) (10.3.7.77)
Requirement already satisfied: nvidia-cusolver-cu12==11.7.1.2 in /usr/local/lib/python3.12/dist-packages (from torch) (11.7.1.2)
Requirement already satisfied: nvidia-cusparselt-cu12==0.7.1 in /usr/local/lib/python3.12/dist-packages (from torch) (0.7.1)
Requirement already satisfied: nvidia-nccl-cu12==2.27.3 in /usr/local/lib/python3.12/dist-packages (from torch) (2.27.3)
Requirement already satisfied: nvidia-nvtx-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch) (12.6.77)
Requirement already satisfied: nvidia-nvjitlink-cu12==12.6.85 in /usr/local/lib/python3.12/dist-packages (from torch) (12.6.85)
Requirement already satisfied: nvidia-cufile-cu12==1.11.1.6 in /usr/local/lib/python3.12/dist-packages (from torch) (1.11.1.6)
Requirement already satisfied: triton==3.4.0 in /usr/local/lib/python3.12/dist-packages (from torch) (3.4.0)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.12/dist-packages (from transformers) (2024.11.6)
Requirement already satisfied: requests in /usr/local/lib/python3.12/dist-packages (from transformers) (2.32.4)
Requirement already satisfied: tokenizers<=0.23.0, >=0.22.0 in /usr/local/lib/python3.12/dist-packages (from transformers) (0.22.0)
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.12/dist-packages (from transformers) (0.6.2)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.12/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.12/dist-packages (from anyio<5.0, >=3.0->gradio) (3.10)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.12/dist-packages (from anyio<5.0, >=3.0->gradio) (1.3.1)
Requirement already satisfied: certifi in /usr/local/lib/python3.12/dist-packages (from httpx<1.0, >=0.24.1->gradio) (2025.8.3)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.12/dist-packages (from httpx<1.0, >=0.24.1->gradio) (1.0.9)
Requirement already satisfied: h11>=0.16 in /usr/local/lib/python3.12/dist-packages (from httpcore==1.*->httpx<1.0, >=0.24.1->gradio) (1.0.9)
Requirement already satisfied: hf-xet<2.0.0, >=1.1.3 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub<1.0, >=0.33.5->gradio) (1.1.3)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas<3.0, >=1.0->gradio) (2.9.0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas<3.0, >=1.0->gradio) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas<3.0, >=1.0->gradio) (2025.2)
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.12/dist-packages (from pydantic<2.12, >=2.0->gradio) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.2 in /usr/local/lib/python3.12/dist-packages (from pydantic<2.12, >=2.0->gradio) (2.33.2)
Requirement already satisfied: typing-inspection>=0.4.0 in /usr/local/lib/python3.12/dist-packages (from pydantic<2.12, >=2.0->gradio) (0.4.0)
Requirement already satisfied: mpmath<1.4, >=1.1.0 in /usr/local/lib/python3.12/dist-packages (from sympy>=1.13.3->torch) (1.3.0)
Requirement already satisfied: click>=8.0.0 in /usr/local/lib/python3.12/dist-packages (from typer<1.0, >=0.12->gradio) (8.2.1)
Requirement already satisfied: shellingham>=1.3.0 in /usr/local/lib/python3.12/dist-packages (from typer<1.0, >=0.12->gradio) (1.5.4)
Requirement already satisfied: rich>=10.11.0 in /usr/local/lib/python3.12/dist-packages (from typer<1.0, >=0.12->gradio) (13.9.4)
Requirement already satisfied: charset-normalizer<4, >=2 in /usr/local/lib/python3.12/dist-packages (from requests->transformers) (3.4.0)
Requirement already satisfied: urllib3<3, >=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests->transformers) (2.5.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2->pandas<3.0, >=1.0->gradio) (1.17.0)
Requirement already satisfied: markdown-it-py>=2.2.0 in /usr/local/lib/python3.12/dist-packages (from rich>=10.11.0->typer<1.0, >=0.12->gradio) (3.0.0)
Requirement already satisfied: pygments<3.0.0, >=2.13.0 in /usr/local/lib/python3.12/dist-packages (from rich>=10.11.0->typer<1.0, >=0.12->gradio) (2.19.2)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.12/dist-packages (from markdown-it-py>=2.2.0->rich>=10.11.0->typer<1.0, >=0.12->gradio) (0.1.2)
```

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

```
model_name = "ibm-granite/granite-3.2-2b-instruct"
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForCausalLM.from_pretrained(
    model_name,
    torch_dtype = torch.float16 if torch.cuda.is_available() else torch.float32,
    device_map="auto" if torch.cuda.is_available() else None
)
```

```
if tokenizer.pad_token is None:
    tokenizer.pad_token = tokenizer.eos_token
```

```
def generate_response(prompt, max_length=512):
    inputs = tokenizer(prompt, return_tensors="pt", truncation=True, max_length=512)
```

```

def generate_response(prompt, return_token_ids = False, max_length=100, min_length=10):
    """Generate a response from the model given a prompt"""

    if torch.cuda.is_available():
        inputs = {k: v.to(model.device) for k, v in inputs.items()}

    with torch.no_grad():
        output = model.generate(
            **inputs,
            max_length=max_length,
            temperature = 0.7,
            do_sample = True,
            pad_token_id = tokenizer.eos_token_id
        )
        response = tokenizer.decode(output[0], skip_special_tokens=True)
        response = response.replace(prompt, "").strip()
        return response

def concept_explnation(concept):
    prompt = f"Explain the concept of {concept} in detail with example"
    return generate_response(prompt, max_length=800)

def quiz_generator(concept):
    prompt = f"Generate 5 quiz question about {concept} with different quies types (multiple choice, true/false, fill in the blank)"
    return generate_response(prompt, max_length=1200)

#create gradio
with gr.Blocks() as app:
    gr.Markdown("# educational AI Assistant")
    with gr.Tabs():
        with gr.TabItem("concept Explanation"):
            concept_input=gr.Textbox(label="Enter a concept", placeholder="e.g., machine Learning")
            explain_btn = gr.Button("Explain")
            explaintion_output = gr.Textbox(label="Explanation", lines=10)


            explain_btn.click(concept_explnation, inputs = concept_input, outputs=explaintion_output)

        with gr.TabItem("Quiz Generator"):
            quiz_input=gr.Textbox(label="Enter a topic", placeholder="e.g., machine Learning")
            quiz_btn = gr.Button("Generate Quiz")
            quiz_output = gr.Textbox(label="Quiz_Questions & answers", lines=15)

            quiz_btn.click(quiz_generator, inputs = quiz_input, outputs=quiz_output)

app.launch(share=True)

```

 /usr/local/lib/python3.12/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). 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. warnings.warn(

tokenizer_config.json: 8.88k/? [00:00<00:00, 441kB/s]

vocab.json: 777k/? [00:00<00:00, 8.08MB/s]

merges.txt: 442k/? [00:00<00:00, 10.1MB/s]

tokenizer.json: 3.48M/? [00:00<00:00, 32.9MB/s]

added_tokens.json: 100% 87.0/87.0 [00:00<00:00, 4.42kB/s]

special_tokens_map.json: 100% 701/701 [00:00<00:00, 35.3kB/s]

config.json: 100% 786/786 [00:00<00:00, 17.8kB/s]

``torch_dtype` is deprecated! Use `dtype` instead!`

model.safetensors.index.json: 29.8k/? [00:00<00:00, 1.84MB/s]

Fetching 2 files: 100% 2/2 [02:18<00:00, 138.54s/it]

model-00002-of-67.1M/67.1M [00:01<00:00, 50.5MB/s]

00002.safetensors: 100%

model-00001-of-5.00G/5.00G [02:17<00:00, 14.3MB/s]

00002.safetensors: 100%

Loading checkpoint shards: 100% 2/2 [00:19<00:00, 8.16s/it]

generation_config.json: 100% 137/137 [00:00<00:00, 11.0kB/s]

Colab notebook detected. To show errors in colab notebook, set debug=True in launch.py

* Running on public URL: <https://1798a736e60af1fc9a.gradio.live>

This share link expires in 1 week. For free permanent hosting and GPU upgrades,

educational AI Assistant

concept Explanation

Quiz Generator

Enter a concept

e.g., machine Learning

Explain

Explanation