

1. Project Goal/Motivation (10%)

- Problem Statement: The project aims to develop a machine learning model to classify news articles and social media posts as real or fake. This capability is crucial for combating the rapid dissemination of misinformation online, which can influence public opinion and affect democratic processes.
- Motivation Behind the Project: The spread of fake news poses significant challenges by misleading the public and distorting truth. Developing a model to detect such falsehoods supports journalistic integrity and helps maintain informed public discourse.
- Relevance of the Project: This project is highly relevant as it leverages advanced NLP techniques to address a critical issue in digital media. It combines educational value with practical application, presenting a robust response to a contemporary challenge affecting public safety and information accuracy.

2. Data Collection or Generation (30%)

fake-and-real-news-dataset:

<https://www.kaggle.com/datasets/clmentbisailon/fake-and-real-news-dataset/data>

fake news classification:

<https://www.kaggle.com/datasets/saurabhshahane/fake-news-classification>

Fake or Real News:

<https://www.kaggle.com/datasets/jillanisofttech/fake-or-real-news>

those two datasets got cleaned and are ready for model training. Additionally a web scraper was built to scrape real articles from the web (1) and the same amount of fake articles (0) was created with gpt. This additional dataset is used to test and validate the model.