**REU project proposal:

Project: LLM Prompt-based misinformation detection**

The widespread dissemination of misinformation and fake news in digital spaces has emerged as a significant concern. With the increasing use of social media platforms, false information can now reach millions within moments, shaping public perceptions, influencing political events, and even impacting personal decisions related to health and safety. Instances such as the COVID-19 pandemic, debates on climate change, and election cycles have underlined the dire consequences of misinformation, including public confusion, societal divisions, and declining trust in credible sources. This critical situation has motivated us to develop an automated system aimed at curbing the spread of misleading content and facilitating large-scale fact-checking efforts.

This project aims to develop a prompt-based misinformation detection framework using large language models like GPT-4 and beyond. The prompt consists of three components: the original sample (original input and its ground-truth label), the prompt goal, and the prompt guidance. The ground-truth label is simply whether the news is fake or real. The prompt guidance incorporates external knowledge into the prompt representation, which ultimately helps in detecting misinformation. The external knowledge is retrieved by combining content from verified article sources and Wikipedia data. Furthermore, the model’s performance can be improved by incorporating features such as sentiment analysis, user profiling, and other metadata.