# Apollo-1

Kumar Shridhar & Iyer Venkatesh

# IDEA

# Idea

---

- ‘**FUCK YOU**!’
- A couple of words can ruin an entire discussion!
- **Machine Learning** to the rescue.

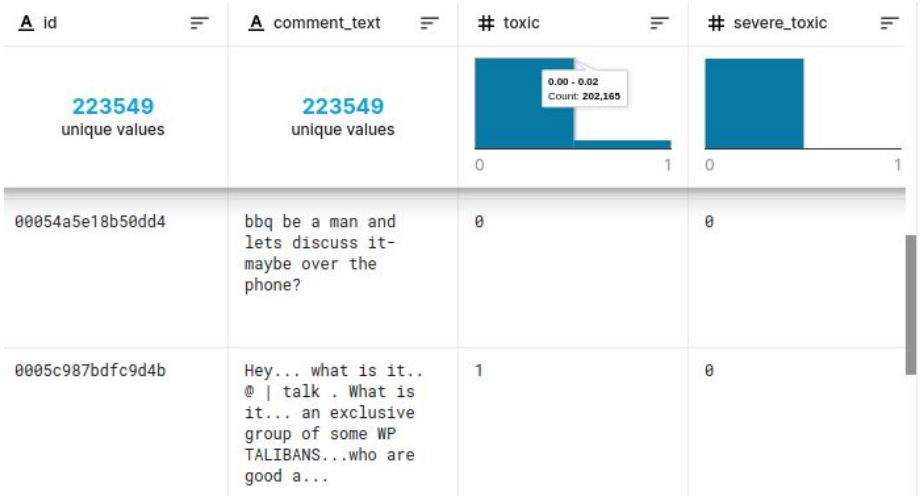
- We propose an idea to identify **toxicity** in online **conversations**.

# DATASET

# Dataset

- **Training** Data
  - Toxic comment Challenge 2018 [1]
  - Unintended Bias Challenge 2019 [2]
    - **Language** : English

- **Testing** Data
  - Wikipedia comments
    - **Language** : Several Languages
    - **Prediction**: 1 for toxic, 0 for non-toxic



| id | comment_text | toxic | severe_toxic |
|---|---|---|---|
| 223549 unique values | 223549 unique values | 0.00 - 0.02 Count: 202,165 (0 … 1) | (0 … 1) |
| 00054a5e18b50dd4 | bbq be a man and lets discuss it-maybe over the phone? | 0 | 0 |
| 0005c987bdfc9d4b | Hey... what is it.. @ \| talk . What is it... an exclusive group of some WP TALIBANS...who are good a... | 1 | 0 |

[1] https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge

[2] https://www.kaggle.com/c/jigsaw-unintended-bias-in-toxicity-classification

# APPROACH

# Approach

- We will use '**XLM-RoBERTa**'[1]
  - XLM-R is a **multilingual** model trained on **100** different languages. Unlike some XLM multilingual models, it does not require *lang* tensors to understand which language is used, and should be able to determine the correct language from the input ids.
  - Available in HuggingFace's Transformers[2] library.

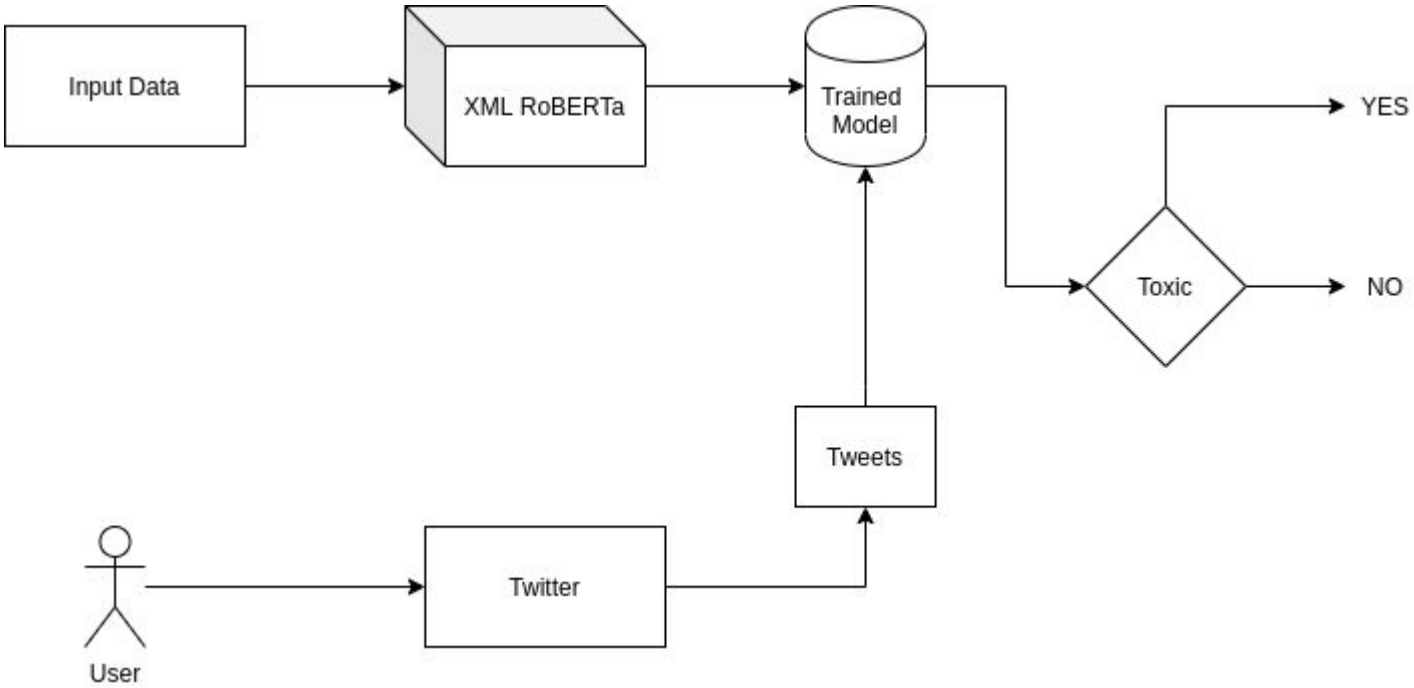
- We propose to use the training data as mentioned previously and infer on Twitter feeds.

[1] https://arxiv.org/abs/1911.02116

[2] https://github.com/pytorch/fairseq/tree/master/examples/xlmr

# PIPELINE

# PIPELINE


Input Data
XML RoBERTa
Trained Model
YES
Toxic
NO
Tweets
Twitter
User

# PROBLEM PRIORITY MATRIX

# Problem Priority Matrix

Scale : **1** (lowest)- **5** (highest)

| Priority Parameter | Problem |
|---|---|
| Problem relevancy | 5 |
| Solution feasibility | 3 |
| Probable solution | 4 |
| Solution complexity | 4 |
| Timeline feasibility | 5 |

# Thanks!