

# Natural Language Processing and Sentiment Analysis for Verbal Aggression Detection; A Solution for Cyberbullying during Live Video Gaming

Natalia Stepanova  
Global Innovation Exchange,  
University of Washington  
natastep@uw.edu

Michael Cross  
Global Innovation Exchange,  
University of Washington  
mccross@uw.edu

Wesley Muthemba  
Global Innovation Exchange,  
University of Washington  
wam98@uw.edu

Nicholas R. Ames  
Global Innovation Exchange,  
University of Washington  
amesr2@uw.edu

Ross Todrzak  
Global Innovation Exchange,  
University of Washington  
rtodrzak@uw.edu

John Raiti  
Global Innovation Exchange,  
University of Washington  
jraiti@uw.edu

## ABSTRACT

Verbal aggression during online multiplayer games is a common occurrence on video gaming platforms. From casual cursing and negative remarks to more targeted and damaging cyberbullying cases, the gaming industry continues to see an increasing number of players engaging in these exchanges. The COVID-19 pandemic has led to a rise in the number of active gamers due to social distancing and stay-at-home orders. As a result, cases of cyberbullying and toxicity have increased by 40% on popular gaming platforms such as Discord [4]. To help resolve this issue, we created an aggression awareness tool called Tempr to discourage verbal aggression in gamers under the age of 18 and reward non-aggressive behavior.

## CCS CONCEPTS

• **Human computer interaction (HCI);** • **Interaction Paradigms;** • **Natural Language Interfaces;**

## KEYWORDS

Natural Language Processing, Sentiment Analysis, Cyberbullying, Video Games

## ACM Reference Format:

Natalia Stepanova, Wesley Muthemba, Ross Todrzak, Michael Cross, Nicholas R. Ames, and John Raiti. 2021. Natural Language Processing and Sentiment Analysis for Verbal Aggression Detection; A Solution for Cyberbullying during Live Video Gaming. In *The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA 2021), June 29–July 02, 2021, Corfu, Greece*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3453892.3464897>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*PETRA 2021, June 29–July 02, 2021, Corfu, Greece*

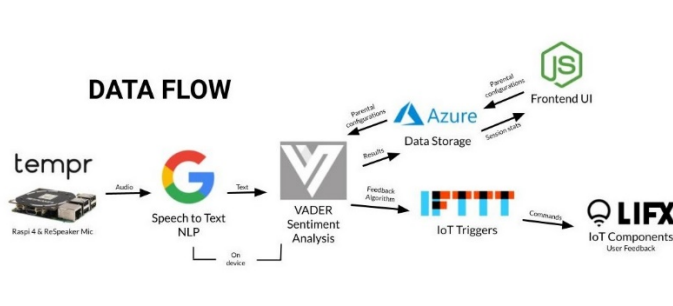
© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-1-4503-8792-7/21/06...\$15.00  
<https://doi.org/10.1145/3453892.3464897>

## 1 INTRODUCTION

Aggressive behavior triggered by stressful situations in which people have little to no control can be very difficult to mediate [6]. Aggression is known to cause feelings of irritability and restlessness, and it can be beneficial for people to recognize the state of anger they are in [1, 5]. Due to COVID-19, more people have been playing video games to pass the time and, in turn, "toxicity," or gamer aggression, has risen [7, 10]. Gamer aggression can impact mood and relationships, and can continue into post-gaming sessions [3, 9]. Moreover, constant harassment from toxic players can cause depression in others playing with them [8]. About 50 million gamers in the United States are kids under the age of 18, and on these gaming platforms, there are over 150 million adults potentially influencing younger gamers with explicit language and behaviors [2]. One of the greatest problems is that gamers struggle with self-monitoring and recognizing their state of mind [5, 6]. To combat this issue, 'Tempr' offers a unique approach through self-awareness and positive reinforcement to help decrease verbal aggression over time.

## 2 MATERIALS AND METHODS

Tempr relies on readily available consumer-oriented technology. For example, Tempr uses smart home IoT controlled LED strips mounted behind a TV (so that visual feedback is line-of-sight and functions as relief lighting), a smart home IoT outlet, a microphone, and various APIs for speech and sentiment analysis. Tempr also integrates a parent-oriented portal for configuration and monitoring. Tempr actively listens to the gamer's speech, transforms it to text, runs sentiment analysis and curse word detection, and generates a score for visual feedback through the LED strip (Figure 1). The LEDs start out green to indicate a neutral state. Increased verbal aggression changes the lights from green to yellow, orange, and red depending on the frequency of curse words. Users have the opportunity to deescalate the colors back to green by saying more positive statements. When curse words are detected, Tempr will blink to alert the user that a curse was detected. Parents can configure a customizable list of inappropriate words and track the number of times per session an inappropriate word has been detected by Tempr before the IoT outlet powers off the console. Parents also can



**Feedback Algorithm** Real-Time Feedback and Reward Structure

SENTIMENT TRIGGERS		SENTIMENT LIGHT SCALE (SLS)	
Trigger (sentiment score ranges from -1.0 to +1.0)	Action	Description	
Neutral or positive sentiment score $\geq -0.05$	Lights: SLS level 1	Neutral (positive sentiment or no aggression detected)	
Average (not median) sentiment score between $-0.5$ and $-0.05$ across 1 minute	Lights: SLS level 2	Low Aggression	
Three consecutive 1 min periods in $-0.5$ to $-0.05$ range	Lights: SLS level 3	Medium Aggression	
Average compound sentiment score over below $-0.5$ over 1 minute	Lights: SLS level 3	Medium Aggression	
Two consecutive 1 min periods below $-0.5$	Lights: SLS level 4	Medium High Aggression	
3rd consecutive period below $-0.5$	Lights: SLS level 5	Max Aggression	

CURSING	TIME ALLOWANCE
<ul style="list-style-type: none"> <li>Lights blink red each time curse detected</li> <li>N-1 curse threshold warning light</li> <li>Exceeding curse threshold powers down console</li> </ul>	<ul style="list-style-type: none"> <li>+2 minutes for positive average sentiment</li> <li>-2 minutes for negative average sentiment</li> <li>-3 minutes for hitting curse threshold</li> </ul>

Figure 1: Tempr data flow diagram and sentiment feedback rule set.

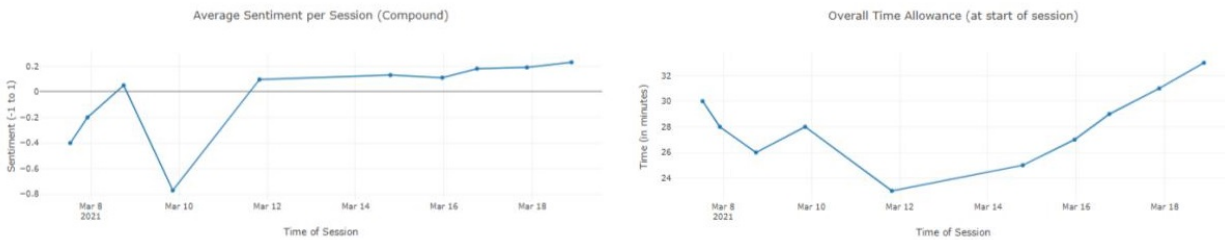


Figure 2: Representation of session sentiment and time allowance over time.

set an initial time-allowance that dictates how long each gaming session will last, and this increases or decreases overtime based on cursing and aggression levels. Reaching the maximum curse tally or a high aggression level will decrease time allowance, whereas positive behavior increases game time allowance. Lastly, parents can override accumulated time allowance and view statistics like sentiment over time, number of curse words per session, and how the game time allowance changes over time through a companion interface (Figure 2).

### 3 CONCLUSION

The Tempr prototype illustrates the feasibility of real-time natural language processing and sentiment analysis to detect verbal aggression during live video game play. Tempr provides visual feedback of sentiment analysis plus data tracking of aggression to help parents and children reduce toxicity during gameplay.

### REFERENCES

[1] Ashcroft, H. (2019, July 17). How To Control 'Gamer Rage' In Yourself Or Your Loved Ones. The Gamer. <https://www.thegamer.com/how-to-control-gamer-rage-in-yourself-loved-ones/>

[2] Entertainment Software Association. (2020, July). 2020 Essential Facts About the Video Game Industry. [https://www.theesa.com/wp-content/uploads/2020/07/2020-ESA\\_Essential\\_facts\\_070820\\_Final\\_lowres.pdf](https://www.theesa.com/wp-content/uploads/2020/07/2020-ESA_Essential_facts_070820_Final_lowres.pdf)

[3] Kowert, R. (2020). Dark Participation in Games. *Frontiers in Psychology*, 11, 29-69. 10.3389/fpsyg.2020.598947

[4] L1ght. (2020, November 19). L1ght Releases Groundbreaking Report on Corona-Related Hate Speech and Online Toxicity: L1ght: Detect & Remove Online Toxicity. L1ght.com. <https://l1ght.com/l1ght-releases-groundbreaking-report-on-corona-related-hate-speech-and-online-toxicity/>.

[5] Lemmens, J. S., Valkenburg, P. M., & Peter, J. (2011). The Effects of Pathological Gaming on Aggressive Behavior. *Journal of youth and adolescence*, 40(1), 38-47. <https://doi.org/10.1007/s10964-010-9558-x>

[6] Maldonado, M. (2016, May 17). Frustration with Video Games Leads to Aggressive Behavior. PsychCentral. <https://psychcentral.com/lib/frustration-with-video-games-leads-to-aggressive-behavior#1>

[7] Nielsen. (2020, June 3). 3, 2, 1 GO! VIDEO GAMING IS AT AN ALL-TIME HIGH DURING COVID-19. Nielsen Insights. <https://www.nielsen.com/us/en/insights/article/2020/3-2-1-go-video-gaming-is-at-an-all-time-high-during-covid-19/>

[8] Ratan, R., Shen, C., & Williams, D. (2020, December 9). Toxicity in Gaming Is Dangerous. Here's How to Stand Up to It. Wired. <https://www.wired.com/story/toxicity-in-gaming-is-dangerous-heres-how-to-stand-up-to-it/>

[9] Smith, N. (2019, February 26). Racism, misogyny, death threats: Why can't the booming video-game industry curb toxicity? The Washington Post. <https://www.washingtonpost.com/technology/2019/02/26/racism-misogyny-death-threats-why-cant-booming-video-game-industry-curb-toxicity/>

[10] Smith, N. (2020, May 12). The giants of the video game industry have thrived in the pandemic. Can the success continue? The Washington Post. <https://www.washingtonpost.com/video-games/2020/05/12/video-game-industry-coronavirus/>