

Ai Powered Story Generator Chat-bot



Team Members

Mohawiz Hamid (Team Lead).

Abdul Rehman.

Adbur Rafey.

Sulman Saeed.

Muhammad.

Abstract:

The development of creative chatbots is motivated by the necessity for natural, engaging, and efficient interactions in our increasingly digital world. This project, presented by Team Alpha NLP, focuses on creating an AI-powered story generation chatbot using advanced language models. Our chatbot leverages models such as GPT-2, GPT-3.5 Turbo, LLaMA 2, and the standout LLaMA 3 to deliver superior natural language understanding and generation capabilities. The project integrates a robust backend with an interactive UI using LangChain, GROQ API, TypeScript, React, Streamlit, and FastAPI. Fine-tuning and prompt engineering enhance performance, ensuring the chatbot generates high-quality stories based on user-defined parameters. The project's future steps include further personalization, multimedia integration, and cross-platform availability, aimed at enriching user engagement and interactivity.

Introduction:

Origin of the Creative Idea:

The impetus behind developing creative chatbots stems from the demand for more natural, engaging, and efficient interactions within digital environments. In today's world, where digital communication is ubiquitous, enhancing user experience through innovative AI technologies is paramount. By blending advanced AI capabilities with creative design, chatbots can transcend traditional boundaries, opening new avenues for user engagement and innovation.

Project Vision and Mission:

Our project aims to develop a fully functional and efficient chatbot that facilitates story generation, utilizing multiple parameters such as length, genre, and plot elements. The vision is to harness cutting-edge AI technologies to create a tool that not only meets current interaction needs but also pushes the boundaries of what is possible in digital storytelling.

Evaluation of LLM Models:

To ensure the highest efficiency and performance, we evaluated several top language models (LLMs), including GPT-2, GPT-3.5 Turbo, LLaMA 2, and LLaMA 3. Our extensive evaluation revealed that LLaMA 3 stands out due to its power, efficiency, and versatility, making it an excellent choice for superior natural language understanding and generation.

Meta Llama 3 Instruct model performance

| | Meta Llama 3 8B | Gemma 7B - It Measured | Mistral 7B Instruct Measured |
|--------------------|-----------------|------------------------|------------------------------|
| MMLU 5-shot | 68.4 | 53.3 | 58.4 |
| GPQA 0-shot | 34.2 | 21.4 | 26.3 |
| HumanEval 0-shot | 62.2 | 30.5 | 36.6 |
| GSM-8K 8-shot, CoT | 79.6 | 30.6 | 39.9 |
| MATH 4-shot, CoT | 30.0 | 12.2 | 11.0 |

| | Meta Llama 3 70B | Gemini Pro 1.5 Published | Claude 3 Sonnet Published |
|--------------------|------------------|--------------------------|---------------------------|
| MMLU 5-shot | 82.0 | 81.9 | 79.0 |
| GPQA 0-shot | 39.5 | 41.5 CoT | 38.5 CoT |
| HumanEval 0-shot | 81.7 | 71.9 | 73.0 |
| GSM-8K 8-shot, CoT | 93.0 | 91.7 11-shot | 92.3 0-shot |
| MATH 4-shot, CoT | 50.4 | 58.5 Minerva prompt | 40.5 |

Development Framework:

To streamline development and leverage large language model capabilities, we adopted the LangChain framework and GROQ API. This approach eliminates the need for downloading bulky pre-trained models, allowing seamless access to LLMs and saving on development time and disk space. The GROQ API also promises future scalability by integrating with state-of-the-art LLMs.



Enhanced Performance:

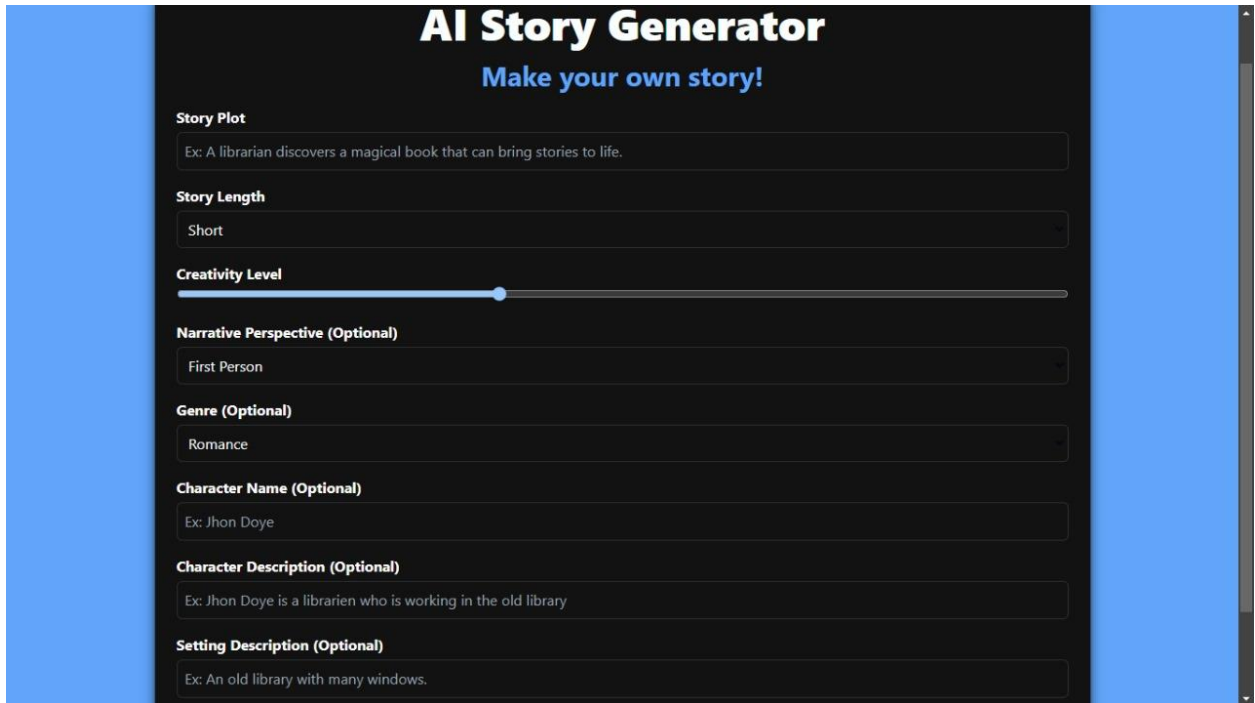
We enhanced the chatbot's performance through fine-tuning and prompt engineering. Fine-tuning involves adapting a pre-trained language model to a specific dataset, improving its performance for particular tasks. Prompt engineering involves designing and optimizing input prompts to guide the language model in generating the most relevant and accurate responses.

Interactive User Interface:

Our project incorporates two types of user interfaces to cater to different needs:

TypeScript/React UI:

This interface is designed for scalable, component-based web applications, offering robust architecture and state management.



The image shows a web application titled "AI Story Generator" with the subtitle "Make your own story!". The interface is dark-themed with blue accents. It features several input fields and a slider for customizing a story. The fields are labeled "Story Plot", "Story Length", "Creativity Level", "Narrative Perspective (Optional)", "Genre (Optional)", "Character Name (Optional)", "Character Description (Optional)", and "Setting Description (Optional)". Each field has a placeholder text starting with "Ex:". The "Creativity Level" is controlled by a horizontal slider. The interface is flanked by blue vertical bars on the left and right sides.

AI Story Generator
Make your own story!

Story Plot
Ex: A librarian discovers a magical book that can bring stories to life.

Story Length
Short

Creativity Level
A horizontal slider bar.

Narrative Perspective (Optional)
First Person

Genre (Optional)
Romance

Character Name (Optional)
Ex: Jhon Doye

Character Description (Optional)
Ex: Jhon Doye is a librarian who is working in the old library

Setting Description (Optional)
Ex: An old library with many windows.

Streamlet UI:

Tailored for data-driven applications in Python, this interface emphasizes simplicity and interactivity, suitable for users without extensive front-end development knowledge.

Options

Choose Task

Generate a New Story

Generation Options

Realistic Creative

0.00 0.50 1.00

Select length

Small

☒ Additional Options

Select genre

Fantasy

Narrative perspective

First-person

Character Name

Ex: Jane Doe

Character Description

Ex: Jane Doe is a timid 30-year-old librarian who has

Story Generation and Completion

Enter your story prompt:

Dog and Cat

Generate Story

Deploy

Backend Integration:

The backend, managed by FastAPI, links our TypeScript/React frontend to the Python backend where our model resides. FastAPI handles requests from the frontend, ensuring smooth data flow, while TypeScript and React maintain code safety and efficiency. This architecture enables a fast, scalable, and maintainable web application suitable for real-time updates and complex data tasks.

Future Steps:

To further enhance our storytelling chatbot, we plan to:

- Improve its natural language understanding.
- Personalize interactions based on user preferences.
- Integrate multimedia elements like pictures and videos.
- Create dynamic stories based on user input.
- Make the chatbot available on multiple platforms.

These steps aim to make the chatbot more engaging, user-friendly, and adaptable to users' evolving needs.

Conclusion:

Our AI-powered story generation chatbot represents a significant step forward in digital storytelling, combining advanced AI with creative design. The ongoing improvements and future plans will ensure it remains a cutting-edge tool for engaging and interactive user experiences.