```python
from google.colab import userdata

API_URL = "https://api-inference.huggingface.co/models/microsoft/Phi-3-mini-4k-instruct"
headers = {"Authorization": userdata.get('AUTH')}

def query(payload):
  response = requests.post(API_URL, headers=headers, json=payload)
  return response.json()

def CustomChatGPT(question,rule):
    i=len(question)+23
    while question !="exit()":
        a=len("<|system|>\n"+rule+"<|user|>\n"+question+"\n <|assistant|>")
        output = query({
            "inputs": "<|system|>\n"+rule+"<|user|>\n"+question+"\n <|assistant|>",
        })
        output=(output[0].get("generated_text")[a+1:]+"\n")
        return output
        question=input("Question:")
        i=len(question)+23


demo = gradio.Interface(fn=CustomChatGPT, inputs = ["text", "text"], outputs = "text", title = "ChattyGPT")

demo.launch(share=True)
```

Colab notebook detected. To show errors in colab notebook, set debug=True in launch()
Running on public URL: https://dd993e35083f1647e3.gradio.live

This share link expires in 72 hours. For free permanent hosting and GPU upgrades, run `gradio deploy` from Terminal to d

ChattyGPT

question

what is a cat

rule

always answer in rhymes

| Clear | Submit |

output

A cat, with fur sleek and fine,
On silent paws and whiskers that twine.
Black as night with marmalade mirth,
Or striped as a tiger, born of earth.