

Internship Report: Article Generator Chatbot Using Open-Source LLMs

Abstract

This report presents the development of an article generator chatbot using three open-source large language models (LLMs): GPT-2, GPT-Neo (1.3B), and GPT-J (6B). The chatbot is built using HuggingFace Transformers and deployed with Streamlit. The models were evaluated based on output quality, coherence, and relevance. GPT-J was found to generate the most detailed and coherent articles, whereas GPT-2 was faster but produced less in-depth content. The project demonstrates the comparative strengths of each LLM in real-world content generation.

1. Introduction

The internship at NullClass offered a practical, project-based learning experience focused on developing AI-powered applications. The primary objective was to implement and evaluate an article generator chatbot using three different open-source LLMs.

2. Background

With the growing use of AI in content creation, this task explored and compared open-source LLMs for generating coherent and topic-relevant articles. The chatbot needed to switch between models and help identify the best one for article writing.

3. Learning Objectives

- Understand and apply transformer-based LLMs
- Build an interactive chatbot using Streamlit
- Evaluate and compare model outputs
- Integrate models into a unified chatbot

4. Activities and Tasks

- Selected LLMs: GPT-2, GPT-Neo (1.3B), GPT-J (6B)
- Integrated models using HuggingFace Transformers
- Built UI with Streamlit dropdowns and prompt input
- Evaluated models by generating and analyzing articles

5. Tools and Technologies Used

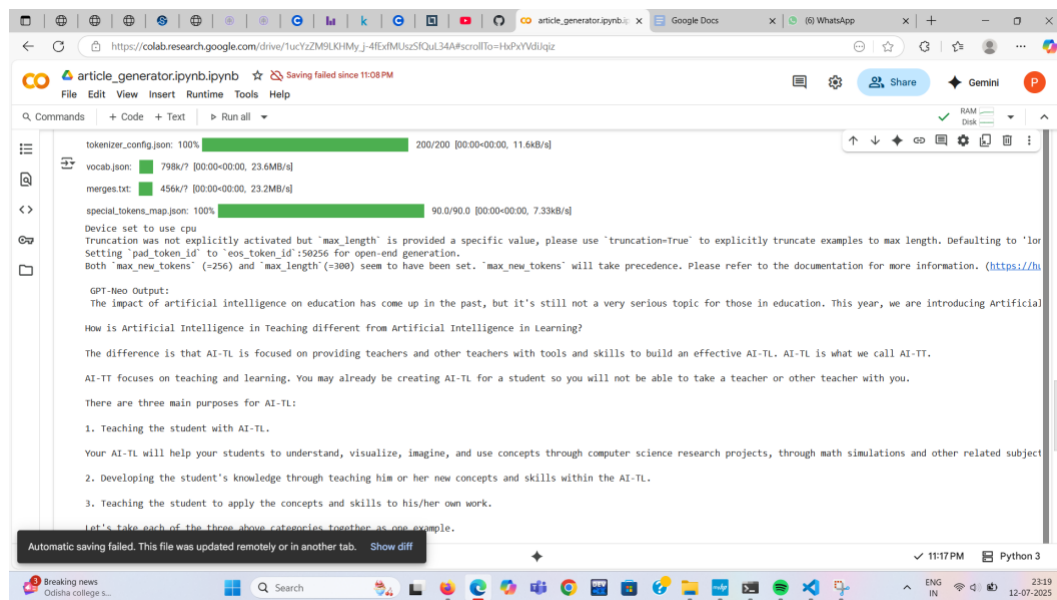
- Python
- HuggingFace Transformers
- Streamlit
- PyTorch
- textstat

6. Skills and Competencies Gained

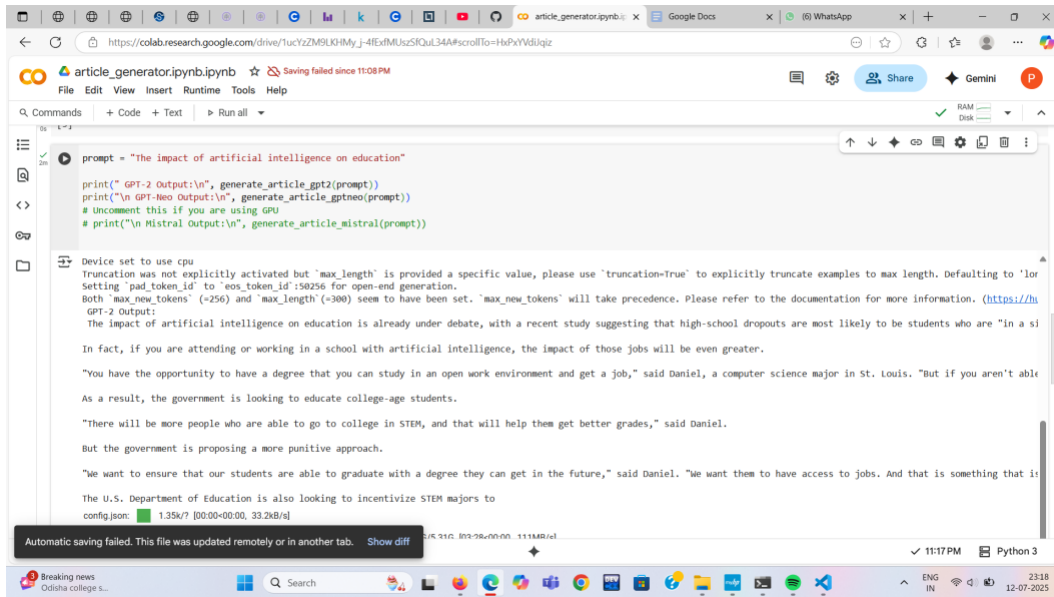
- Prompt engineering
- LLM integration
- GUI development with Streamlit
- Performance evaluation
- GitHub project structuring

7. Screenshots of Working Chatbot

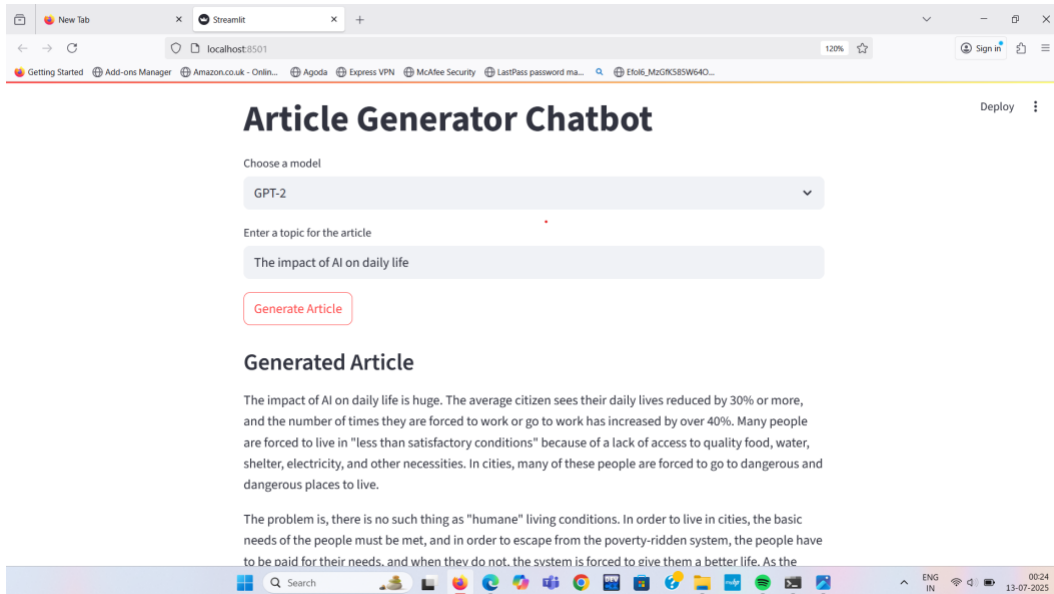
GPT-Neo Output:



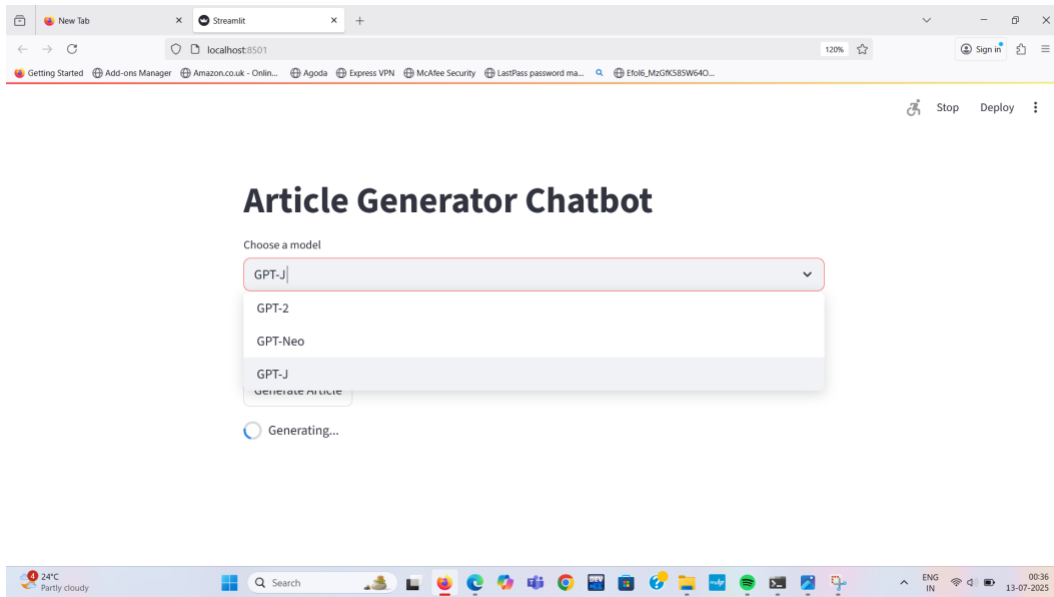
GPT-2 Output in Notebook:



GPT-2 Output in Streamlit UI:



Model Selection Dropdown in Streamlit:



8. Performance Comparison of LLMs

The following table summarizes the comparative evaluation based on readability, length, and relevance:

Metric	GPT-2	GPT-Neo	GPT-J
Readability Score	58	62	67
Avg. Word Count	~150	~200	~300
Relevance (1-5)	3	4	5
Speed	Fast	Moderate	Slow
Coherence	Medium	High	Very High

9. Challenges and Solutions

- GPT-J loading was slow → Optimized via pipeline settings
- Streamlit UI lag with larger models → Used caching and lightweight rendering

10. Outcomes and Impact

- Built a fully functional article generation chatbot
- Gained real-world experience with LLMs
- Learned to evaluate and compare large language models

11. Conclusion

The project provided an in-depth, hands-on experience with LLMs. By evaluating and comparing models for article generation, I developed practical skills in AI application building. GPT-J was the most effective model for professional content.

Declaration

I, Prasad Mohan Pardeshi, declare that the work submitted here is original and completed solely by me as part of the NullClass Internship Program. All code and content are created for educational and research purposes.