PROJECT REPORT

PREPARED FOR

Vishwareena Vanoo

PREPARED BY

Nicholas Tan

EXECUTIVE SUMMARY

This report outlines a project aimed at developing an AI-based web page content generation system. The system utilizes advanced machine learning algorithms to generate customized, search engine optimized web pages based on user input. It identifies relevant industry, target audience and long-tail keywords of the product/service.

Technical challenges faced during the project included ensuring the AI model's understanding and correct interpretation of user input, guaranteeing the quality and consistency of the generated content, and addressing scalability issues as the system must cater to multiple users with large volumes of content requests. These challenges were tackled through planning and continuous refinement and testing.

Price analysis shows the cost of each stage of an average run, with the total price calculated based on the number of tokens processed by the model. The total cost for an average run was found to be $0.008652, based on the GPT 3.5 Turbo Model's price of $0.002 per 1k tokens.

The project was structured into several milestones, including analysis, development, testing, and documentation stages. Each milestone encompassed specific tasks necessary for the project's success, with iterative testing and optimization to ensure the system's performance and functionality.

1. Project Overview

The primary focus of this project was to develop an AI-based web page content generation system. The system is designed to provide users with suitable content in a webpage format based on user details and keywords. By leveraging the capabilities of advanced machine learning algorithms, the system finds similar long-tail keywords, the target audiences, industry of product/service and generates a customized web page with unique content using AI to optimize and enhance their search visibility on the search engine.

2. Technical Obstacles

While developing an AI-based content generation system, there are several technical challenges that may need to be addressed:

Natural Language Understanding: Ensuring that the AI model truly understands and correctly interprets the user input is a significant challenge. This includes understanding the context, nuances, and subtleties of language, which can be difficult to achieve.

Quality of Generated Content: The quality of the content generated by the AI model is another significant challenge. The content should not only be grammatically correct and coherent but also engaging, relevant, and valuable to the reader. Ensuring this level of quality consistently can be difficult.

Consistency and Accuracy: The consistency of the AI generated content is tricky at times as the results generated might differ for every run. For example, the generated web page would be incorrectly styled or contains no styling. Therefore, another stage of prompt engineering is needed to refine the consistency of the results.

Scalability: As the system needs to cater to multiple users with possibly large volumes of content generation requests, scalability can be a significant issue. The system needs to be designed to handle growing amounts of work in a capable manner.

These challenges require careful consideration and planning, as well as continuous refinement and testing, to ensure the system functions as intended and provides value to its users.

3. Cost Analysis

GPT-3.5-Turbo was used in this project, table below shows the cost analysis of each stage of an average run:

Stage	Tasks	Prompt Tokens	Completion Tokens	Total Tokens	Price ($)
1	Industry Search	77	4	81	0.000162
2	Target Audience Search	80	73	153	0.000306
3	Keyword Clusters Search	90	15	105	0.00021
4	Titles Generation(5)	90	15	105	0.00021
5	Content Generation(5)	268	409	677	0.001354
6	HTML Conversion(5)	530	526	1056	0.002112
7	Styles and Components(5)	778	1374	2149	0.004298
Total				4326	0.008652

Price for GPT 3.5 Turbo Model: $0.002 / 1k tokens

4. Software

In the provided script, the following software and libraries are used:

1. **Python**: The primary programming language for this project. All methods and integration are written in Python

2. **OpenAI GPT-3 API**: This is the core AI model used for keyword find, content generation and web page conversion

3. **Threading**: This module is used to multiprocess the content generation, website

4. **re**: This is a Python built-in module for working with regular expressions. It's used in this script for tasks like removing unwanted characters from strings.

5. **dotenv**: This is a third-party Python library used for loading environment variables from a .env file. In this script, it's used to load the OpenAI API key.

5. Key Achievements

Automated Content Generation: The AI-based system successfully automates the process of content generation. It takes as input user details and relevant keywords, and outputs a fully formed webpage complete with unique, SEO-optimized content.

SEO Optimization: A significant feature of the system is its ability to incorporate SEO strategies into the content creation process. By intelligently using keywords throughout the generated content, the system helps improve the webpage's visibility on search engines.

Integration: Users can easily input their details and desired keywords, and receive search-engine-optimized keywords, content and a fully generated web page in return.

6. Milestones and Reporting

Milestone	Tasks
1 - Analysis	
1.1	Analysis and design stage, gather data
1.2	Architecture design
2 - Development	
2.1	Create codebase
2.2	Keyword and Industry Identification
2.3	Keyword Clusters and Title Generation
2.4	Multi-threaded Content Generation
2.5	HTML Content Generation
2.6	Adding Styles and Components
3 - Testing	
3.1	Testing Industry, Keyword Clusters and Titles Generation
3.2	Testing Content Generation
3.3	Testing HTML Generation
3.4	Testing Style and Components Generation
3.5	Optimization and refinement
4 - Documentation and Final Review	
4.1	Documentation and review of the code

Links:

https://github.com/Nick-1357/SEO-Content-Generation/tree/main