Defending Fun

Using Deep Learning and Natural Language Processing to Tackle Toxicity in Esports Titles

Thomas Vernon 1907240

What's the problem?

- Toxicity is still a problem in Esports titles.
- Players from different countries, backgrounds and cultures will find themselves competing against and with one another to rise the ranks.
- This competitive atmosphere creates the possibility for toxicity to occur between players.

● *MUERTO*(Antiterrorista) twitch.tv/tense1983 : gg are noobs
● *MUERTO*(Antiterrorista) twitch.tv/tense1983 : very noobs
● (Antiterrorista) twitch.tv/tense1983 en Base AT : green retard
● (Antiterrorista) twitch.tv/tense1983 en Base AT : me say alone ramp
● (Antiterrorista) twitch.tv/tense1983 en Base AT : u come ramp

bastard| ENVIAR

What's being done about it?

- Many games have systems in place to try and mitigate toxicity, but not all of them are effective.
- Counter Strike's built-in system is poor and relies on player reports.
  - o Repercussions are not very severe.
- Valorant and Faceit* both use an AI-based solution that are very effective.
  - o Their punishments are very harsh, which demonstrates their effectiveness.

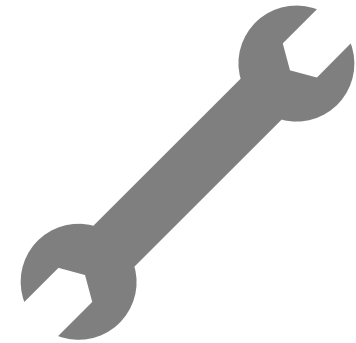*a third-party matchmaking service for Counter Strike
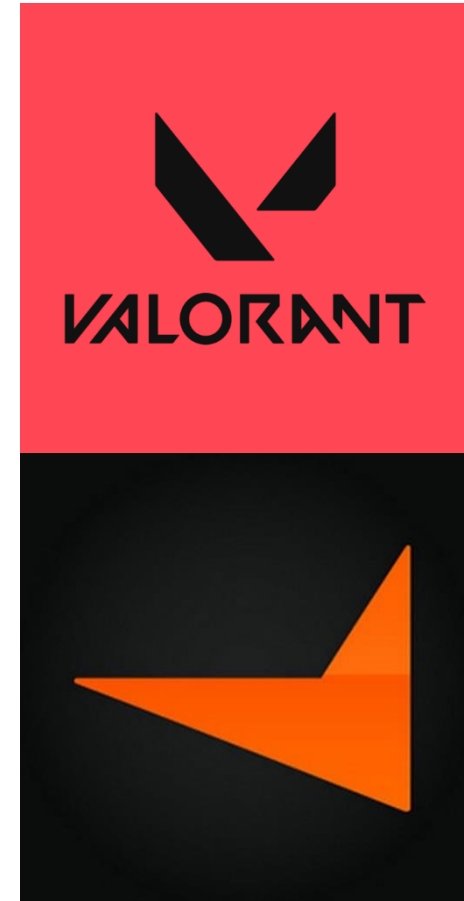
Objectives

- To create my own AI-based solution that can detect toxic players using the text chat from a Counter Strike game.
- Use the experience gained from development and other users to determine the effectiveness of the software, and what improvements need to be made.

What Did I Use?

- Counter Strike used due to its open platform.
- Python Notebooks with Anaconda for great AI workflow and easy library management.
- "Awpy" demo file parser:
    - Counter Strike can represent any past match played as a recorded "demo file".
    - This can be parsed for data analysis.
- "Detoxify" deep learning model:
   - A model trained for detecting different forms of toxicity in strings of text.

My Solution

- Parses a .dem file from a Counter Strike match using Awpy.
- Processes the JSON into a useable data structure.
- Cleans the corpus using a dictionary.
- Uses Detoxify to generate toxicity scores for every chat message.
- These scores are summed for each player, resulting in a "toxicity score" for each player in that match.
  - A weighting system was implemented but not changed from uniform weighting when testing the software.
- The data from four example matches were recorded and presented to Counter Strike players to judge and give their thoughts.

Counter Strike → *demo file* → **Awpy**

Awpy → *JSON* → **Acronym Replacement**
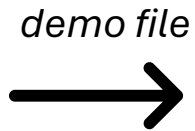
Chat Extraction → *organised chat array* → **Acronym Replacement**

**Acronym Replacement** → *cleaned chat* → **Score Calculation**

Detoxify → *predicted features* → **Score Calculation**

Player 1 | SteamID: 76561199143972579
['tdekt nana', 'sonte', 'GO KILL URSELFT'
THE MESSAGE', ':dddddddddddddddddddddddd
'GG', 'ORANGE I HOPE YOUR FAMILY DIE\\',
'OUR ORANGE', 'SHUT THE FUCK UP', 'OK', '
FAMILY TOO PURPLE', 'LONG', '3 LONG', 'LO
'GG', 'LOWER', 'LONG', 'LOWER', 'B', '\\'
'GG', 'G', 'G', 'G', 'G', 'LONG', 'DOORS'
GAME', 'PLEAES', 'EAH UR LAST IN TAB', 'H
STILL\\A', 'AH OKAY', 'STILL 10 KILLS', '
SITE', 'GREEN', 'CONGRATULATYIONS', 'HAHA
'TS', 'TS', 'TS']
Player 2 | SteamID: 76561199191984504
['hooom', 'gg', 'gg', 'lasrt short', 'so
team hiding on long', 'classic lvl 5 game
mongodanske', 'AA VAAAO DENSKO AAA VAAO',
LAAO', 'a vaoo sa oo', 'aa vaao ', 'aa va
' AAA VAAO LLAAO VAOO FORDI IGG AA VOOA',
vooa", 'aa vaao mand??', 'mænd aa vaoo sa
'a vao sa du nuo', 'last think ure in t s
insane aim', 'AHAHAHAH', 'HAHAHA', '1 top
'a idiot from my team top mid', 'who is b

Game ID: 9af30566-56f5-44d8-a4bb-ef634c0bc5d1.dem
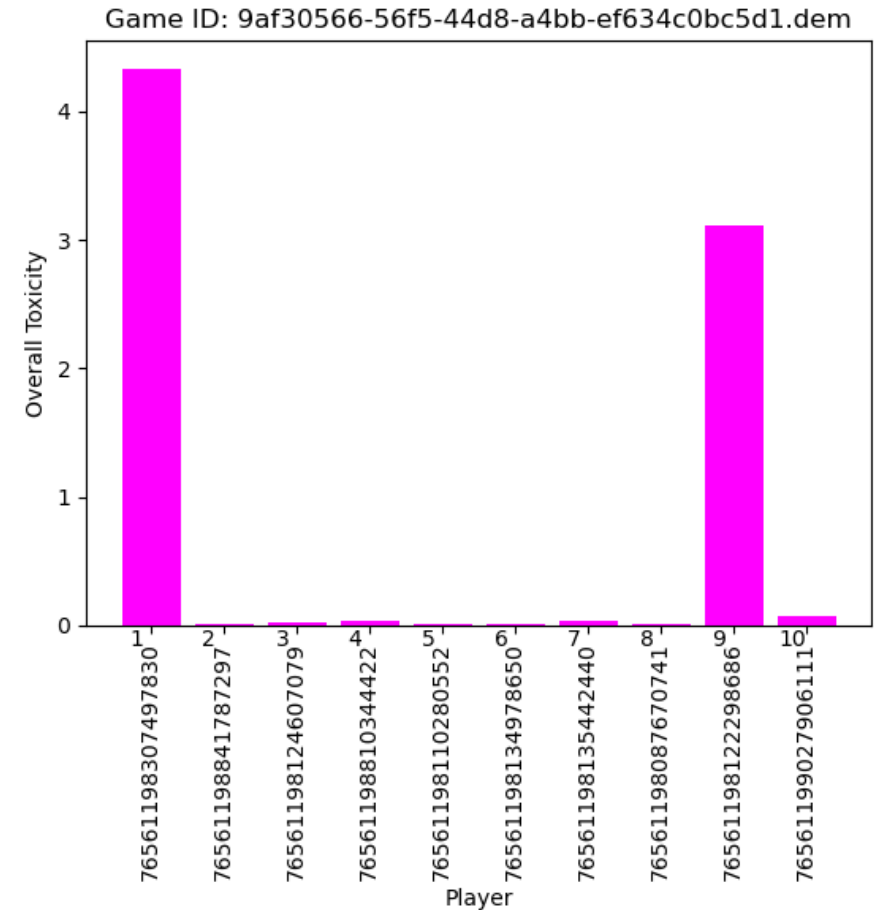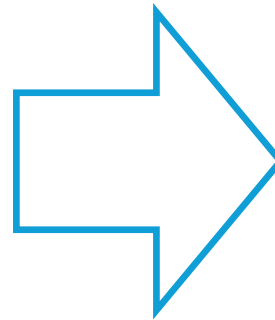
Findings

- Application of a deep learning model to detect toxicity is very feasible.
- Fine tuning the model is essential to maximise accuracy, context of a general model is not enough.
  o Publicly available data for training is very limited
- Careful tweaking of parameters and cleaning input data is important.
- Extra data is needed to detect more subtle forms of toxicity such as "ghosting".
- All very important if the model is to be used seriously and to administer harsh punishments.