

```
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
!nvidia-smi
```

Wed Oct 1 18:41:15 2025

+-----+ NVIDIA-SMI 550.54.15										Driver Version: 550.54.15										CUDA Version: 12.4										+-----+																			
GPU Name										Persistence-M										Bus-Id										Disp.A										Volatile Uncorr. ECC									
Fan Temp Perf										Pwr:Usage/Cap										Memory-Usage										GPU-Util										Compute M.									
																																								MIG M.									
=====										=====										=====										=====										=====									
0 Tesla T4										Off										00000000:00:04.0 Off										0																			
N/A 37C P8										10W / 70W										0MiB / 15360MiB										0%										Default									
																																								N/A									
+-----+										+-----+										+-----+										+-----+																			
+-----+										+-----+										+-----+										+-----+																			
Processes:																																																	
GPU GI CI										PID Type										Process name										GPU Memory																			
ID ID																														Usage																			
=====										=====										=====										=====										=====									
No running processes found																																																	
+-----+										+-----+										+-----+										+-----+																			

```
import torch
from transformers import AutoModelForCausalLM, AutoTokenizer
tokenizer = AutoTokenizer.from_pretrained("stabilityai/stable-code-instruct-3b", trust_remote_code=True)
model = AutoModelForCausalLM.from_pretrained("stabilityai/stable-code-instruct-3b", torch_dtype=torch.bfloat16, trust_remote_code=True)
model.eval()
model = model.cuda()

messages = [
    {
        "role": "system",
        "content": "You are a helpful and polite assistant",
    },
    {
        "role": "user",
        "content": "Write a simple website in HTML. When a user clicks the button, it shows a random joke from a list of 4 jokes."
    },
]
```

```

]

prompt = tokenizer.apply_chat_template(messages, add_generation_prompt=True, tokenize=False)

inputs = tokenizer([prompt], return_tensors="pt").to(model.device)

tokens = model.generate(
    **inputs,
    max_new_tokens=1024,
    temperature=0.5,
    top_p=0.95,
    top_k=100,
    do_sample=True,
    use_cache=True
)

output = tokenizer.batch_decode(tokens[:, inputs.input_ids.shape[-1]:], skip_special_tokens=False)[0]

```

/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>), set it as secret in your Colab secrets. 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 or datasets.

```

warnings.warn(
tokenizer_config.json:      9.35k/? [00:00<00:00, 227kB/s]

tokenizer.json:      2.12M/? [00:00<00:00, 25.4MB/s]

special_tokens_map.json: 100%                               587/587 [00:00<00:00, 20.8kB/s]

config.json: 100%                                           738/738 [00:00<00:00, 26.0kB/s]
`torch_dtype` is deprecated! Use `dtype` instead!
model.safetensors.index.json:      29.4k/? [00:00<00:00, 2.21MB/s]

Fetching 2 files: 100%                                     2/2 [01:52<00:00, 112.00s/it]

model-00002-of-00002.safetensors: 100%                     610M/610M [00:39<00:00, 13.2MB/s]

model-00001-of-00002.safetensors: 100%                     4.98G/4.98G [01:51<00:00, 64.5MB/s]

Loading checkpoint shards: 100%                             2/2 [00:00<00:00, 1.30it/s]

generation_config.json: 100%                               132/132 [00:00<00:00, 11.7kB/s]

Setting `pad_token_id` to `eos_token_id`:0 for open-end generation.

```

```
print(output)
```

Sure, here's a simple HTML page with JavaScript code that displays a random joke when a button is clicked:

```
```html
<!DOCTYPE html>
<html>
<head>
  <title>Random Joke Generator</title>
  <script type="text/javascript">
    // List of jokes
    var jokes = [
      "Why don't scientists trust atoms? Because they make up everything!",
      "Why did the chicken go to the seance? To talk to the other side!",
      "Why did the scarecrow win an award? Because he was outstanding in his field!",
      "Why did the tomato turn red? Because it saw the salad dressing!"
    ];

    // Function to generate a random joke
    function getRandomJoke() {
      var randomIndex = Math.floor(Math.random() * jokes.length);
      return jokes[randomIndex];
    }

    // Function to display a joke
    function displayJoke() {
      var joke = getRandomJoke();
      document.getElementById("joke").innerHTML = joke;
    }
  </script>
</head>
<body>
  <h1>Random Joke Generator</h1>
  <p id="joke"></p>
  <button onclick="displayJoke()">Show Joke</button>
</body>
</html>
```
```

This code defines a list of jokes, a function to generate a random joke from the list, and a function to display the joke on the webpage  
<|endoftext|>

```
!nvidia-smi
```

```
Wed Oct  1 18:50:40 2025
```

```
+-----+
| NVIDIA-SMI 550.54.15                Driver Version: 550.54.15          CUDA Version: 12.4          |
+-----+-----+-----+-----+
| GPU   Name                               Persistence-M | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf              Pwr:Usage/Cap |      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+-----+-----+-----+-----+
```

|                            |       |    |  |           |                  |              |         |            |  |
|----------------------------|-------|----|--|-----------|------------------|--------------|---------|------------|--|
|                            |       |    |  |           |                  | MIG M.       |         |            |  |
| =====                      |       |    |  |           |                  | +            |         | =====      |  |
| 0                          | Tesla | T4 |  | Off       | 00000000:00:04.0 | Off          | 0       |            |  |
| N/A                        | 52C   | P8 |  | 10W / 70W | 0MiB / 15360MiB  | 0%           | Default |            |  |
|                            |       |    |  |           |                  |              |         | N/A        |  |
| -----                      |       |    |  |           |                  | +            |         | -----      |  |
|                            |       |    |  |           |                  |              |         |            |  |
| +-----                     |       |    |  |           |                  |              |         |            |  |
| Processes:                 |       |    |  |           |                  |              |         |            |  |
| GPU                        | GI    | CI |  | PID       | Type             | Process name |         | GPU Memory |  |
|                            | ID    | ID |  |           |                  |              |         | Usage      |  |
| =====                      |       |    |  |           |                  |              |         |            |  |
| No running processes found |       |    |  |           |                  |              |         |            |  |
| +-----                     |       |    |  |           |                  |              |         |            |  |

Start coding or [generate](#) with AI.