DLOps Project Proposal

1. **Title**: Implementation of GPT-3 for Hate Speech Detection and Deployment as a Web Application

2. **Group Members:**
   I. **Ravi Shankar Kumar (Roll No. M21AIE247)**
   II. **Debonil Ghosh (Roll No. M21AIE225)**
   III. **Saurav Chowdhury (Roll No. M21AIE256)**

3. **Introduction:**

   Hate speech is a growing problem on social media platforms, and it can have serious consequences for marginalized groups. In recent years, there has been a lot of interest in using machine learning techniques to detect hate speech automatically. In this project, we propose to implement the code from the research paper "Detecting Hate Speech with GPT-3" by Ke-Li Chiu, Annie Collins, and Rohan Alexander. We will use OpenAI's GPT-3 language model to identify sexist and racist text passages with zero-shot, one-shot, and few-shot learning.

4. **Related Work:**

   There have been several studies on using machine learning techniques for hate speech detection. Some studies have used traditional machine learning algorithms such as SVMs and Naive Bayes classifiers, while others have used deep learning models such as CNNs and LSTMs. However, few studies have explored the use of large language models like GPT-3 for this task.

5. **Project Plan:**
   A. **Data Collection:** We will collect a dataset of text passages that contain sexist or racist content.

   B. **Pre-processing:** We will pre-process the data by removing stop words, stemming/lemmatizing words, and converting all text to lowercase.

   C. **Model Training:** We will train the GPT-3 model using zero-shot, one-shot, and few-shot learning techniques.

   D. **Model Evaluation:** We will evaluate the performance of the model using metrics such as accuracy, precision, recall, and F1 score.

   E. **Web Application Development:** We will develop a web application that allows users to input text and get a prediction of whether it contains sexist or racist content.

   F. **Deployment**: We will deploy the web application on a free hosting site such as Heroku or PythonAnywhere.

6. **Expected Outcome:**

   We expect to develop a web application that can accurately detect sexist and racist content in text passages using GPT-3. The application will be user-friendly and accessible to anyone with an internet connection. We hope that this project will contribute to the ongoing efforts to combat hate speech online and promote more inclusive communication.