

Pattern Recognition

Mitigating Bias in Real-Time Cyberbullying Detection

CSE424

Group 02

- 
- 01/ Maliha Binte Masud
 - 02/ Abdur Rahman Shafi
 - 03/ Rakibul Hasan Rafi
 - 04/ Amina Zannat Nurhan
 - 05/ Sabbir Hossain Mirza
 - 06/ Fahmid Hasan Chowdhury
 - 07/ Radito Dhali
 - 08/ Zakaria Ibne Rafique

Introduction

The exponential growth of social media has brought with it a serious challenge: cyberbullying. While numerous solutions have been proposed to combat this issue, achieving unbiased detection of online harassment in real-time remains a significant hurdle. Our project aims to strike a critical balance – effectively mitigating biases based on factors like gender, language, recency, religion, race, and nationality – while simultaneously classifying text that constitute harassment with exceptional accuracy.

O1 - Investigation

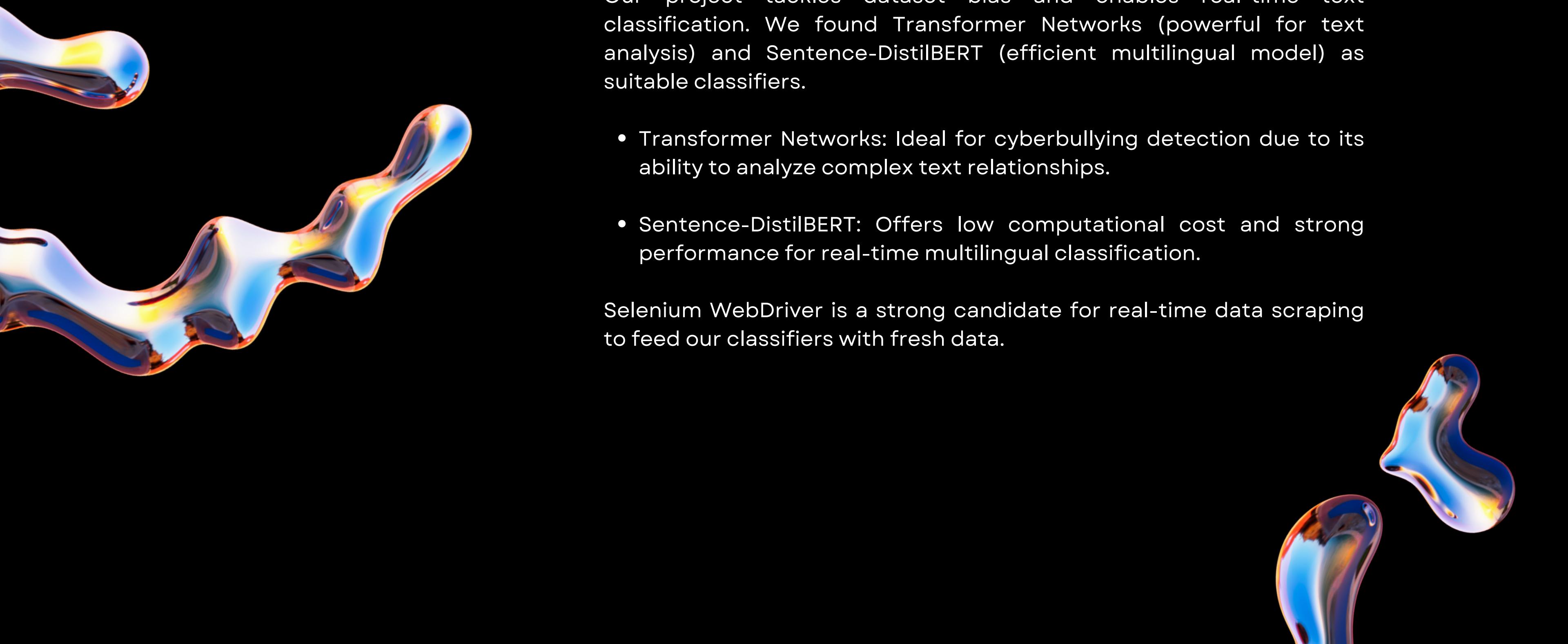
01/ Investigation

- Unintentional biases in cyberbullying detection based on gender, language, recency, religion and race.
- Studies are mostly based on datasets from specific social media platform.
- Imbalanced datasets hampers accuracy.
- The model was based on lexical analysis and only tackles textual form of bulling.



03 - Analysis

02/ Analysis

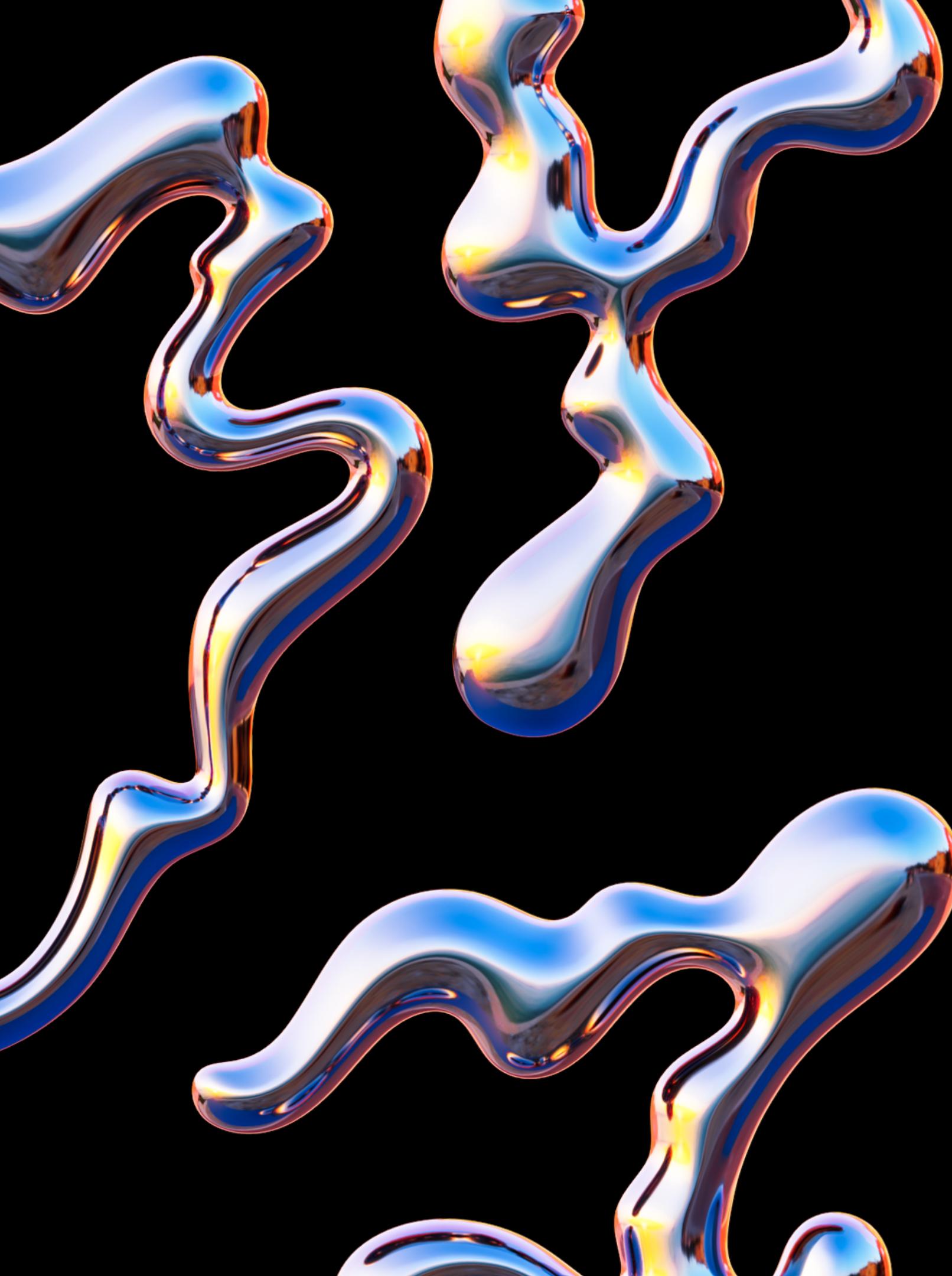


Our project tackles dataset bias and enables real-time text classification. We found Transformer Networks (powerful for text analysis) and Sentence-DistilBERT (efficient multilingual model) as suitable classifiers.

- Transformer Networks: Ideal for cyberbullying detection due to its ability to analyze complex text relationships.
- Sentence-DistilBERT: Offers low computational cost and strong performance for real-time multilingual classification.

Selenium WebDriver is a strong candidate for real-time data scraping to feed our classifiers with fresh data.

04 -Conclusions

A large, abstract graphic on the left side of the slide features several organic, flowing shapes resembling liquid metal or plasma. These shapes are highly reflective, with bright highlights and deep shadows, creating a metallic and futuristic appearance against a solid black background.

04/ **Conclusions**

Our project proposes a real-time cyberbullying detection system using Transformer Networks for accuracy and Sentence-DistilBERT for efficient multilingual processing. By combining these with Selenium WebDriver for data scraping, we aim to achieve real-time and unbiased classification of cyberbullying text.

CSE 424

Thanks

Pattern Recognition