**FAKE NEWS GENERATOR & DETECTOR USING GENERATIVE AI AND NLP**

Submitted by:

Muskan Varshney

Under the supervision of:

Mr. Gaurav Singh

# Abstract

In today’s digital era, the rapid spread of misinformation poses a serious threat to public opinion and democracy. This project aims to tackle this issue using Generative AI and Natural Language Processing (NLP). It involves building two models — one that generates fake news using GPT-2 and another that detects fake news using a machine learning classifier. This dual approach not only helps understand how fake news can be generated, but also equips us with the tools to detect it effectively.

# Introduction

Fake news has become a critical problem with the rise of social media and digital news consumption. The ability to detect fake news automatically is vital in maintaining information integrity. This project explores both the creation and detection of fake news using advanced NLP models. The generator is based on OpenAI’s GPT-2 model, while the detector uses TF-IDF vectorization and Logistic Regression.

# Problem Statement

To develop a system that can generate and detect fake news articles using state-of-the-art Natural Language Processing and Machine Learning techniques.

# Technology Stack

- Python  
- Jupyter Notebook  
- Transformers (HuggingFace)  
- Scikit-learn  
- Pandas, Numpy, Matplotlib  
- TF-IDF Vectorizer  
- GPT-2 (Generative Pre-trained Transformer 2)

# Dataset Overview

The fake news detection model is trained using a publicly available dataset containing labeled real and fake news articles. The dataset is cleaned, vectorized using TF-IDF, and split into training and test sets. It includes fields such as title, text, and label.

# Project Workflow

1. \*\*Fake News Generator\*\* using GPT-2

- Load the pre-trained GPT-2 model  
 - Prompt the model with a topic or headline  
 - Generate a fake news article using language generation

2. \*\*Fake News Detector\*\* using Logistic Regression

- Load dataset  
 - Preprocess and clean the data  
 - Apply TF-IDF Vectorizer  
 - Train Logistic Regression model  
 - Evaluate using accuracy, classification report, and confusion matrix

# Implementation Steps

1. Set up virtual environment and install required packages  
2. Build the news generator using HuggingFace GPT-2  
3. Load and process the fake news dataset  
4. Vectorize text using TF-IDF  
5. Train a Logistic Regression classifier  
6. Evaluate performance metrics  
7. Test the detector with custom inputs

# Results

The fake news detector achieved an accuracy of approximately 90.8% on the test dataset. The classification report and confusion matrix further confirm the model’s reliability in distinguishing between real and fake news articles.

# Screenshots

Below are placeholders for output screenshots from the project:

[Insert Screenshot of Fake News Generator Output]

[Insert Screenshot of Fake News Detector Accuracy and Confusion Matrix]

[Insert Screenshot of Final Prediction Result]

# Conclusion

This project successfully demonstrates the use of Generative AI and NLP in both generating and detecting fake news. The combination of GPT-2 for generation and TF-IDF with Logistic Regression for detection provides an effective framework to tackle misinformation in online media. Future improvements could include integrating deep learning classifiers for even higher accuracy.