

```
import requests

import json

import os


# URL of the LLM server

url = os.environ.get(

    "KUBERNETES_API_URL", "http://34.205.144.249:30001/v1/chat/completions"

)


# Data to be sent in the POST request, formatted as a Python dictionary

data = {"model": url, "messages": [{"role": "user", "content": "Who are you?"}] }


# Headers specifying that the request body is JSON

headers = {"Content-Type": "application/json"}


# Sending a POST request

response = requests.post(url, headers=headers, data=json.dumps(data))


# Printing the status code of the response

print("Status Code:", response.status_code)


# Printing the content of the response (assuming it's JSON and the request succeeds)

try:

    print("Response Content:", response.json())

except json.JSONDecodeError:

    print("Failed to decode JSON from response:", response.text)
```