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

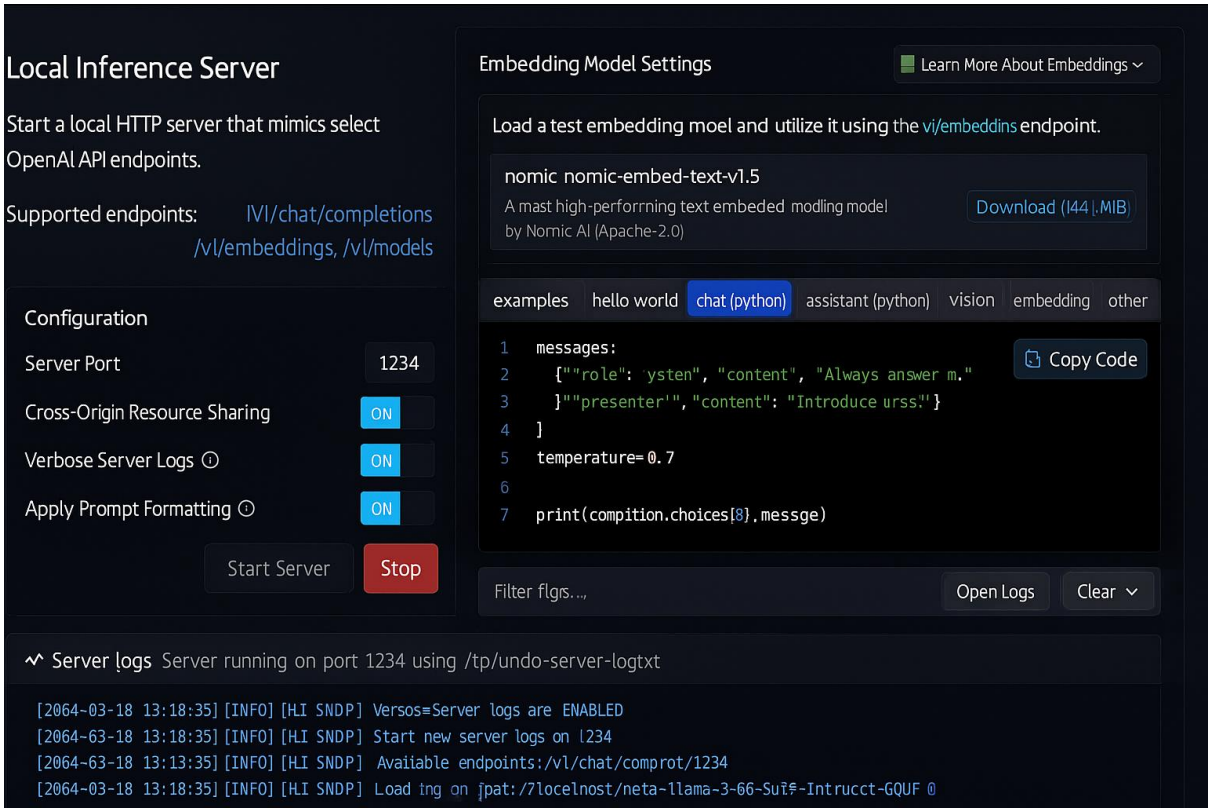

Figure 1 local server

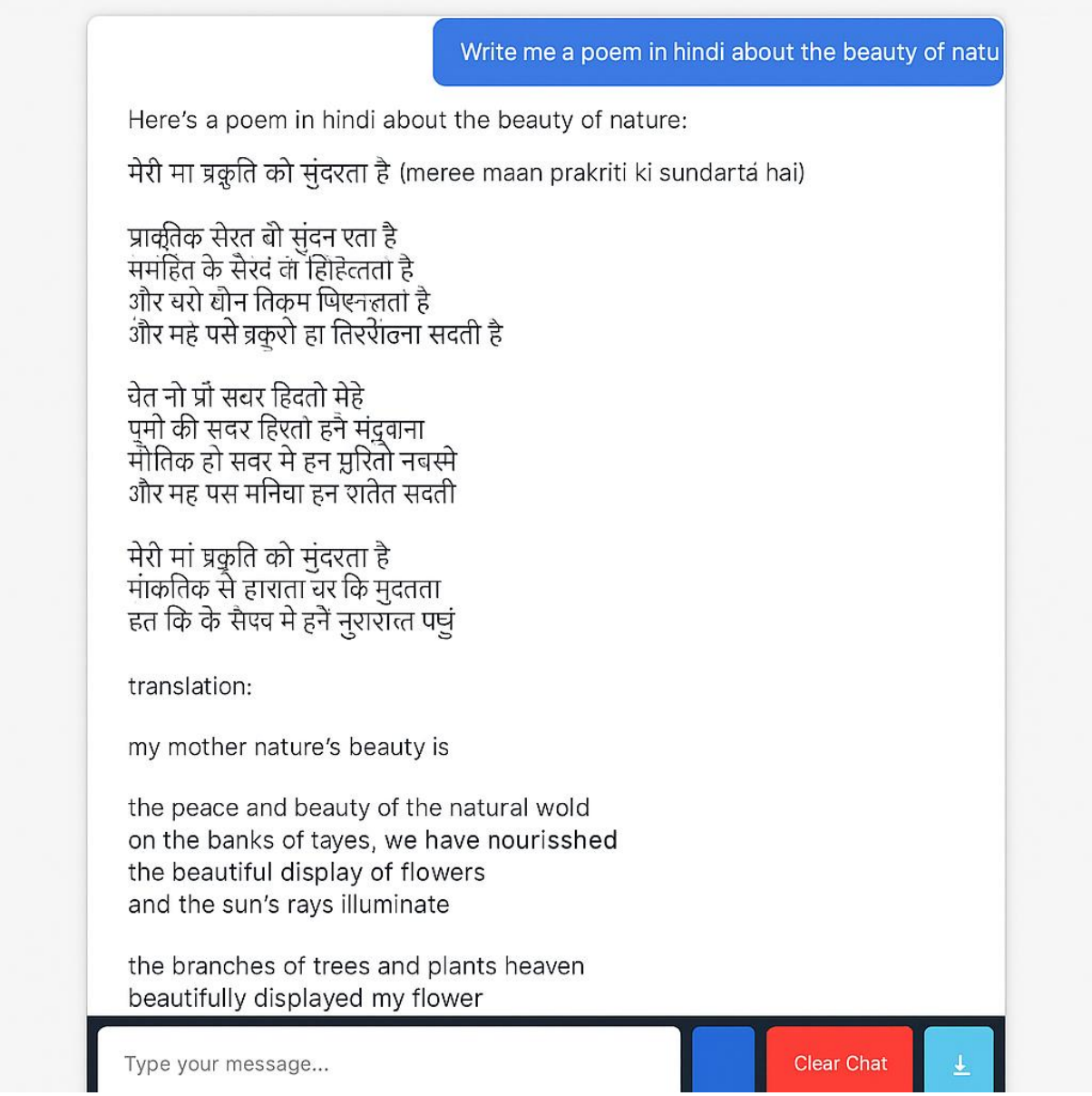

Figure 2 Light mode

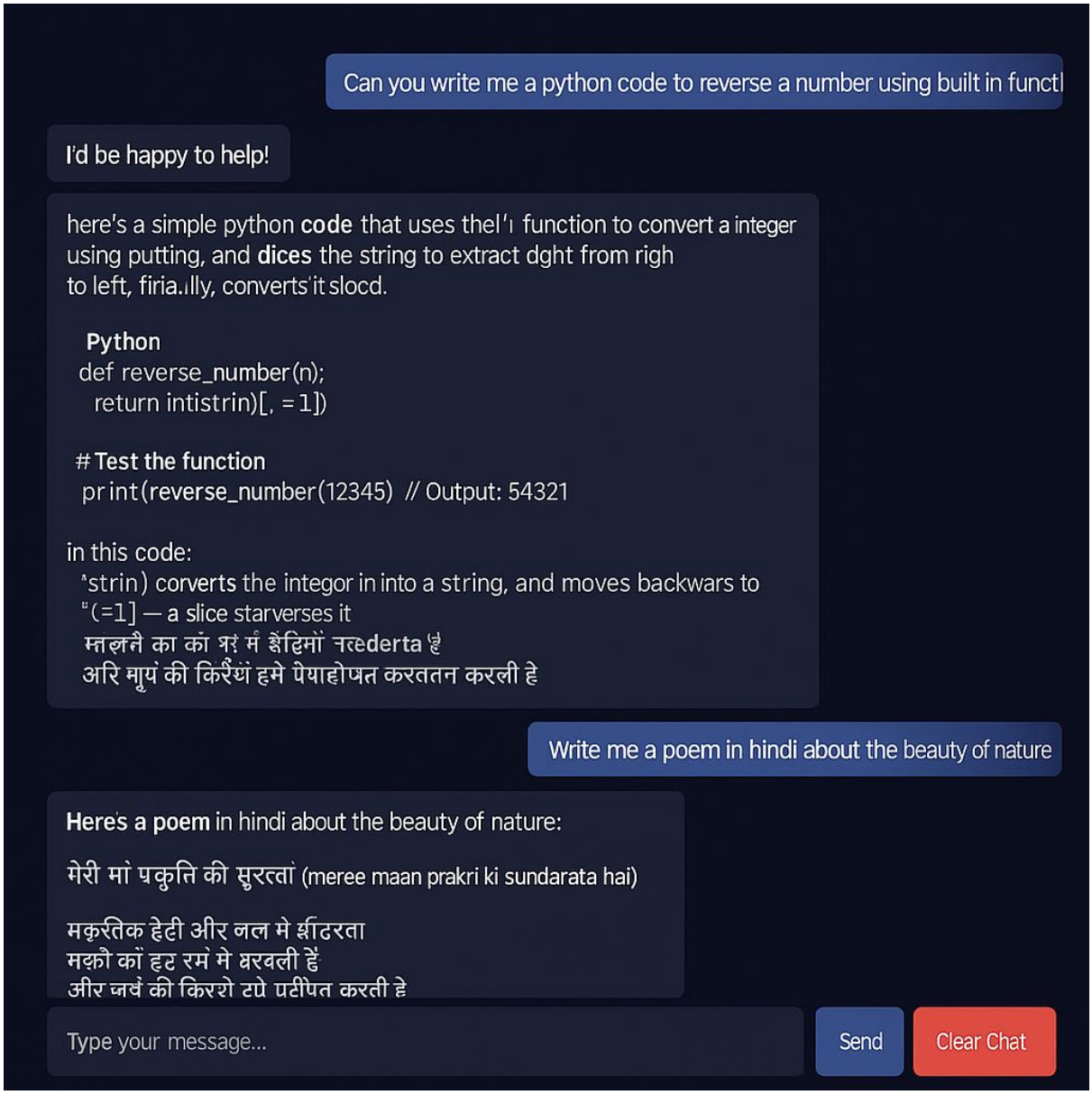

Figure 3 Dark mode

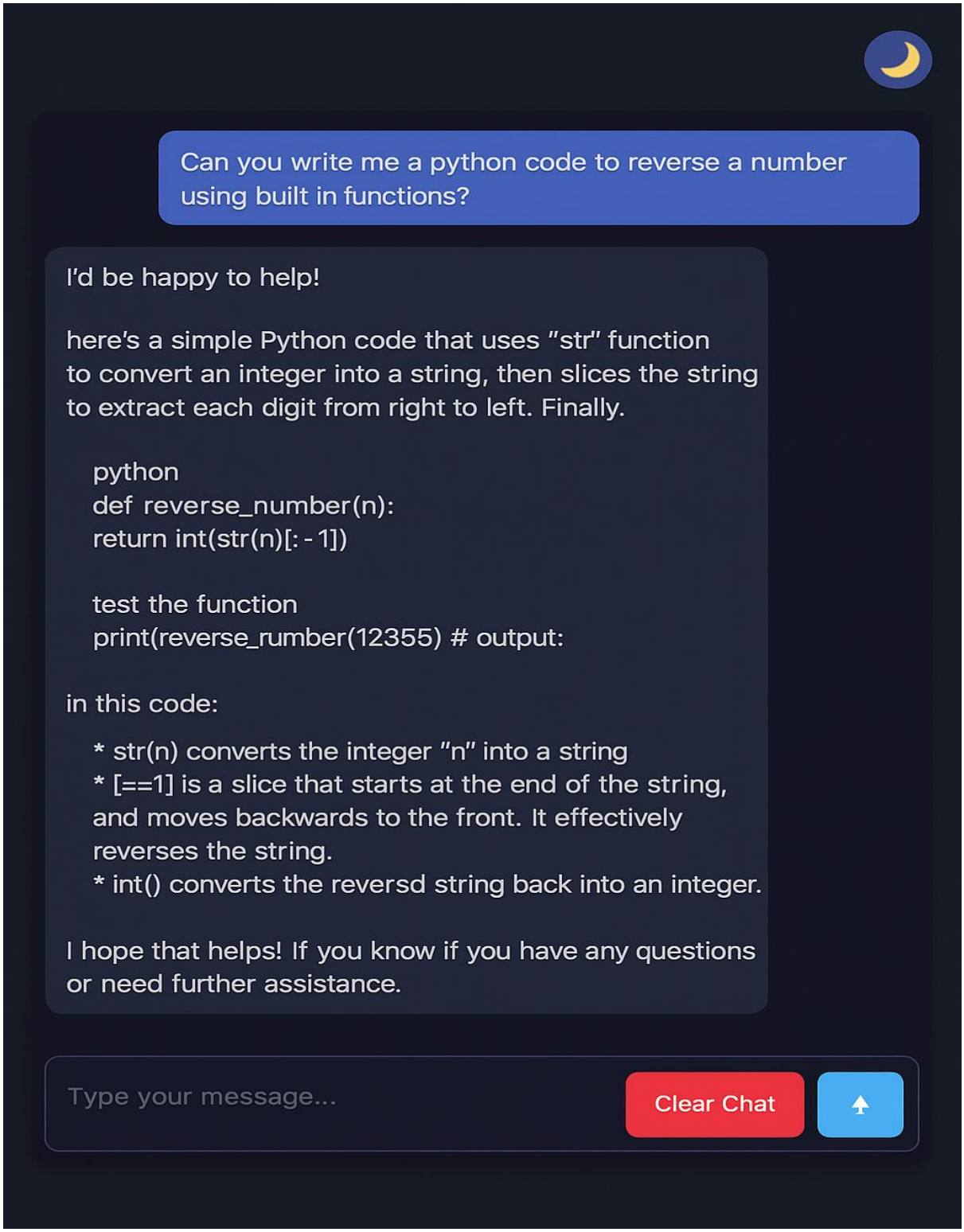

Figure 4 Input $ Output