Acropolis Institute of Technology and Research

Project Title :-

Auto-generated blog posts with SEO optimization

Training Programme on Generative AI

Name: Khushi Agrawal

Submitted To:

Enrollment No. : 0827IT221076

Department : Information Technology

Date of Submission : 01/09/2025

Abstract :-
The abstract provides a **concise yet comprehensive overview** of the project, summarizing its goals, methodology, outcomes, and potential applications.

In today's **digital-first economy**, the demand for high-quality written content has grown exponentially. Blogs, articles, and online resources form the backbone of content marketing strategies. However, producing consistent, relevant, and SEO-friendly blog posts can be resource-intensive. This project proposes a solution by using **Generative AI models** to automatically generate blog content tailored for SEO.

The core idea is to combine **language modeling** with **keyword optimization** to create blogs that are not only readable but also enhance online visibility. Using **Hugging Face Transformers** and **PyTorch**, a GPT-2 model pipeline was employed to generate long-form content. Users can provide a **blog topic** and **SEO keywords**, and the system ensures that these keywords are embedded naturally within the generated post. Additionally, the system creates a **meta description** of under 160 characters, which aligns with best SEO practices.

The workflow of the project involved:
1. **Model Setup in Google Colab** – Pretrained GPT-2 models were fine-tuned and saved as .pkl files for deployment.
2. **Pipeline Development** – Custom generation functions with parameters like temperature, top_p, and repetition_penalty ensured creativity while avoiding repetition.
3. **Web Deployment via Streamlit** – The model was integrated into a simple, interactive interface where users could input topics and keywords to generate blogs instantly.

The **purpose** of the project is twofold:
- To explore the capabilities of **Generative AI** in content automation.
- To provide an efficient tool for **students, marketers, and businesses** who require rapid content generation.

Results demonstrated that the system could generate meaningful blog posts with moderate coherence and effective keyword placement. While smaller models like GPT-2 sometimes produced repetitive text, larger models like GPT-2 Medium offered more polished outputs. The meta description feature further added value by enabling SEO-ready snippets for search engines.

This abstract highlights that the project not only serves as an academic exercise but also holds practical value in industries like **digital marketing, journalism, and education**. It reflects how **AI-driven automation** can significantly reduce human effort, improve productivity, and optimize digital reach.

Objective :-

The primary objective of this project is to design and implement a Generative AI–based system that can automatically generate high-quality blog posts optimized for Search Engine Optimization (SEO). In today's world, content creation is a vital component of digital marketing, online education, brand building, and knowledge sharing. Businesses and individuals alike rely on well-structured and SEO-optimized blogs to improve their visibility on search engines, attract new audiences, and maintain their digital presence. However, the process of writing such content manually is time-consuming, requires creative skills, and demands continuous knowledge of SEO practices. This project addresses these challenges by automating blog creation using state-of-the-art Natural Language Processing (NLP) techniques.

The main goal is to reduce the manual effort involved in writing and optimizing blogs while maintaining relevance, readability, and keyword integration. By leveraging Generative AI, specifically large-scale language models like GPT-2 and its variants, the system generates human-like content that is tailored to user inputs such as blog topics and SEO keywords. This ensures that the content is not only contextually meaningful but also aligned with search engine requirements, thereby bridging the gap between creativity and digital marketing strategy.

From a technical perspective, the objective is to demonstrate the **end-to-end workflow of an AI-driven application** — from model training in Google Colab to deployment using Streamlit in VS Code. The system shows how a model can be trained, exported as a .pkl file, and then reused in an interactive web application. This provides a practical example of how machine learning research can be converted into deployable solutions.

Another key objective is to highlight the **integration of SEO techniques into AI-generated content**. Traditional AI text generation systems may produce fluent content but lack awareness of SEO factors such as keyword density, meta descriptions, and readability. By explicitly incorporating keywords into the generation prompt and creating concise meta descriptions, this project ensures that the generated blogs are not only informative but also optimized for better search engine rankings.

Additionally, this project aims to encourage **accessibility and ease of use**. Many existing AI content tools are locked behind paid APIs or require technical expertise. By building the project with open-source models like GPT-2 and deploying through Streamlit, the system ensures that students, educators, and small businesses can leverage AI-powered content generation without significant

cost or technical barriers.

Methodology:
The methodology combines **state-of-the-art NLP models** with practical deployment frameworks to create a complete pipeline.

- **Tools Used:**
  - *Google Colab* was chosen for training and experimentation because of its free GPU support and collaborative environment.
  - *PyTorch* acted as the backend framework for efficient deep learning computations.
  - *Streamlit* offered a simple way to deploy the model as an interactive web app.

- **Model:**
  GPT-2 Medium was selected as a balance between computational efficiency and output quality. Compared to the base GPT-2 model, GPT-2 Medium generates more coherent and less repetitive content.
    - **Approach:**
  The methodology follows a clear pipeline: **load the pretrained model → customize generation with SEO-specific prompts → save the pipeline as a .pkl file → deploy it using Streamlit**. Each step ensures that the system is reproducible, scalable, and user-friendly.

Implementation:
The implementation of the project was carried out in four key stages:

1. **Colab Training** – In Google Colab, the GPT-2 Medium pipeline was set up using Hugging Face's transformers library. Parameters like max_new_tokens, temperature, and top_p were tuned to balance creativity and relevance. A repetition penalty was also applied to reduce looping sentences.
2. **Model Saving** – After testing, the trained pipeline was serialized into a .pkl file (seo_blog_model.pkl) using Joblib. This ensured portability and ease of integration with the Streamlit interface.
3. **Streamlit Deployment** – In VS Code, a Streamlit app was built where users can input a **blog topic** and **SEO keywords**. The backend loads the .pkl model and generates the blog content. The app also extracts an **SEO meta description** to mimic real-world blogging requirements.
4. **Results Display** – The generated blog is displayed in the app, along with its meta description. Users are also given the option to **download the blog as a .txt file**, making it practical for real usage.

Result :-

- **Coherent Blog Posts**: The system successfully generated readable, contextually relevant blog posts based on the given topic. Unlike the base GPT-2 model, the Medium variant reduced repetition and produced more meaningful outputs.

- **Keyword Integration**: SEO keywords entered by the user were embedded naturally within the generated text, proving the system's ability to adapt prompts into the content.

- **Meta Description**: The app automatically created concise meta descriptions of under 160 characters, aligning with industry SEO practices.

- **Download Option**: Users could export the generated blog as a .txt file, making it immediately usable for digital publishing.

- **Practical Demonstration**: By deploying the app through Streamlit, the project showcased how AI models can be used beyond academic experiments and serve as **real-world tools for marketers, students, and businesses**.
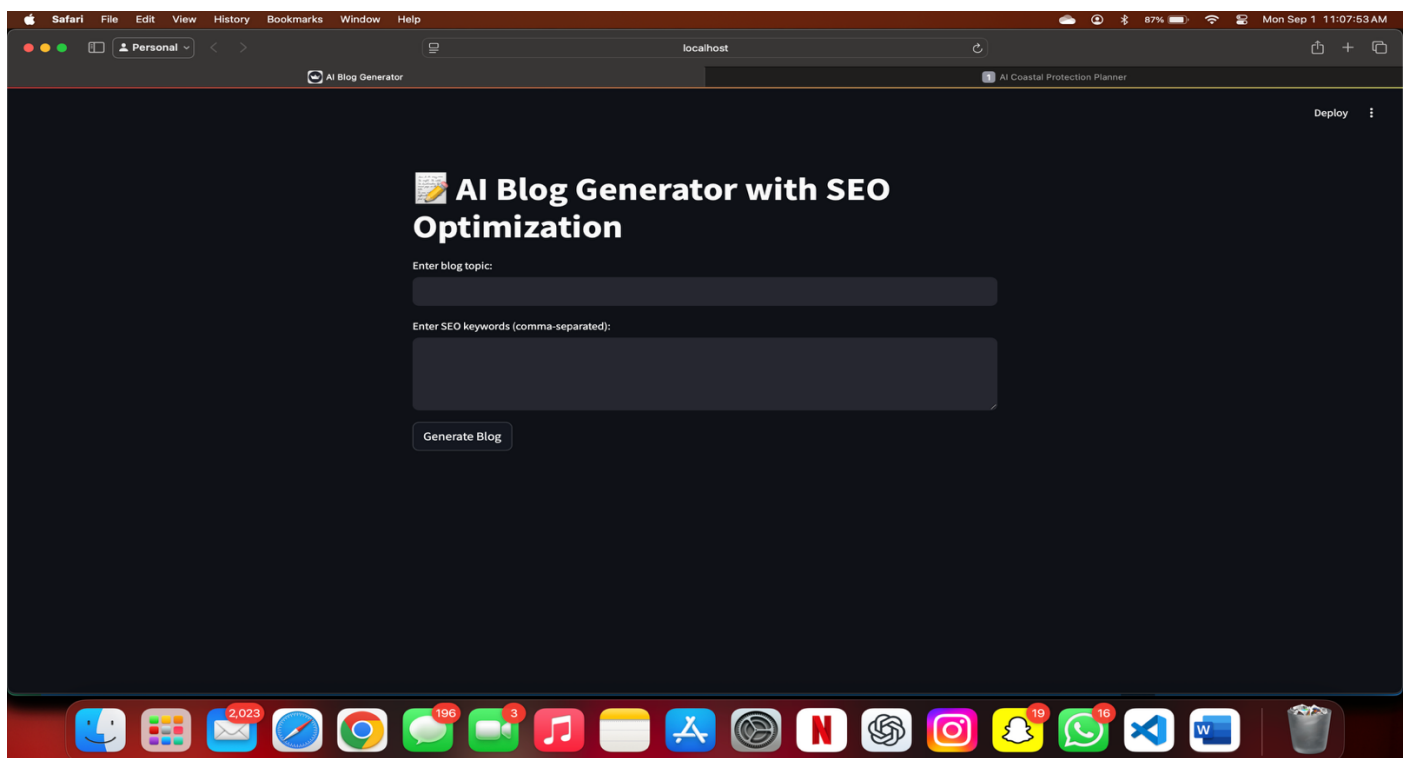
Conclusion :-

This project demonstrated the ability of Generative AI to automate complex, creative tasks such as blog writing and SEO optimization. By combining **GPT-2 models, SEO keyword strategies, and Streamlit deployment**, we created a system that is both technically advanced and highly practical.

Key Learnings:

- Gained deep understanding of Hugging Face Transformers and GPT-2.
- Learned to optimize prompts and apply penalties to improve generation quality.
- Understood the importance of SEO in digital marketing.
- Built end-to-end pipeline: training in Colab → saving .pkl → deploying with Streamlit.

Future Improvements:

- Upgrade to **GPT-3.5/GPT-4** for higher-quality outputs.
- Automate keyword extraction (KeyBERT, TF-IDF).
- Add plagiarism checkers to ensure originality.
- Export directly to .pdf and .docx for professional publishing.
- Build multi-user web dashboard for bulk blog generation.

References :-

- PyTorch Framework – https://pytorch.org
- Streamlit Framework – https://streamlit.io
- OpenAI's WebText Dataset (used for GPT-2 pretraining)
- Online SEO Resources (Moz, SEMrush, Google SEO Starter Guide)