Chapter 16
AI–Driven Consent Mechanisms:
Enhancing Parental Control While Preserving Child Autonomy in Digital Spaces

Anjali Rawat

https://orcid.org/0009-0006-7171-6016

University of Extremadura, Spain

Anand Rajavat

Shri Vaishnav Vidyapeeth Vishwavidyalaya, India

Purvee Bhardwaj

Faculty of Science, Rabindranath Tagore University, India

Prathamesh Muzumdar

https://orcid.org/0000-0002-9000-363X

Mangalayatan University, India

ABSTRACT

This study introduces an AI-based consent framework designed to strengthen parental control in digital gaming while supporting child autonomy. Using data from the 2023 Video Game Behavior Dataset—comprising over 10,000 interactions from popular online multiplayer games—the approach applies Natural Language Processing and Transformer models (e.g., BERT, GPT-4) to detect harmful content, offer context-aware feedback to parents, and adjust controls based on a child's behavior and maturity. The system empowers children to make supervised digital decisions through a real-time consent model. Evaluation reveals a 98.45% accuracy

in harmful content detection with a 2.7% false-positive rate. The consent feature showed 92% parental satisfaction and led to a 20% rise in positive child engagement, all while maintaining safety. These results suggest AI-driven tools can effectively balance digital safety and child independence in gaming environments.

1. INTRODUCTION

The rapid integration of digital gaming into the lives of children and adolescents has introduced a new set of challenges for parents aiming to balance safety with autonomy in online environments. With the increasing prevalence of online multiplayer games, concerns about exposure to harmful content, cyberbullying, and digital addiction are more pronounced than ever. As a result, there is a growing need for advanced tools that allow parents to implement effective control mechanisms while preserving their child's autonomy in the digital space. Traditional parental controls, such as manual content filtering or time restrictions, often fail to address the complexity of modern gaming environments and the evolving behaviors of young users.

This study introduces an innovative AI-driven (Dhawan et al.,2025) consent mechanism designed to enhance parental control in digital gaming environments while maintaining the child's autonomy. By leveraging cutting-edge Natural Language Processing (NLP) (Bhardwaj et al., 2024) techniques and Transformer-based models like BERT and GPT-4, the proposed framework enables real-time content moderation, context-aware feedback, and adaptive control settings. These models are fine-tuned using the Video Game Behavior Dataset (Rismayanti, 2024) (D'Angelo et al., 2024) (Wang et al., 2024) (Sree et al., 2024) (Faraz et al., 2022) (Hughes et al.,2019), which includes over 10,000 user interactions from a wide range of popular online multiplayer games. The dataset features gameplay data, chat logs, and in-game decisions, which provide rich insights into player behaviors and interactions.

The approach combines real-time monitoring and machine learning algorithms to detect harmful content, such as cyberbullying (Mascari et al.,2025) (Nahar et al.,2023) and inappropriate language, while offering parents the flexibility to adjust control settings according to the child's gaming patterns and maturity level. One of the key innovations of this system is its ability to implement a consent-driven framework (Pithawa et al., 2023), allowing children to make certain decisions about their gaming experience under parental guidance. This dynamic approach is based on a child's individual behavior patterns and learning progress, allowing for a personalized balance between autonomy and safety.

In preliminary evaluations, the system demonstrated an impressive detection accuracy of 98.45% for harmful content with a false-positive rate of just 2.7%. Furthermore, the user satisfaction rate among parents was 92%, highlighting the

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/ai-driven-consent-mechanisms/384745

Related Content

Transformation Through Participatory Action Research in a Community School of Nepal
Roshani Rajbanshiand Niroj Dahal (2023). *Implementing Transformative Education With Participatory Action Research (pp. 130-140).*
www.irma-international.org/chapter/transformation-through-participatory-action-research-in-a-community-school-of-nepal/329916

Effectiveness of Zero Tolerance Policies and Suggestions for Improvement to Reduce School Violence
Danielle Marie Carkin Lacorazzaand Michel-Ange Siaba (2023). *Addressing Violence in the U.S. Public School System (pp. 65-93).*
www.irma-international.org/chapter/effectiveness-of-zero-tolerance-policies-and-suggestions-for-improvement-to-reduce-school-violence/323001

Information Ethics in the Context of Current Developments
Yusuf Esmerand Aye Nihan Arba (2024). *Research Anthology on Business Law, Policy, and Social Responsibility (pp. 1016-1038).*
www.irma-international.org/chapter/information-ethics-in-the-context-of-current-developments/335745

Decoding Lesson Study: Narratives From Assistant Language Teachers in Japan
Gerard Marchesseau (2024). *Inquiries of Pedagogical Shifts and Critical Mindsets Among Educators (pp. 27-51).*
www.irma-international.org/chapter/decoding-lesson-study/339800

Hybrid Warfare: New Implications for NATO's Deterrence and Defense – Asymmetric Challenge
Eka Beraia (2021). *NATO and the Future of European and Asian Security (pp. 33-43).*
www.irma-international.org/chapter/hybrid-warfare/286716