

Violating and Offensive Content on Social Media

Name: Kshitij Mohan

Branch: CSE

Institute Name: Indraprastha Institute of Information Technology Delhi

Contact No.: 9910295298

City: New Delhi

E-Mail: kshitij19054@iiitd.ac.in





Content Synopsis

OVERVIEW OF KEY IDEAS

Introduction

Problem Statement

My Solution

Implementation Plan



Introduction

The term Social Media generally refers to the modern means of communication such as Facebook, Twitter, Instagram, YouTube, etc. They were initially developed to connect and allow an effective exchange of information between people.

Today, Social Media has gradually evolved into an integral aspect of people's daily lives. The new generation is using it much more efficiently due to which its dependence has drastically increased. Social Media influences a large number of people especially teenagers.

In the current situations, the dependence of social media has increased even more as it has become a major source of entertainment. People have become more keen to develop their online presence and connect with people they cannot meet in real life.



Problem Statement

Teenagers today have access to social networking sites that enable them to communicate and have fun. Because of the large number of social media platforms, not all of them have a safe atmosphere for children. In such situations, the negative effects of social media on teenagers become more pronounced as the use of offensive and foul language in social interactions rises. This increase can cause anger, depression, and a significant shift in their behaviour. Many social media platforms have already identified this issue and have started banning and hiding these contents. But still a lot of them pass through.





My Solution

I believe that it is our responsibility to spread awareness amongst the new generation to stop using these offensive or violating words. If many of us try to not use these words and tell people near not to use them, We can bring a great change in our society. Taking the above points in consideration, I have planned a 2 stage solution to this problem:

- **Identification:** First, We need to identify these words and the emotions the person is in while using these words. This will help us understand why do people generally use these words.
- **Spreading Awareness:** Then, We need to make people aware of the possible consequences of their actions. We need to help express themselves without using these words and even provide them with counselling that can help them if they are facing some serious problems.



Implementation Plan



The most efficient method to achieve the above mentioned ideas is to design an app that provides a keypad to the users. The app will have an AI model will sense any use of foul, offensive and violating words and provide the user with better alternatives to express their emotions. There will be another model that will predict the emotions of the user using the words typed and in case of emergency can automatically recommend assistance.

Identification

The identification can be easily done by using Natural Language Processing and Deep Learning concepts. By using these concepts a model can be created that will automatically detect the use of foul, offensive and violating words.

Spreading Awareness:

Counselling and assistance can be provided to the users that tend to use these words more frequently by connecting them to Counsellors. This can be done by providing them an AI chatbot (within the app itself) with which they can share their problems while maintaining their anonymity and it will connect them to the counsellors.