


# Chapter 15


## Empowering Child Safety Using AI and ML to Address Online Risks

**Seena Thomas Kaithathara**

 <https://orcid.org/0000-0002-8503-1468>

*King Abdulla International Medical Research Center, Riyadh, Saudi Arabia*

**Amrutha Jose**

 <https://orcid.org/0000-0002-1192-2217>

*ICMR-National Institute of Immunohaematology, Mumbai, India*

### **ABSTRACT**

*Children face escalating online risks, including cyberbullying, grooming, and privacy violations. This chapter examines AI/ML's role in mitigating threats through real-time detection of harmful content and enhancing safety measures like behavioral analysis and adaptive moderation. Analysis of a cybersecurity behavior reveals gaps in children's risk awareness, emphasizing the need for targeted interventions. While AI/ML offers scalable protection, ethical challenges, algorithmic bias, privacy trade-offs, and over-reliance on automation, highlight risks of censorship and eroded human agency. A balanced framework is proposed, integrating AI tools with policy reforms and digital literacy programs. Collaborative governance among technologists, educators, and policymakers is urged to prioritize child-centric design, transparency, and ethical safeguards. This approach harmonizes innovation with education, fostering safer digital ecosystems without compromising children's autonomy or developmental needs.*

DOI: 10.4018/979-8-3373-5132-2.ch015

## **INTRODUCTION**

### **Overview of the Growing Online Risks for Children**

Children today face an increasingly complex digital landscape where online risks such as cyberbullying, grooming, exposure to inappropriate content, and exploitation are on the rise. Cyberbullying, defined as intentional and repeated harmful behavior via digital platforms, affects nearly 37% of young people globally, with severe consequences for mental health, including anxiety, depression, and even suicidal ideation (UNICEF, 2021). Additionally, predatory grooming—where offenders build trust with minors to exploit them—has surged with the growth of social media and gaming platforms. Reports indicate that 1 in 5 children encounter unwanted sexual solicitations online (ECPAT International, 2022). Meanwhile, algorithms on video-sharing and social media platforms often expose children to violent, extremist, or sexually explicit content, despite age restrictions (Livingstone et al., 2023). The commercialization of child data and the rise of AI-generated deep fakes further exacerbate privacy and exploitation risks, making robust protective measures essential.

The convergence of anonymity, widespread internet access, and advanced digital tools has amplified threats like sextortion, where children are blackmailed with intimate images, and child sexual abuse material (CSAM) distribution via encrypted channels. According to the National Center for Missing & Exploited Children (NCMEC, 2023), reports of online child exploitation have increased by 90% since 2020, with AI-generated synthetic media complicating detection efforts. Furthermore, gaming and metaverse environments introduce new vectors for exploitation, as predators exploit in-game chats and virtual interactions (Interpol, 2023). While legislation like the UK's Online Safety Act (2023) and the EU's Digital Services Act (DSA) aim to enforce platform accountability, the rapid evolution of threats necessitates adaptive AI-driven solutions to safeguard children effectively.

### **The Dual Role of Technology in Children's Online Risks**

Technology has played a paradoxical role in both amplifying and mitigating online risks for children. On one hand, the widespread adoption of social media, gaming platforms, and instant messaging has expanded opportunities for cyberbullying, predatory grooming, and exposure to harmful content. Algorithms designed to maximize engagement often inadvertently promote extreme or inappropriate material, while end-to-end encryption and anonymous platforms enable predators to operate with reduced detection (Livingstone & Stoilova, 2021). Additionally, emerging technologies like deepfake AI and virtual reality pose new threats, such as

36 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/empowering-child-safety-using-ai-and-ml-to-address-online-risks/386107](http://www.igi-global.com/chapter/empowering-child-safety-using-ai-and-ml-to-address-online-risks/386107)

## Related Content

---

### A Conceptual Methodology for Dealing with Terrorism "Narratives"

Gian Piero Zarri (2010). *International Journal of Digital Crime and Forensics* (pp. 47-63).

[www.irma-international.org/article/conceptual-methodology-dealing-terrorism-narratives/43554](http://www.irma-international.org/article/conceptual-methodology-dealing-terrorism-narratives/43554)

### Development of Secured Log Management System Over Blockchain Technology

Sagar Shankar Rajebhosale and Mohan Chandrabhan Nikam (2019). *International Journal of Cyber Research and Education* (pp. 38-42).

[www.irma-international.org/article/development-of-secured-log-management-system-over-blockchain-technology/218896](http://www.irma-international.org/article/development-of-secured-log-management-system-over-blockchain-technology/218896)

### A Comparison of Cyber-Crime Definitions in India and the United States

Himanshu Maheshwari, H.S. Hyman and Manish Agrawal (2011). *Cyber Security, Cyber Crime and Cyber Forensics: Applications and Perspectives* (pp. 33-45).

[www.irma-international.org/chapter/comparison-cyber-crime-definitions-india/50712](http://www.irma-international.org/chapter/comparison-cyber-crime-definitions-india/50712)

### A Novel IDS Securing Industrial Control System of Critical Infrastructure Using Deception Technology

Shaobo Zhang, Yuhang Liu and Dequan Yang (2022). *International Journal of Digital Crime and Forensics* (pp. 1-20).

[www.irma-international.org/article/a-novel-ids-securing-industrial-control-system-of-critical-infrastructure-using-deception-technology/302874](http://www.irma-international.org/article/a-novel-ids-securing-industrial-control-system-of-critical-infrastructure-using-deception-technology/302874)

### Empowering Child Safety Using AI and ML to Address Online Risks

Seena Thomas Kaithathara and Amrutha Jose (2026). *Child Protection Laws and Crime in the Digital Era* (pp. 361-398).

[www.irma-international.org/chapter/empowering-child-safety-using-ai-and-ml-to-address-online-risks/386107](http://www.irma-international.org/chapter/empowering-child-safety-using-ai-and-ml-to-address-online-risks/386107)