

# Chapter 2

## Automating Child Protection: The Future of AI, ML, and Digital Safety

**Hitesh Rawat**

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

*University of Extremadura, Spain*

**Anjali Rawat**

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

*University of Extremadura, Spain*

**Prathamesh Muzumdar**

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

*Mangalayatan University, India*

**Shyam Gehlot**

*Shri Vaishnav Vidyapeeth Vishwavidyalaya, India*

**Chandrapal Singh Dangi**

*Manipal University Jaipur, India*

### ABSTRACT

*With increasing digital exposure, safeguarding children online is more important than ever. This research introduces an AI-driven framework, ChildGuard-CNN, designed to detect and prevent harmful online content targeting children. The system combines Word2Vec embeddings with a Convolutional Neural Network (CNN) to analyze and classify digital content. Training and validation were conducted using two public datasets: the 2023 Cyberbullying Detection Dataset (45,000+ social media*

DOI: 10.4018/979-8-3373-2716-7.ch002

*posts) and the YouTube Open Dataset for Content Moderation (60,000 instances of metadata flagged for safety concerns). Word2Vec transforms text into semantic vectors, which are processed by the CNN for feature extraction and classification. The model achieved 95.37% accuracy, with 94.89% precision, 95.78% recall, and a 95.33% F1-score. It also supports real-time detection with an average processing time of 1.3 seconds. Future improvements aim to enhance multilingual support and broaden platform compatibility for greater reach and effectiveness.*

## **1. INTRODUCTION**

With the rapid expansion of digital ecosystems (Krutzinna, 2024), children are increasingly exposed to a broad spectrum of online risks, including cyberbullying, exposure to explicit content, online grooming, and psychological manipulation. Reports by UNICEF (2024) and the Internet Watch Foundation (IWF) highlight that over 60% of children aged 9–17 have encountered harmful or inappropriate content online. Additionally, cyberbullying incidents (Rawat et al., 2025) surged by 25% globally between 2021 and 2023, signaling an urgent need for intelligent digital safety frameworks.

Traditional parental control systems—relying on keyword filtering and manual settings (Faraz et al., 2024)—are no longer sufficient to handle the dynamic, contextual, and often concealed nature of harmful content. The advent of Artificial Intelligence (AI) (Singh and Nambiar, 2024) and Machine Learning (ML) (Bhardwaj et al., 2024) offers a transformative shift in content moderation and child protection strategies. Recent advancements in Natural Language Processing (NLP) (Dhawan et al., 2025) (Rajavat et al., 2024), particularly context-aware embeddings, neural classifiers, and sentiment analysis (Mascari et al., 2025), have shown strong potential for real-time risk detection (Nahar et al., 2023) (Pithawa et al., 2023).

### **1.1 Secure Guidelines for Kids to Protect from Cybercrime (Singh and Nambiar, 2024) (Faraz et al., 2024)**

With children's internet usage rising by 75% over the last five years (Statista, 2024), the threat of cybercrime targeting young users is at an all-time high. Research shows that 1 in 3 internet users worldwide is under 18 (UNICEF, 2023). Therefore, teaching kids' cybersecurity essentials is more important than ever. Here are essential secure guidelines.

- a) Never Share Personal Information (0% Public Disclosure) (Rawat and Rajavat, 2024a)

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/automating-child-protection/384731](http://www.igi-global.com/chapter/automating-child-protection/384731)

## Related Content

---

### Methodology

(2026). *Miscarriage in Higher Education: Systemic Harms, Accountability, and Institutional Response* (pp. 27-36).

[www.irma-international.org/chapter/methodology/408365](http://www.irma-international.org/chapter/methodology/408365)

### Challenges and Ethical Considerations in Bookless Libraries: Digital Divide, Data Privacy, and Content Authenticity

N. Rajkumar, C. Viji, Mahmood Hussain Mir, A. Mohanraj, Jayavadivel Raviand Balusamy Nachiappan (2026). *Book Bans, Classroom Restrictions, and the Fight for Educational Access* (pp. 359-388).

[www.irma-international.org/chapter/challenges-and-ethical-considerations-in-bookless-libraries/412739](http://www.irma-international.org/chapter/challenges-and-ethical-considerations-in-bookless-libraries/412739)

### Transforming Sanitation and Hygiene Behaviour of Basic-Level Students Through Participatory Classroom Pedagogy in a Public School in Nepal

Govinda Prasad Devkota, Bhimsen Devkota, Manoj Kumar Pandeyand Shyam Krishna Maharjan (2023). *Implementing Transformative Education With Participatory Action Research* (pp. 141-155).

[www.irma-international.org/chapter/transforming-sanitation-and-hygiene-behaviour-of-basic-level-students-through-participatory-classroom-pedagogy-in-a-public-school-in-nepal/329917](http://www.irma-international.org/chapter/transforming-sanitation-and-hygiene-behaviour-of-basic-level-students-through-participatory-classroom-pedagogy-in-a-public-school-in-nepal/329917)

### The Role of Technology in Environmental Governance

Manas Kumar Jha, Dilip Kumar Markandey, Ruchi Guptaand Pranavi Mishra (2025). *Intersecting Environmental Governance With Technological Advancements* (pp. 287-322).

[www.irma-international.org/chapter/the-role-of-technology-in-environmental-governance/363225](http://www.irma-international.org/chapter/the-role-of-technology-in-environmental-governance/363225)

### Reverse Logistics and Solid Waste: Challenges for the National Waste Policy (PNRS) in Brazil

Hermes de Andrade Júnior (2020). *Developing Eco-Cities Through Policy, Planning, and Innovation: Can It Really Work?* (pp. 84-113).

[www.irma-international.org/chapter/reverse-logistics-and-solid-waste/236491](http://www.irma-international.org/chapter/reverse-logistics-and-solid-waste/236491)