Deepfake Text Detection Using NLP-Based Neural Networks

Abstract:
Deepfake technology, which manipulates media content, is increasingly used to spread misinformation. Detecting deepfake-generated text (e.g., AI-generated fake news or manipulated statements) is crucial for combating this threat. This project proposes a system for **deepfake text detection** using **NLP-based neural networks** to identify AI-generated text that mimics human writing.

Methodology:
The system will employ **NLP techniques** such as **text classification** and **style analysis** to capture linguistic patterns indicative of deepfake text. Features like sentence structure, vocabulary use, and coherence will be extracted from text samples. The system will use **neural networks** like **LSTM**, **CNNs**, and **Transformer-based models** (e.g., **BERT** or **RoBERTa**) to detect anomalies that are typically present in machine-generated text.

Outcome:
The expected outcome is an AI-driven deepfake text detection system that can reliably differentiate between human-written and AI-generated content. This system will help combat the spread of fake information in the media, social networks, and online platforms, improving content credibility.