International Journal of Computer Science
and Engineering (IJCSE)
ISSN (P): 2278–9960; ISSN (E): 2278–9979
Vol. 13, Issue 2, Jul–Dec 2024; 1181–1226
© IASET

International Academy of Science,
Engineering and Technology
IASET Connecting Researchers; Nurturing Innovations

LEVERAGING LLMS TO DETECT VIOLATIONS AND ENHANCE CONDUCT MONITORING IN SOCIAL GAMING ENVIRONMENTS

Rahul Jain[1] & Akshun Chhapola[2]

[1]Texas A&M University, College Station, TX 77840, United States

[2]Delhi Technical University, Delhi, India

ABSTRACT

The rapid growth of social gaming environments has brought about significant challenges in ensuring fair play and maintaining a healthy community. As the complexity of player interactions increases, the traditional methods of conduct monitoring are often insufficient to address the scale and diversity of violations. This paper explores the potential of leveraging Large Language Models (LLMs) for detecting and mitigating rule violations in social gaming platforms. By analyzing player-generated content, including in-game chats, posts, and interactions, LLMs can be trained to identify harmful behaviors such as harassment, cheating, and toxic language in real-time. The paper examines how LLMs can be fine-tuned to understand context, detect subtle nuances in communication, and improve the accuracy of violation detection without heavily relying on predefined rule sets. Furthermore, the integration of sentiment analysis and context-aware models enhances the ability to differentiate between harmless banter and serious misconduct. The system also proposes an automated reporting and escalation mechanism for identified violations, ensuring a seamless user experience and timely intervention. Ultimately, the paper demonstrates that LLMs can significantly enhance conduct monitoring systems by providing scalable, efficient, and adaptable solutions that not only detect violations but also foster a safer and more enjoyable environment for players. The findings suggest that LLMs offer a promising approach to addressing the complexities of player behavior in modern gaming ecosystems, contributing to both the development of fairer gaming experiences and better community management.

KEYWORDS: *Large Language Models, Social Gaming, Conduct Monitoring, Violation Detection, Player Behavior, Harassment Detection, Toxic Language, Real-Time Monitoring, Sentiment Analysis, Community Management, Automated Reporting, Cheating Prevention, Gaming Environment, AI-Driven Moderation.*

Article History

Received: 11 Dec 2024 | Revised: 18 Dec 2024 | Accepted: 20 Dec 2024