

NLP POWERED CHAT ASSISTANT

Introduction

This paper presents the development of an advanced chatbot powered by the Llama-3-8B-Instruct-GGUF model, utilizing its 8 billion parameters for precise and context-aware responses.

The chatbot runs on a local server via LM Studio, enabling fine-tuning and efficient model interaction without relying on external APIs. The backend is built using Python and Flask, ensuring reliable performance and smooth integration.

Key features include,

- Dynamic thinking animations
- Typing indicators
- Chat history management
- Downloadable conversation logs
- Customizable themes

These enhancements create a fluid and engaging user experience. Designed for flexibility, the system supports real-time response handling and persistent sessions. Its architecture allows easy adaptation across various domains such as customer support, education, and productivity tools. This project showcases the practical use of large language models in building intelligent, responsive, and user-friendly chat assistants.

Screenshot of the work done

Local Inference Server

Start a local HTTP server that mimics select OpenAI API endpoints.

Supported endpoints: [/v1/chat/completions](#), [/v1/embeddings](#), [/v1/models](#)

Configuration

Server Port

Cross-Origin Resource Sharing ☒

Verbose Server Logs ☒

Apply Prompt Formatting ☒

Embedding Model Settings

[Learn More About Embeddings](#)

Load a test embedding model and utilize it using the [v1/embeddings](#) endpoint.

nomic-nomic-embed-text-v1.5

A fast high-performing text embedding model by Nomic AI (Apache-2.0)

[Download \(144 MiB\)](#)

examples hello world **chat (python)** assistant (python) vision embedding other

```
1 messages:
2   [{"role": "system", "content": "Always answer m."}
3   {"role": "user", "content": "Introduce urself."}]
4 ]
5 temperature=0.7
6
7 print(completion.choices[0].message)
```

[Copy Code](#)

Filter logs...

Server logs Server running on port 1234 using /tp/undo-server-log.txt

```
[2064-03-18 13:18:35] [INFO] [HLI SNDP] Versos=Server logs are ENABLED
[2064-63-18 13:18:35] [INFO] [HLI SNDP] Start new server logs on 1234
[2064-63-18 13:13:35] [INFO] [HLI SNDP] Available endpoints: /v1/chat/comprot/1234
[2064-03-18 13:18:35] [INFO] [HLI SNDP] Load ing on [pat:/7locelnost/neta-1lama-3-66-Suif-Intrucct-GQUF @
```

Figure 1 local server

Write me a poem in hindi about the beauty of nature

Here's a poem in hindi about the beauty of nature:

मेरी मा प्रकृति को सुंदरता है (meree maan prakriti ki sundarta hai)

प्राकृतिक सेरत बी सुंदन एता है
समर्हित के सैरद वी हिहेल्लता है
और चरो खौन तिकम पिस्नलता है
और महे पसे ब्रकुरी हा तिरशेवना सदती है

चेत नो प्रौ सघर हिदतो मेहे
पुमी की सदर हिरतो हने मंदुवाना
मौतिक हो सवर मे हन प्ररितो नबस्मे
और महे पस मनिचा हन शतित सदती

मेरी मां प्रकृति को सुंदरता है
मांकतिक से हाराता चर कि मुदतता
हत कि के सैषच मे हने नुरारात्त पधुं

translation:

my mother nature's beauty is

the peace and beauty of the natural world
on the banks of taves, we have nourished
the beautiful display of flowers
and the sun's rays illuminate

the branches of trees and plants heaven
beautifully displayed my flower

Type your message...

Clear Chat



Figure 2 Light mode

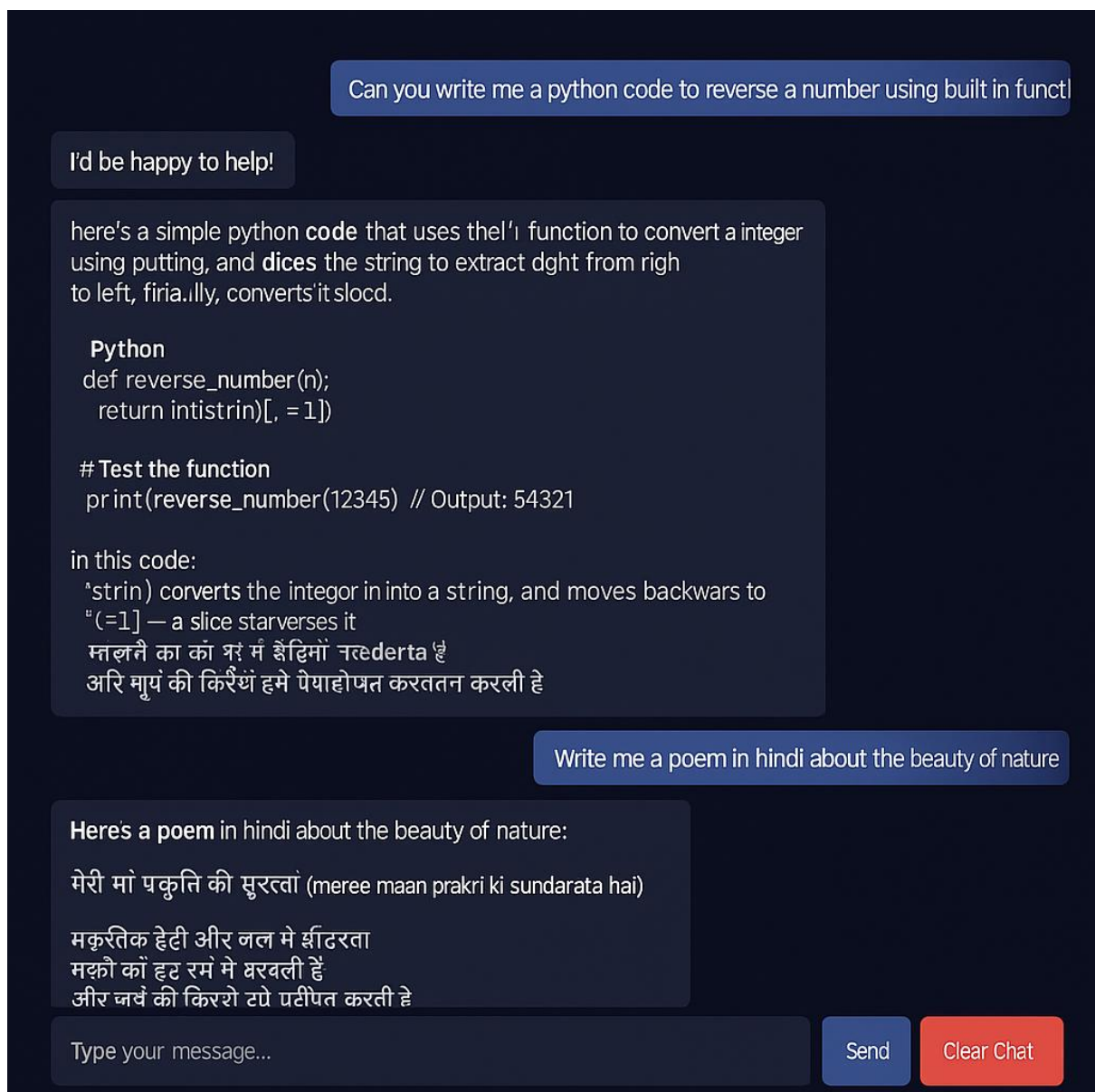


Figure 3 Dark mode

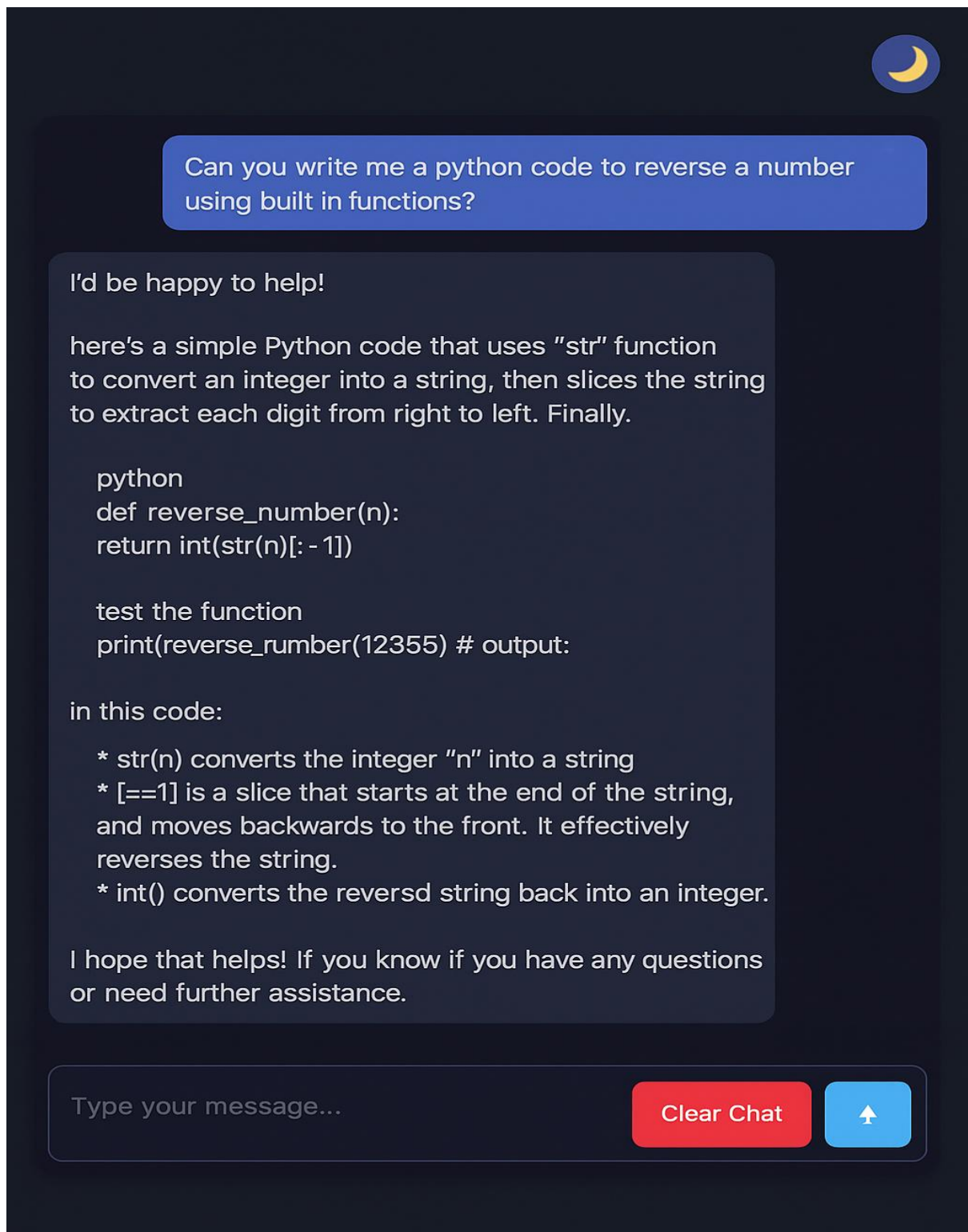


Figure 4 Input \$ Output