

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

Natalia Stepanova, Wesley Muthemba, Ross Todrzak, Michael Cross, Nicholas Ames, John Raiti

## INTRODUCTION

Due to COVID-19, online gaming has grown, and in turn, "toxicity" and cyberbullying have risen. Exhibiting toxicity can impact mood and relationships continuing into post-gaming sessions. Moreover, experiencing constant harassment from toxic players can lead to depression in others. 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. One of the greatest problems is that gamers struggle with self-monitoring and recognizing their state of mind. To combat this issue, Tempr provides a unique approach through self-awareness and positive reinforcement to help decrease verbal aggression and cyberbullying over time.

## MATERIALS AND METHODS

Tempr uses IoT controlled LED strips mounted behind a TV so that visual feedback is line-of-sight and functions as relief lighting, an IoT outlet, a microphone, and various APIs for speech and sentiment analysis. Tempr also provides a parent-oriented portal for configuration and monitoring (Fig 1).

Tempr actively transcribes gamer speech-to-text and runs sentiment analysis and inappropriate word detection to generate a score for visual feedback through the LED strip (Fig 2) and parental feedback through the parental companion app. Aggression and curse words affect current gameplay through visual feedback, including session termination, and positive/negative behavior increase/decrease future gaming sessions. Lastly, parents can override accumulated time allowance and view statistics like sentiment and time allowance over time, and number of curse words per session.

## 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

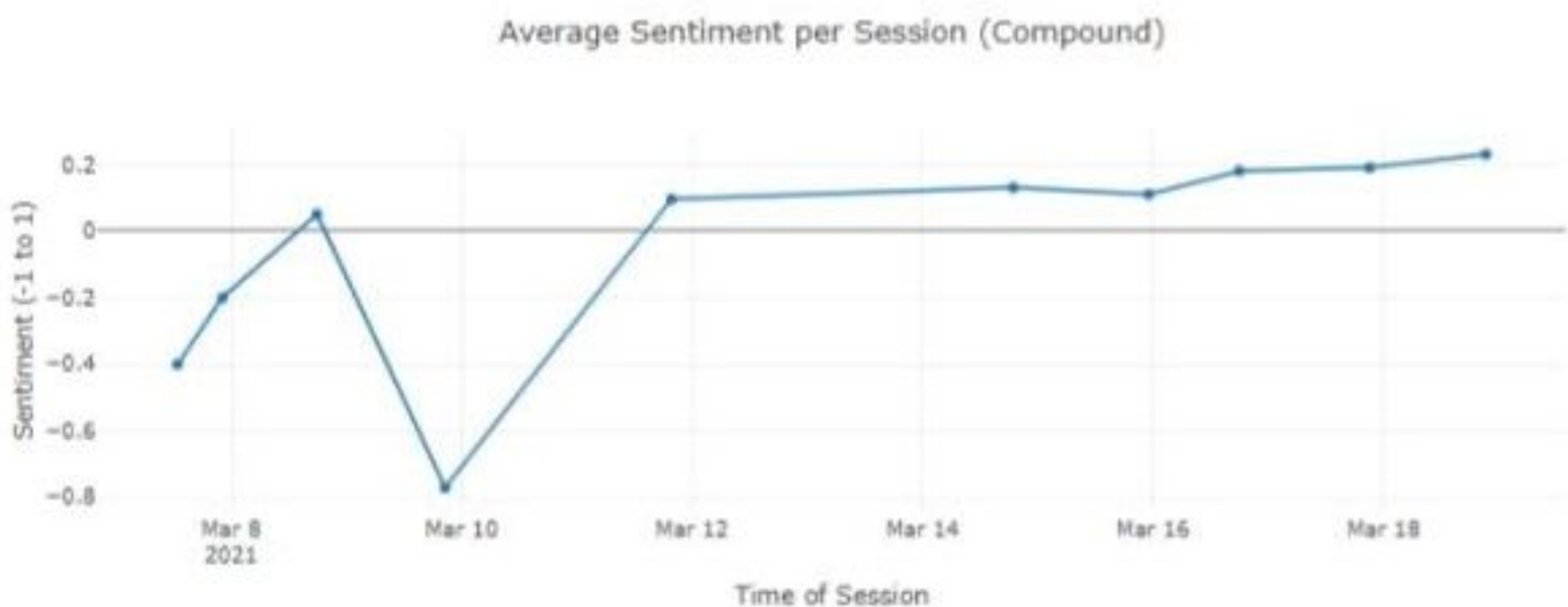


Figure 3: Sentiment Trend Over Time

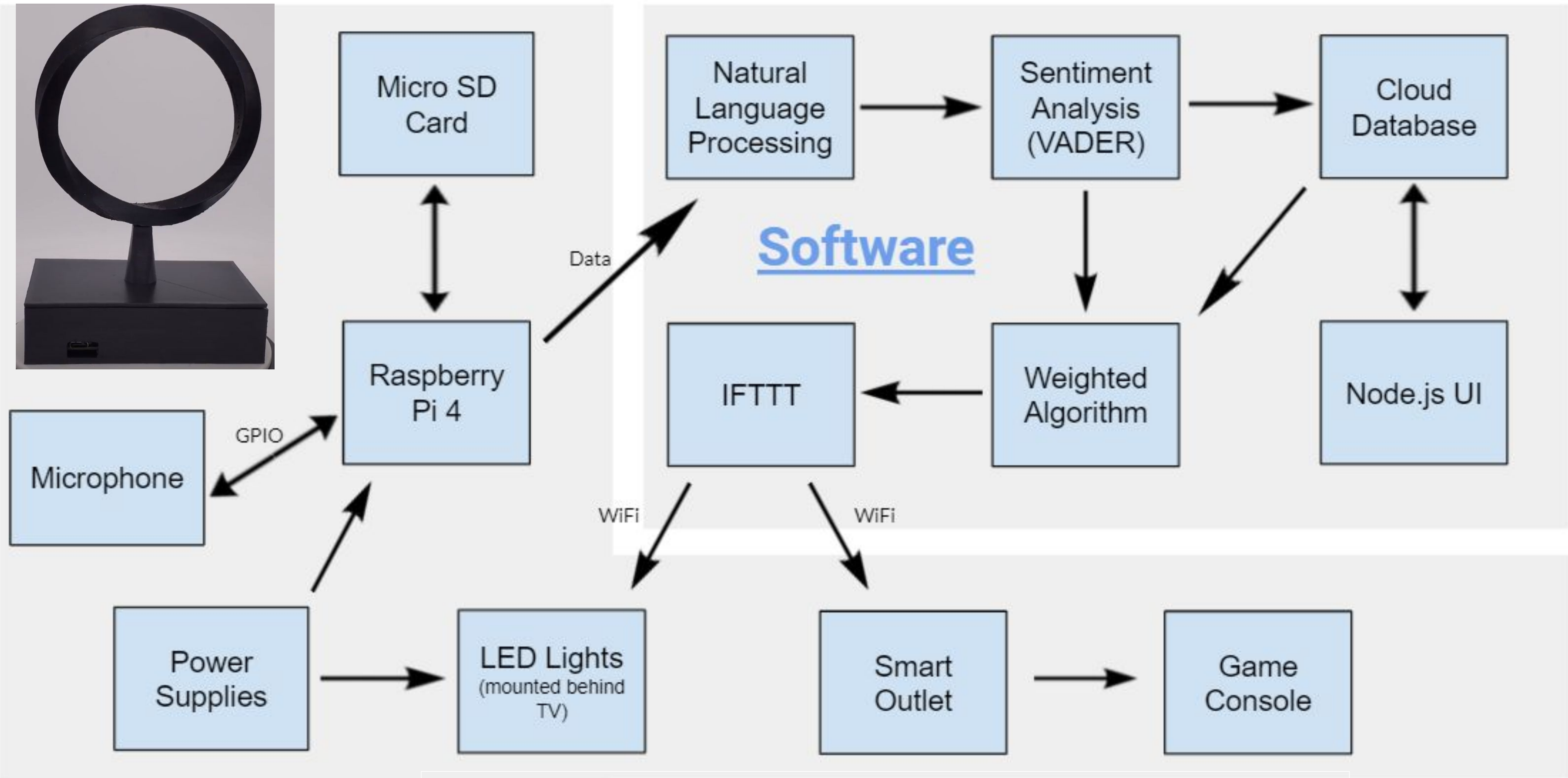


Figure 1: Hardware and Software Architecture

	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 <b>below</b> $-0.5$ over 1 minute	Lights: SLS level 3	Medium Aggression
	<b>Two</b> consecutive 1 min periods <b>below</b> $-0.5$	Lights: SLS level 4	Medium High Aggression
	<b>3rd</b> consecutive period <b>below</b> $-0.5$	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 2: Sentiment Feedback Ruleset