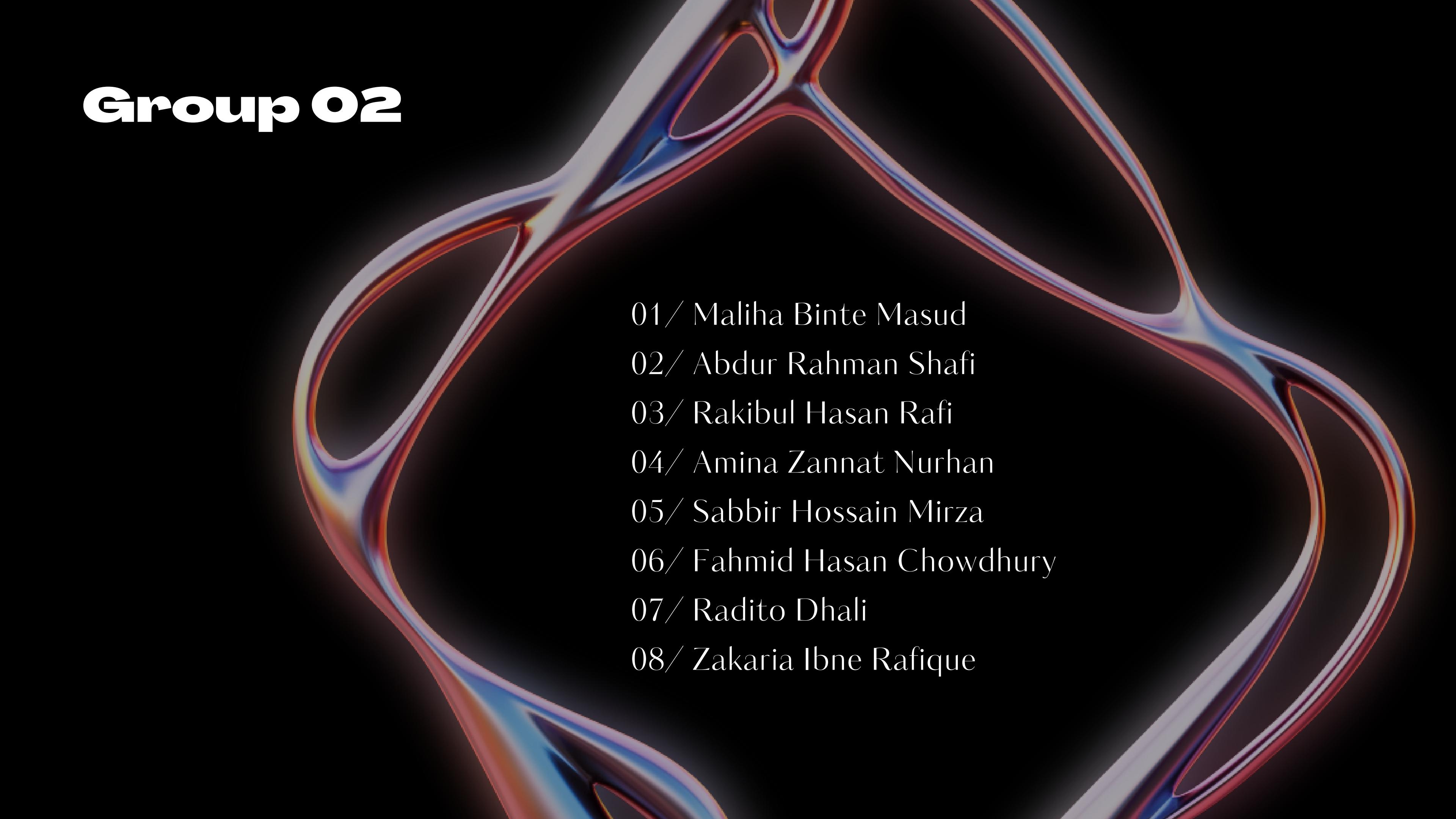


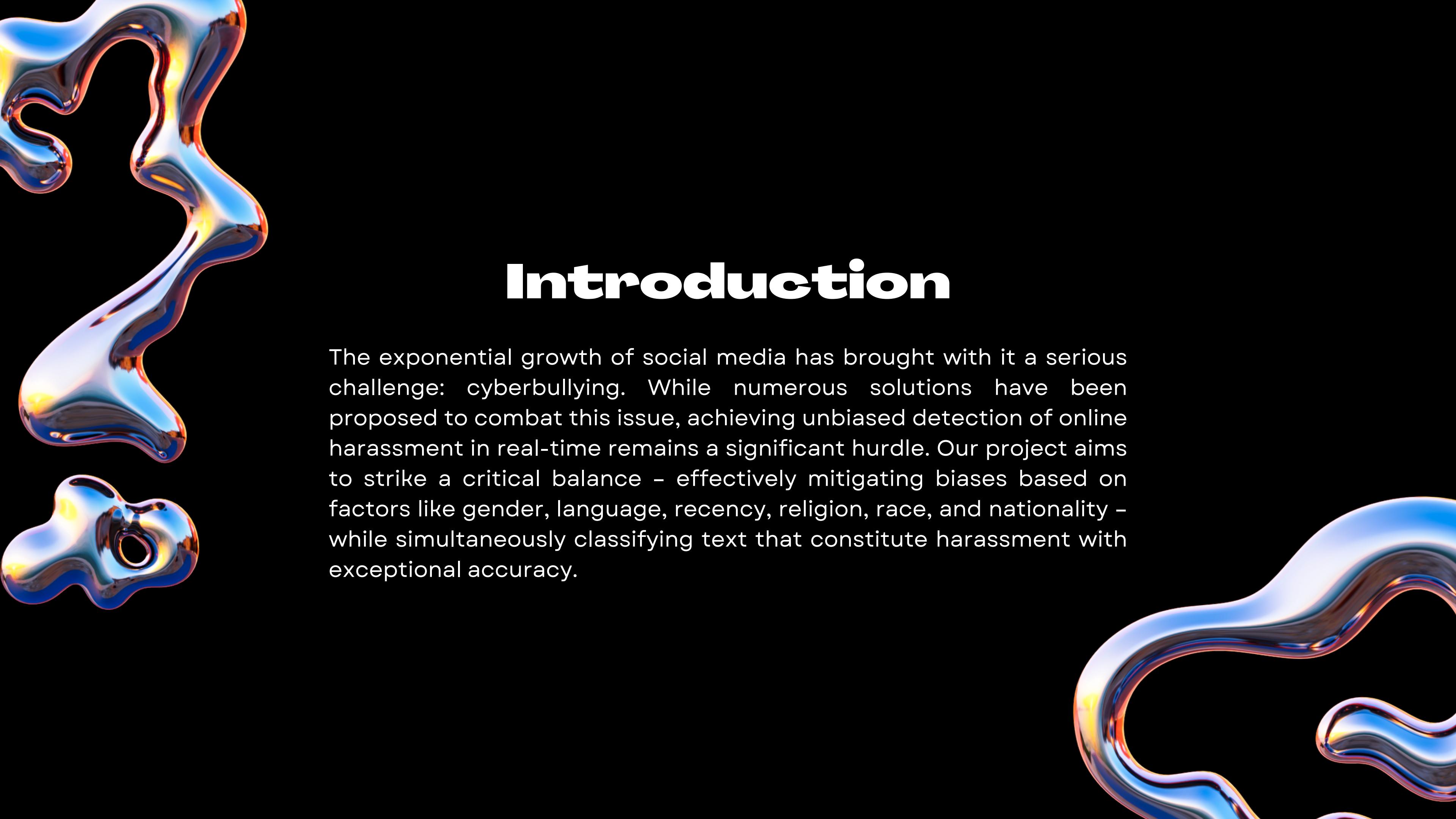
Pattern Recognition

# 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

The background features abstract, flowing metallic liquid shapes in shades of blue, silver, and gold against a black background. These shapes are organic and fluid, creating a sense of movement and depth.

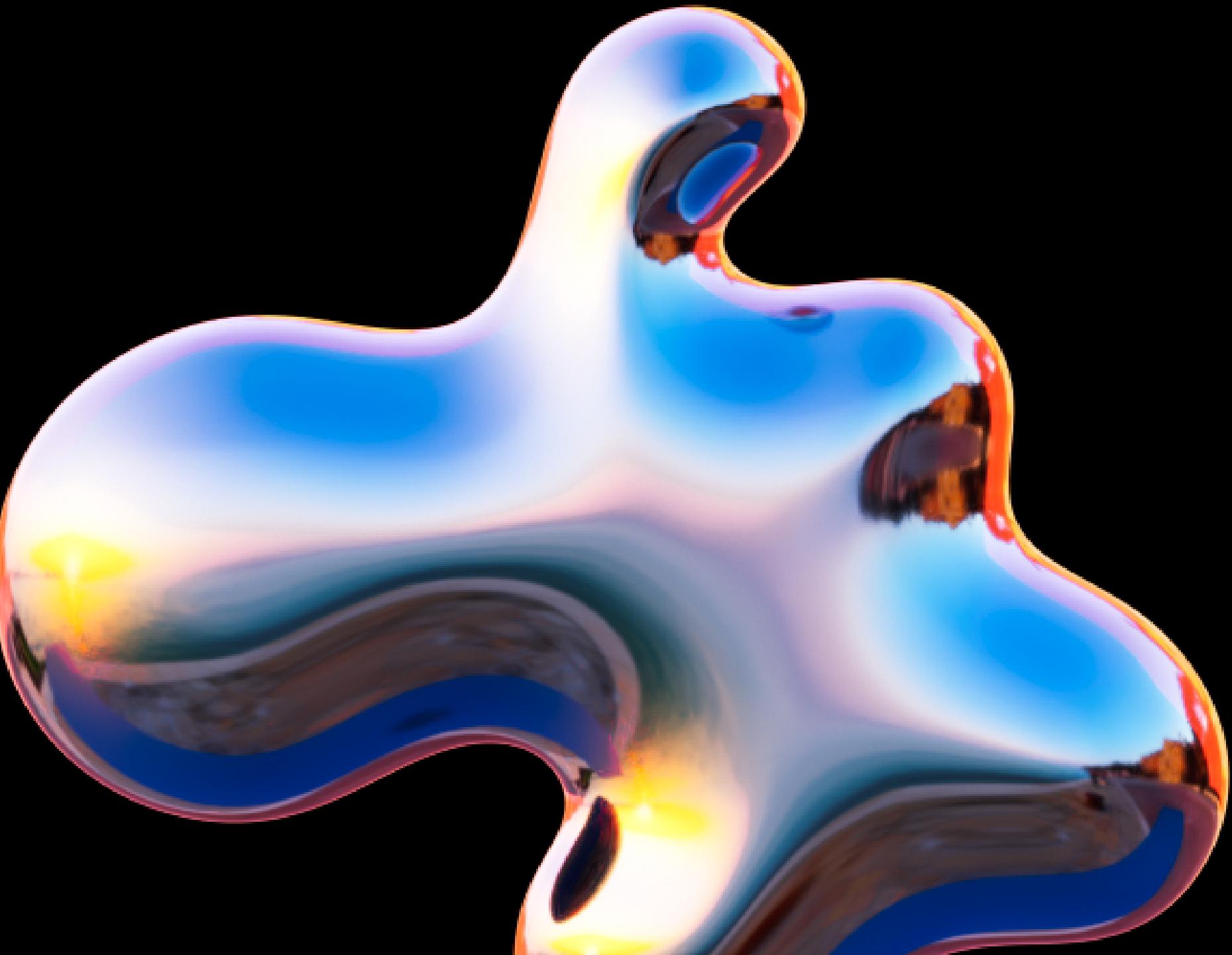
# Introduction

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**

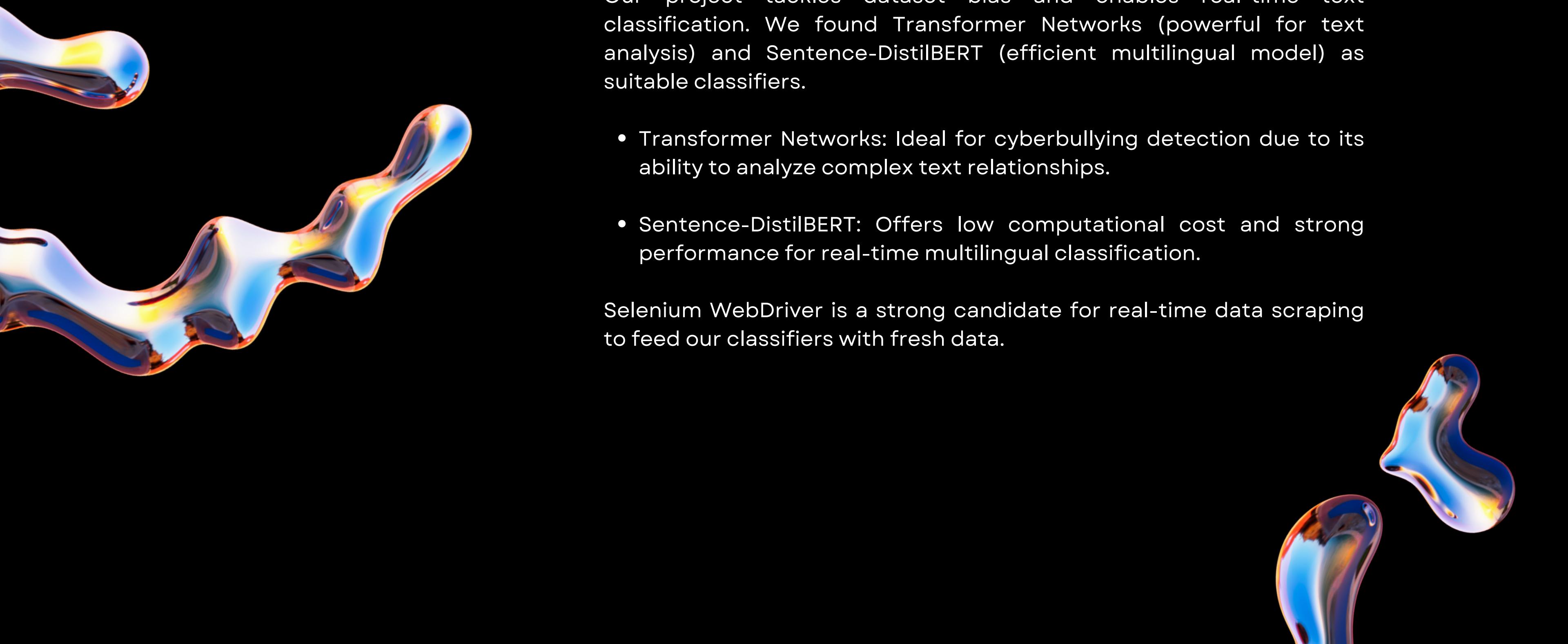
# 01/ Investigation

- 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.
- Studies that specialized in identifying various manifestations of cyberbullying on Twitter, but limited to processing text written in English fonts.



# 03 - Analysis

# 02/ Analysis

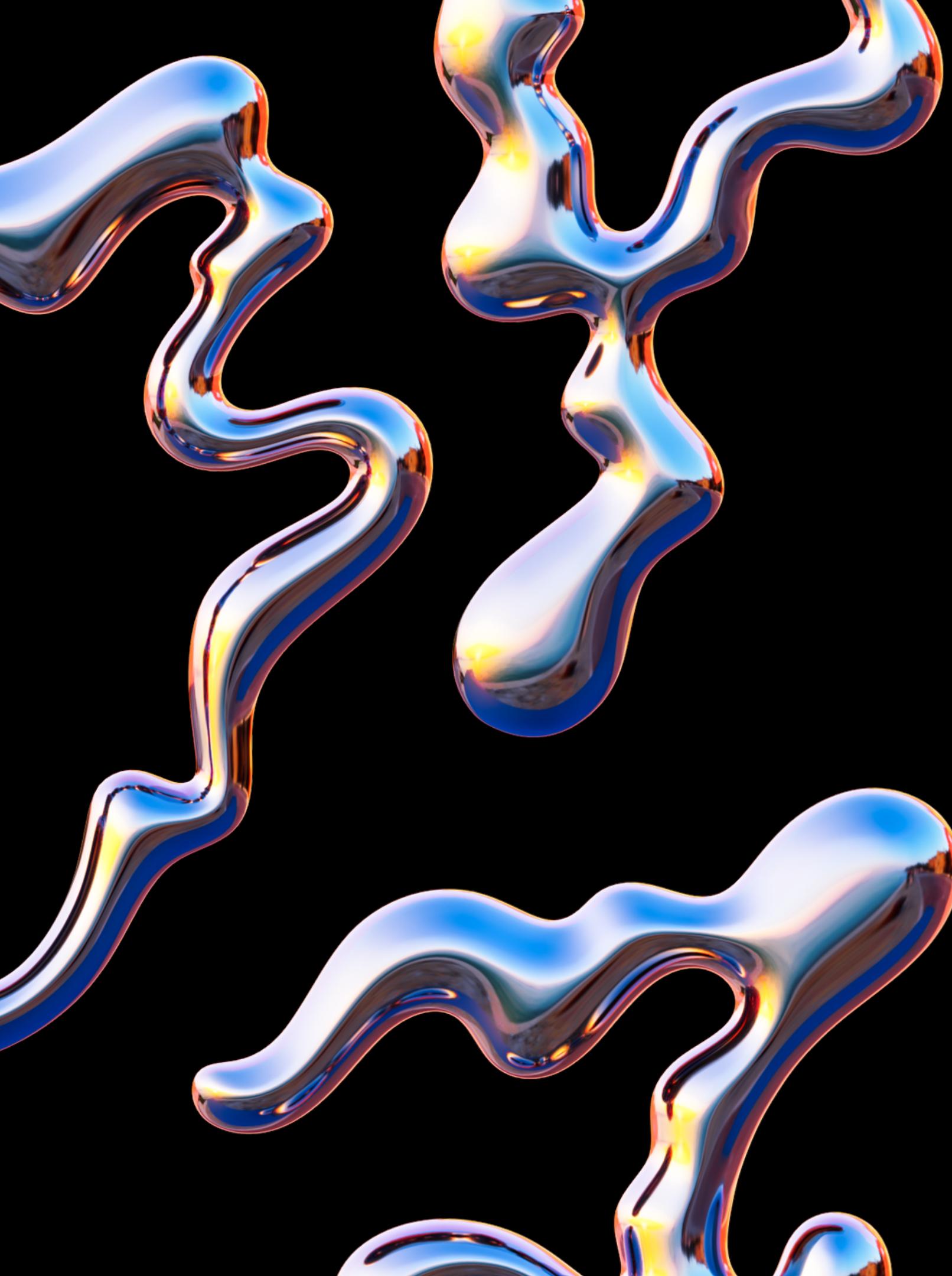


Our project tackles dataset bias and enables real-time text classification. We found Transformer Networks (powerful for text analysis) and Sentence-DistilBERT (efficient multilingual model) as suitable classifiers.

- 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**



## 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