

Project 6: An Interactive Human-in-the-Loop Collaborative Storytelling using Retrieval Augmented Generation

Project Description

This project aims to create an "Interactive Human-in-the-Loop Collaborative Storytelling" system where users and a Large Language Model (LLM) co-develop stories interactively. The system will dynamically adapt story suggestions based on user input, providing contextually relevant ideas, prompts, and text completions to enhance creativity and maintain narrative coherence. Using a Retrieval-Augmented Generation (RAG) approach, the system will retrieve similar story fragments from a large text database to ground generated suggestions in relevant contexts. Users can provide feedback on the model's suggestions to guide and refine the storytelling experience.

Note: You have the flexibility to select the models and datasets that align best with your project's goals. The steps and tools outlined here are merely suggestions to guide you; feel free to tailor, modify, or expand upon these ideas to make the project uniquely your own.

Project Goals

1. **Human-AI Collaboration:** Facilitate collaborative storytelling where the user and the LLM work together to create engaging, imaginative narratives.
2. **Retrieval-Augmented Generation (RAG):** Integrate retrieval-based methods to pull relevant story snippets, enhancing the quality of generated suggestions.
3. **Interactive Feedback:** Allow users to interactively rate or modify the LLM's suggestions to refine content alignment with their creative goals.
4. **Creative Skill Building:** Encourage users to explore narrative structures, character development, and imaginative storytelling through interaction with the LLM.

Implementation Resources

1. Datasets

- **Primary Dataset:** Use a curated dataset of short stories or literary texts to provide rich content for the retrieval system. **Options** include:
 - **BookCorpus** (via Hugging Face): Contains a large collection of books, ideal for diverse storytelling themes.
 - **ROC Stories Dataset** ([Link](#)): Short, five-sentence stories useful for structuring narrative coherence.
 - **FairytaleQA** ([Link](#)): Contains 278 children-friendly fairytale narratives.
- **Preprocessing:** Tokenize, clean, and segment the data into manageable chunks for retrieval.

2. Technologies and Tools

- **LLM for Generation:**
 - **Example** models include **GPT-4** or **GPT-3.5** (via OpenAI API) for high-quality text generation. You can also use Mistral, LLaMa-3, ...
- **Retrieval Model:**
 - **Options** include **Dense Passage Retrieval (DPR)** via Hugging Face for embedding passages and retrieving contextually similar content.

- **FAISS** (Facebook AI Similarity Search): Efficient similarity search across a large number of passages, enabling fast retrieval from the story dataset.
- **RAG Model Integration:**
 - **Retrieval-Augmented Generation (RAG)** pipeline to combine retrieval and generation components effectively.
- **Interactive Interface: Options** include:
 - **Streamlit:** An intuitive, Python-based framework for creating a web app to enable user interaction.
 - **Gradio:** Another interactive interface that allows users to provide input, rate generated content, and see story progressions in real-time.
- **RAG Pipeline:**
 - **Retrieval Step:** User inputs a prompt (e.g., an initial story sentence or character setup), and the system retrieves related story passages using **DPR** and **FAISS**.
 - **Generation Step:** The LLM receives the retrieved context and generates the next part of the story. The prompt is refined based on user feedback, and further story suggestions are retrieved as needed.
- **Feedback Loop (Optional but impactful):**
 - **User Feedback Mechanism:** Implement a simple thumbs-up/thumbs-down or slider for user feedback on suggestions.
 - **Adaptive Model Adjustment:** Track feedback in real-time and adjust generation preferences based on user responses (e.g., using a “rating score” to bias prompts for more aligned results).