UDC
DOI:

GENERATIVE AI AND THE CYBERBULLYING OF CHILDREN IN INDIA: LEGAL GAPS, RISKS, AND PRACTICE-BASED POLICY SAFEGUARDS

Raahul JAIN

Ph.D. Research Scholar, School of Law, Bennett University, India
E-mail: rahuljain.bennett@gmail.com

Gyandeep CHAUDHARY

Assistant Professor, School of Law, University School of Legal Studies,
G.G.S. Indraprastha University, Delhi, India
E-mail: gyan.2889@gmail.com

Abstract

The rapid evolution of generative artificial intelligence (Gen-AI) has fundamentally transformed digital interactions, while simultaneously intensifying the scale and complexity of cyberbullying against children. In India, where digital adoption among youth is rapidly expanding, Gen-AI has enabled new forms of online harm, including deepfakes, AI-driven impersonation, misinformation, and the generation of exploitative content. Despite a growing body of literature on cyberbullying and AI, the literature on their intersection within the Indian legal and policy framework remains limited. This study adopts a qualitative doctrinal methodology supplemented by a comprehensive overview of key legislative frameworks, including the Information Technology Act (IT), 2000, Information Technology (Intermediaries Guidelines and Digital Media Ethics Code) 2021, the Protection of Children from Sexual Offences Act (POCSO), 2012, and the Bharatiya Nyaya Sanhita (BNS), 2023. It also evaluates secondary literature to examine both normative structures and practical enforcement challenges. In addition, the study undertakes a comparative assessment of regulatory approaches in jurisdictions such as the European Union, the United Kingdom, the United States and Australia to situate India within emerging global governance trends. The findings reveal a significant regulatory gap, as existing legal frameworks remain inadequately equipped to address AI-driven cyberbullying, particularly in relation to synthetic content,

anonymity, and algorithmic amplification. Structural challenges such as weak enforcement, limited digital literacy, and a lack of AI-specific safeguards further exacerbate children's vulnerability, resulting in psychological harm, social exclusion, and long-term developmental consequences. This study contributes to the emerging field of AI and child protection by (i) offering a conceptual integration of Gen- AI risks with cyberbullying frameworks, (ii) providing a comparative legal critique of existing regulatory regimes, and (iii) proposing a multi-layered, child-centric policy framework that combines legal reform, platform accountability, and ethical AI governance.

Keywords: *Gen- AI, Cyberbullying, Child Protection, Deepfakes, Digital Safety, AI Regulation, India*

1. Introduction

Since its introduction in November 2022, ChatGPT has brought AI to a higher global stage, especially as Gen-AI rapidly advances. Gen- AI has completely redefined discussions of AI and its societal impact in a remarkably short period. With the rapid advancement and widespread accessibility of Gen-AI tools, it is now easy to generate materials such as text, images, audio, speech, code, 3D model data, and video from simple text inputs (Gupta et al., 2025)

Children worldwide are progressively being introduced to AI to harness its capabilities, primarily for educational and entertainment purposes (Reeder & Lee, 2024). AI-driven educational tools, chatbots, and virtual tutors provide personalised learning experiences, with content, exercises, and feedback tailored to enhance academic outcomes (Zhou & Hou, 2024). Nevertheless, when the advantages of AI for children's learning experiences are considered, the technology still carries an inherent risk of abuse (O'Higgins Norman, 2020). The prominent threat to the harmful use of this Gen AI is the creation and distribution of extreme child sexual exploitation and abuse material (CSEAM), including cyberbullying, hate speech, misinformation and disinformation (Mishna et al., 2021). The Internet Watch Foundation (IWF) released a report in October 2023 revealing the rise of AI-facilitated CSEAM, which is impeding efforts to address it.

A total of 20,254 AI-generated images were posted on a single extremely dark web CSEAM forum in a single month. In their 2023 annual report, the IWF team noted a 22% increase in web pages containing CSEAM sites, with 51,369 URLs identified in 2022 (Internet Watch Foundation, 2023). These figures show that, in many cases, AI can produce content inappropriate for children, despite measures in place to protect them. Likewise, addressing the

implications of Gen-AI applies to all sectors. When engaging together, governments, technology sectors, teachers, parents, law enforcement, international organisations, and United Nations agencies need to collaborate to generate synergy and develop strategies and/or policies to address the negative consequences of Gen-AI.

Several existing studies have explored the application of Gen-AI among children (He & Lu, 2024), pathways of cyberbullying among children (DePaolis & Williford, 2019), levels, patterns and predictors of cyberbullying among children (Ranjith et al., 2023), and consequences of cyberbullying on children (Mishna et al., 2016). At the same time, a growing body of literature has highlighted the need for effective laws and policies to protect children from the adverse effects of AI, given its rapid evolution (Badawy, 2025; McStay & Rosner, 2021). However, despite this growing body of literature, there remains limited research that systematically explores the intersection between Gen-AI and cyberbullying of children**,** particularly within the Indian context.

The disparity in this regard is concerning, particularly for India, which, in addition to having the most significant number of young people in the world, also has one of the fastest rates of digital technology adoption among its youth. Various laws, such as the IT Act (2000), the POCSO Act (2012), and the Bharatiya Nyaya Sanhita (2023), have acknowledged that online harms take many forms; they have not gone into detail regarding AI. This lack of consideration puts children's lives at risk from exploitation and abuse through new AI technologies in India.

The study aims to address this gap through a systematic literature review. Its purposes are first, to explore how Gen AI influences children's lives, including both positive and negative aspects; second, to examine the absence of laws and policies that can effectively address AI-driven cyberbullying; and third, to develop feasible, child-focused recommendations for protecting the digital environment. By exploring Gen AI as a central topic in discussions of child protection and digital safety, this study makes a meaningful contribution to the policy debate. It examines how technological innovation can contribute to children's well-being when properly managed.

1.1 Defining child

Article 1 of the ‘Convention on the Rights of the Child’ defines a child as a person below the age of 18 years, except in cases where a particular state has legally defined a lower age at which an individual is considered an adult (UN General Assembly, 1989). The Committee on the Rights of the Child, which monitors the convention, has urged states to revise the age of majority, which is currently 18 years, and to provide better protection for persons below that age. In India, various legislations, including the Juvenile Justice Care and Protection of Children Act, 2015 and the POCSO Act, 2012, define a child as

someone under the age of 18. The Committee on the Rights of the Child has further emphasised the need for States to strengthen safeguards.

2. Literature review

2.1 Gen-AI

Gen- AI is a means by which machine learning creates entirely new digital content of various kinds, such as text, images, audio, video, and multimodal experience simulations (Table 1). This branch of machine learning is often referred to as AI. Hence, Gen-AI is a technology that generates content in response to user-assigned prompts (Chowdhury & Lakshmi, 2023). The primary distinction between Gen-AI and other AI forms is that Gen-AI can generate entirely new outputs, whereas other AI forms typically make predictions or classify inputs (Goodfellow et al., 2014). Although Gen-AI models utilise human-like parameters that serve as guidelines or constraints to determine how tasks are completed or produced, the neural networks underlying them have been trained on massive datasets (Bommasani et al., 2021). These algorithms employ probabilistic techniques to synthesise new instances that closely resemble the original data, with some behaviours extending even beyond the explicitly defined parameters (Brown et al., 2020). The advent of large learning models (LLMs) and diffusion models gave rise to Gen-AI. LLMs were trained to understand the relationship between words. They are thus employed to process and generate human-language text, and large language datasets are used for training. (Malsia & Loku, 2024). In contrast, diffusion models excel at synthesising images and videos from random noise, having been trained on vast datasets of images and videos (Reznikov, 2024).

Gen-AI has both advantages and disadvantages. It provides many opportunities and creative avenues in the arts, music, and design (Otis et al., 2024). However, this technology has also been exploited for malignant effects, from committing fraud via cyberbullying to creating massive amounts of hateful false content and disseminating disinformation and misinformation (Grassini, 2024; Milosevic et al., 2023). Children are increasingly turning to AI technologies for educational and entertainment purposes. AI educational tools, bots, and virtual teachers offer personalised learning through customised content, tasks, and feedback for optimal learning. However, the usual misuse of Gen-AI can lead to the generation, and dissemination of AI-generated CSEAM, cyberbullying, hate speech, and the propagation of misinformation and disinformation. (Milosevic et al., 2023).

Text generation	The Gen-AI model generates text responses to user prompts or questions. Examples: ChatGPT, Google Bard (Gemini), Jasper Chat, Perplexity, Claude
Image generation	The Gen-AI model generates highly realistic, high-quality images in response to text prompts. Examples: Dall-E, NightCafe, Midjourney
Video creation	The Gen-AI model generates realistic videos of scenarios described by the user. It can also generate videos from user-provided images. Examples: DALL-E 2, Synthesia, descriptive, expression, Stable Diffusion,
Voice generation	A Gen- AI model generates speech from text or speech prompts. Examples: LOVO, Synthesys, Replica Studios, VALL-E

Table 1. Major Categories of Gen-AI and Their Applications

2.2 Cyberbullying

Cyberbullying refers to various forms of intentionally harmful behaviours that can include offensive messages, posts, comments, and other activities on digital platforms (Hinduja & Patchin, 2015). Cyberbullying is typically a repeated behaviour, but a one-off post that can be viewed and re-shared by more people can also be cyberbullying. Cyberbullying often involves a power imbalance, where the perpetrator holds some form of advantage over the victim. This might involve physical strength in offline settings, but the online definition is more complex (O'Higgins Norman, 2020). This imbalance can stem from greater digital skill in bullying or from possessing greater social capital, such as popularity or a larger number of followers (Kowalski & McCord, 2020; Smith, 2016). However, even youths with significant social capital, such as influencers, may become targets of perpetrators. This makes it difficult to consistently apply the criteria for an unbalanced power dynamic. Perpetrators also take advantage of being anonymous via usernames as a major contributor to the prevalence of cyberbullying. Despite this, studies have indicated that cyberbullying frequently occurs within the context of offline relationships, such

as in school environments, where victims are often aware of their bullies' identities (Mishna et al., 2021).

Prioritising the moderation of bullying content based on the number of "views" is a notable example that highlights the types of scenarios in which cyberbullying is challenging to address (Bickert, 2020). For instance, when a company determines that bullying reports involving a larger audience or greater public visibility should be treated as a higher priority, it presupposes that such incidents necessarily cause greater harm than cases witnessed by only a few individuals. Consequently, incidents involving limited visibility may become underprioritized, even where the affected child is experiencing severe psychological distress or requires immediate care and intervention. Such an approach is problematic because the seriousness of bullying should not be assessed solely based on audience size or public exposure, but also on the potential emotional, mental, and physical impact upon the child concerned, particularly in situations that may be time-critical and demand urgent protective measures. The child may incur extensive damage in the event of the set case. Cyberbullying is typically prohibited on social media platforms, and companies stipulate this in their policies (Gillespie, 2018). Keeping in mind the vast amount of cyberbullying content on platforms, social media are struggling to moderate or process cyberbullying cases, and they are increasingly relying on AI or algorithmic tools intended to help automate the task of moderation, which leverages natural language processing (NLP), machine learning (ML), and deep learning (DL) (Gorwa et al., 2020). Social media users, for example, can inform a specific platform about cyberbullying by dealing with the issue after it has taken place, and AI is gradually being implemented for a proactive engagement in checking and supervising content that has not yet been reviewed. This process is detailed in the transparency reports of some large companies, which report the number or percentage of detected and proactively removed bullying content (Milosevic et al., 2022).

2.3 Risks and negative impacts of Gen-AI in children

Although Gen-AI offers many advantages to education, the economy, and technology, the potential risks to children cannot be ignored. The following are some ways in which children may be harmed by using Gen-AI, such as deepfakes, the creation of sexualised synthetic content, impersonation, and indirect consequences related to the technology's limitations, including inaccurate output, the fabrication of data or information, and the amplification of misinformation or disinformation. (INHOPE, 2022)

Increasingly, there is consensus among educational institutions and child safety stakeholders about the potential long-term, detrimental effects of digital technology (OECD, 2025; UNICEF Innocenti, 2025). Digital technology presents unique challenges for children and adolescents because they often use

it without fully understanding the risks. Therefore, developing a strategy for managing exposure to digital technology requires implementing additional precautions to support younger users.

Research has also shown the difficulties that parents, teachers and caregivers face due to their limited knowledge of digital literacy, technology or supervision tools to protect children online. A lack of adult understanding of how children use AI systems, social media, and the internet will lead to poor-quality parental guidance and school supervision (Yu et al., 2025).

A further limitation concerns communication between trusted adults and children in both home and school contexts. If there is a lack of communication, the child will be more likely to be at risk. Many children share devices or log in to their device(s) using the same login information as other family members, but do not have a specific child-safe profile on the device. Hence, they may engage in online activity such as browsing, and their parents may not be aware of what they are doing, and be unable to supervise them, for example, when the child clears their browsing history, or uses a platform discreetly. Therefore, children may be exposed to inappropriate, misleading, or dangerous content through Gen-AI technologies (Lakatos et al., 2023).

Understanding how children interact with Gen-AI is therefore essential to analysing the specific harms discussed in Sections 2.3.1 to 2.3.5, which examine the relationship between Gen-AI technologies, child safety, and emerging digital threats.

2.3.1 Misuse of Deepfakes

Deepfake is a term that combines the words deep learning and fake, meaning that AI uses deep learning to create false representations of what a human has actually said or done. This can take the form of synthesised audio, image or video. The term gained prominence in 2017 following the online spread of AI-generated manipulated media, particularly non-consensual sexual content. Since then, rapid advances in Gen-AI have significantly heightened concerns about privacy, misinformation, identity misuse, and digital safety. Deepfake systems typically analyse a target's facial features, voice patterns, gestures, and expressions to generate realistic synthetic media. Detection is increasingly difficult because modern tools employ high-resolution rendering, facial mapping, voice cloning, and masking techniques (Chakrabarty & Chattopadhyay, 2024).

Deepfakes are often associated with blackmail, humiliation, harassment, and damage to reputations, and children and youth in particular, teenage girls and young women may be particularly at risk as a result of these negative consequences. A case was reported in August 2025 in Howrah, West Bengal,

where two Class X students were suspected of using AI tools to produce synthetic pornography by morphing images of their female classmates and distributing them through the internet (Times of India, 2025). This incident led to public outrage, protests, and calls to file a First Information Report (FIR). This incident indicates the extent of psychological and social damage inflicted by synthetic sexualised images on minor victims.

The Global Trends also reflects Misusing AI, such as public figures like Taylor Swift are vocalising their frustrations with both the unauthorised AI-generated explicit images of them, as well as the use of their identity through deepfakes violating their rights, which will ultimately hurt the lives of everyday people, who will be affected by abusive deepfake activity (Reuters, 2025; Time, 2024). Unfortunately, the negative impact of deepfake abuse will result in even worse consequences for children, such as cognitive trauma, bullying, damage to their reputations, and long-term emotional distress.

As deepfake tools become more widely available, demands for increased accountability from platforms on which they are hosted, timely removal of content, digital literacy instruction, and better regulations for synthetic sexual abuse are on the rise.

2.3.2 Impersonation: Gen- AI Fake Profiles and Online Deception.

The use of AI-powered fake accounts for malicious purposes is considered a major threat to social media users. Using Gen-AI tools, users can create highly realistic profile pictures, biographical information, chat responses, and behaviours to make a fake account seem legitimate. These accounts can be used for identity theft, impersonation, deception, harassment, and the dissemination of false information.

According to Goyat (2015), the negative impacts associated with false or impersonating accounts are illustrated by a reported incident that has taken place at the Panjab University, Chandigarh, in which an account was created on Instagram in the name of a student and pictures of them in their private moment were shared without consent. This experience has been described as violating the victim's privacy through humiliation and distress, and was present prior to the introduction of the current generation of AI technology, but provides an example of the form of abuses that can be perpetuated using AI technology and that are likely to have longer-lasting impacts than what has occurred in previous years.

Due to their frequent online use, socially adventurous nature, and less familiarity with impersonation strategies, children and adolescents may be especially at risk of experiencing these negative effects.

In addition, research indicates that numerous fake social media accounts are associated with political manipulation, conspiracy theories, fraudulent schemes, generating spam, and organised efforts to influence public opinion. Investigators have reported a significant increase in the number of internet accounts created by AI. Most of these accounts are expected to have been created within hours using automated services, raising new concerns about digital trust, the integrity of online platforms, and online security (Yang et al, 2024).

A related challenge is the growing difficulty in distinguishing between real and artificial content. For example, AI-generated fraudulent posts and fake endorsements may be disseminated via large networks such as Facebook or Instagram to trick people into visiting harmful websites or releasing personal data. As AI-generated content generation capabilities grow, it becomes much harder to tell whether your online identity is real or fabricated.

2.3.3 Gen- AI for Creating Child Sexual Exploitation and Abuse Material (CSEAM)

New risks in relation to the production, proliferation, and normalisation of CSEAM have emerged as a result of Gen-AI. The Supreme Court of India made a ruling regarding the use of the term '*child sexual exploitation and abuse material*' instead of outdated terms that downplay the abusive nature of such content in September 2024. The Court further clarified that sexually explicit material depicting children, whether produced using AI tools or digitally altered to appear as if it depicts an actual child, may result in criminal charges irrespective of the method used to create the material (Just Rights for Children Alliance, 2025).

Recent data show an increase in incidents involving such material. According to NCRB (National Crime Records Bureau) statistics, the number of cases increased from 88 in 2018 to 1,902 in 2023, indicating a drastic increase in reported incidents. According to the U.S.-based National Center for Missing and Exploited Children (NCMEC), there are millions of reported suspected electronic transmissions involving this material associated with India (Tyagi, 2025). The increased number of reports and detections of such material may explain the rapid rise; however, the magnitude of the increase raises serious doubt about more significant digital exploitation than previously accounted for.

There are various ways to use Gen-AI to construct a synthetic CSEAM. First, by training models on actual or unlawful exploitative materials and developing models to create child-like exploitative images. Second, by using adult sexual content, which may be dependent on the use of prompts, fine-tuning, or age-

regression models to create child-like exploitative images. Therefore, the composition and filtering of training datasets are critical in both scenarios.

A well-known example was reported in December 2023 by the Stanford Internet Observatory, who found that Stable Diffusion 1.5 had the capability of producing erotic or sexually explicit pictures of children due to being trained on LAION-5B - a very large open-source dataset created by performing a massive number of automated crawls of the Internet (Thiel, 2023), which may have inadvertently included images of children who were abused or depicted in sexual ways. The LAION-5B dataset has been reported to have been used to train multiple publicly available image generators, thereby raising concerns about child safety due to weaknesses in data governance (Beaumont, 2022; Levine, 2024a).

The appearance of synthetic CSEAMs creates substantial hurdles for law enforcement agencies engaged in combating child exploitation crimes. Even where a victim was not actually photographed, such commercially produced materials may support the sexual interest in children; promote community forums that serve as a vehicle for grooming behaviours; lead to re-traumatisation of child victims through the graphical manipulation of their image; and hinder investigative processes. Consequently, this highlights a requirement to audit datasets for accuracy immediately, implement robust safety filters, develop hash-matching methods, hold platforms accountable for reporting, and introduce clear legislative frameworks to regulate the use of AI in the creation of exploitative content.

2.3.4 AI-Generated Misinformation and Disinformation

Gen AI technology is helping create authentic-looking but fabricated or manipulated media that may potentially mislead and exploit children, increasing their susceptibility to deceit, bullying, and reputational damage. Over the past decade, children have been the target audience of social media platforms, increasing the speed and reach of this form of communication. Therefore, the existence of Generated Media poses a serious and emergent danger to users' overall safety, perceptions, and decision-making capabilities. In 2024, an incident in Moradabad, Uttar Pradesh, exemplified the potential for AI to be misused in the production and dissemination of obscene, manipulated photographs of a female instructor captured on video for social media; the two young men involved were in the 9th grade. An FIR was filed under the applicable subsections of the IT Act, 2000. This incident demonstrates how the abuse of AI can result in targeted harassment and represent an existential risk to personal privacy and dignity, as well as psychological harm to others. (NDTV, 2024; JustAI, 2024).

In addition to this particular situation, the episode highlights a broader danger posed by AI-generated false or manipulated content: damage to one's reputation, emotional distress, or the potential to provoke panic or hostility. The increasing number of children using online platforms for news, communication, and information may make it difficult for younger users to distinguish authentic from false or fabricated media as Gen-AI content becomes more prevalent, thereby heightening concern about misinformation and disinformation (Tamboer et al., 2022).

2.3.5 AI Bias and Discriminatory Content

Gen- AI systems tend to be opaque, but they can still create large-scale exclusionary or discriminatory outcomes. Children are particularly susceptible to these consequences, as bias can create a digital environment that is biased from a very young age. When AI systems employ unrepresentative training data, utilise a non-contextual understanding, or generate automated recommendations without meaningful human involvement, those systems tend to perpetuate negative stereotypes and discriminatory treatment. Children and youths in minority communities, especially, are disproportionately affected by these outcomes, which may limit their opportunities and negatively impact their self-worth and their future development. (Hitanshi & Chaudhary, 2024).

Gen -AI has created tremendous opportunities for India's economy and its ongoing digital transformation. However, ethical considerations regarding the use of Gen- AI, given the lack of an overarching legislative/regulatory structure governing AI, will be significantly heightened. AI-driven bias and misinformation may adversely affect education, recruitment, and online service platforms. Several scholars have argued that many AI models are insufficiently adapted to India’s socioeconomic and cultural realities because they are primarily trained on data reflecting Western contexts. Research by the Indian Council for Research on International Economic Relations (ICRIER) has indicated that AI-generated outputs may reflect gender, caste, linguistic, and regional biases. In a country as socially and culturally diverse as India, the lack of inclusive data representation risks reinforcing existing inequalities unless effective safeguards are introduced (Sood, 2023).

The preceding discussion has examined the principal forms of misuse that Gen-AI may facilitate, including exploitation, deception, manipulation, and discriminatory harm in digital environments. However, the significance of these developments cannot be confined to the immediate modalities of abuse alone. Gen- AI also has wider implications for children’s social experience, developmental trajectory, psychological well-being, and future opportunities. It is therefore necessary to examine the broader impact of Gen AI on children.

2.4 Impact of Gen-AI on Children

2.4.1 Impact on jobs

The growing adoption of Gen-AI technologies raises concerns about job displacement, particularly in artistic fields such as content writing, photography, and graphic design. Publicly, easy access to such tools and platforms has also raised questions about the credibility and trustworthiness of creative professionals' capabilities. (Bastian, 2024)

2.4.2 Impact on cognitive development

The impact of Gen-AI on children's cognitive development can, on the one hand, be viewed as fruitful for education. On the contrary, it can also be detrimental due to its various risks. Personalisation of learning through AI can meet individual needs. It can promote interactions that help individuals become better critical thinkers, problem solvers, and even more creative thinkers.

Excessive dependence on AI technologies may hinder cognitive skill development, making it challenging for a child to perform tasks independently. This can lead to a loss of human interaction and impede the development of social skills, empathy, and communication. This may hinder opportunities for problem-solving by providing ready-made answers, and children do not learn to question, analyse, and examine the forms that influence their judgments (Kanders et al., 2024).

Advances in Gen-AI technologies have raised concerns that jobs will be displaced, particularly in hands-on creative fields such as content writing, photography, and graphic design. There are questions about the honesty and integrity of creative professionals when such tools and platforms are made publicly available. This situation raises concerns about the long-term influence of AI, which could have a considerable impact on younger generations. This may lead to the disappearance of certain professions and necessitate new types of work.

2.4.3 Psychological effect

Technology, particularly Gen-AI in the form of chatbots and virtual companions, can be so engaging that children spend excessive time with it, losing touch with real-world relationships and responsibilities. Such interactions may harm mental health (Zsila & Reyes, 2023).

The advent of AI companions that do not require human presence, thereby reducing human-to-human interaction and hindering the development of interpersonal skills, Emotional Intelligence and other aspects of Social Justice. On that continuum, immersive AI-enabled virtual environments may make

confident children interested in interactions with AI less likely to socialise with their peers, which may adversely affect their social skills and relationships. Increased screen time and reliance on AI devices can also reduce physical activity, leading to weight gain, eye strain, and other health issues. In addition, Gen-AI reflects user preferences and thus provides similar content to engage users, which might shorten children's attention spans and influence their cognitive development. These Gen-AI tools may prompt existential questions and uncertainty about the value of human effort, and destabilise psychological well-being.

The foregoing discussion has shown that the effects of Gen AI on children extend beyond technological novelty and encompass questions of safety, dignity, development, and future opportunity. These consequences make clear that the challenges posed by AI-driven harms are not merely social or educational concerns, but also legal ones. It therefore becomes necessary to examine the legal framework governing Gen-AI-driven cyberbullying against children.

3. Legal Framework of Gen-AI-driven Cyberbullying Against Children

The IT Act 2000 and its 2008 amendments have established a legal framework to address the risks posed by next-generation AI-driven technologies, including those related to crime, cyberbullying, and child exploitation via deepfakes (see Table 2). Section 67 criminalises the publication of obscene content through electronic means while extending this application to AI-generated content designed to deceive, harass, or otherwise manipulate a person. More safeguards are provided in section 67A, which forbids any sexually explicit content and states that deepfake pornography or obscene AI-generated impersonations of real or fictitious persons will be treated as serious offences. However, Section 67B, which was introduced in 2008, defines child pornography. It prohibits the creation, retention, sharing, or distribution of any CSEAM concerning a minor, in any form, including those generated by AI. In addition, this provision provides punishment for grooming, extortion, and predatory behaviour enabled by AI for child abuse. (Information Technology Act, 2000).

The IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, have been promulgated under the Information Technology Act, 2000, to regulate a subset of online intermediaries, particularly social media platforms. These regulations have been framed to address concerns regarding the abuse of Gen-AI by deepfakes, cyberbullying, and child exploitation. The purpose of these regulations is to create an online environment that is safe, responsible, and accountable by requiring proactive content moderation and enforcement. Rule 3(1)(b) requires intermediaries to specifically define prohibited content, including defamatory, obscene, harmful, or privacy-invasion material, to alert

online platforms to such issues and to implement timely and effective prevention methods. This includes deep fake harassment perpetrated with the help of AI and any online abuse against children. Rule 3(1)(d) orders an intermediary to remove illegal content within 36 hours of receiving actual knowledge or direction from a legitimate authority. Rule 3(2) provides the mechanism for grievance redressal, requiring every intermediary to appoint a grievance officer to address complaints regarding AI-enabled cyberbullying of a user within 24 hours and to resolve such complaints within 15 days. Rule 4(4) recognises the augmentation of digital risk by AI and requires significant social media intermediaries (SSMIs) to have automated tools to process CSEAM and to delete deepfake pornography and other AI-generated content. Rule 4(7) encourages voluntary account verification to counter the proliferation of fake identities used for impersonation or fraud. (Intermediary Guidelines and Digital Media Ethics Code Rules, 2021).

The POCSO Act, 2012, is a progressive and consolidated legal statute that potentially provides maximum protection against sexual abuse and exploitation. While the ambit of the Act is limited to physical offences, it does provide recourse against cyberstalking or cyberbullying committed with the aid of AI, especially in matters of child abuse or sexual exploitation, when offensive material comes within the realm of sexual abuse and exploitation. Section 11 includes any sexual gesture, act, or presentation of a child through the media for sexual gratification, punishable by the Act, or through AI in the process of production or distribution of child pornography. Similarly, Section 13 criminalises the use of a child for pornographic purposes, which extends to AI-generated CSEAM. (Prevention of Children from Sexual Offences Act, 2012).

Although the BNS does not mention Gen-AI, its general provisions apply to offences committed using AI-generated content. Section 75 of BNS prescribes punishment for sexual harassment with the use of AI-generated deepfake images or videos against a woman of any age, intended to demean or humiliate her. Section 78 constitutes cyberstalking when using deepfake techniques to threaten, harass, or extort a woman of any age. It also includes defamation perpetrated through AI, which affects a woman's dignity and reputation. Section 79 criminalises words, gestures, or acts that are aimed at insulting women's modesty and may also apply to AI-generated content used to abuse women of any age. Section 294 criminalises the sale, distribution, or circulation of obscene material, extending to AI-generated obscene images or videos of children. However, there are no clear provisions addressing AI-driven cyberbullying, AI-generated CSEAM, or platform-based exploitation, as the framework remains primarily content-centric. Besides, its gender-specificity in sections 75, 78 and 79 excludes male and transgender children, a key legislative loophole. As such, the BNS is a supplementary but vital reminder of the need for a comprehensive, child-centric, gender-neutral legal framework to mitigate

emerging forms of AI-driven cyberbullying against children. Section 95 of the BNS criminalises hiring, employing, or using a child to commit an offence, and its explanation includes the use of a child for sexual exploitation or pornography within its scope. However, it is primarily limited to offline child exploitation and does not explicitly address AI-generated CSEAM or other emerging forms of digital exploitation (Bharatiya Nyaya Sanhita, 2023).

While the legislative framework provides the formal legal basis for responding to AI-driven cyberbullying against children, the existence of legal provisions does not by itself ensure effective regulation in practice. The rapid evolution of Gen- AI, the cross-platform nature of online abuse, and the limits of existing enforcement mechanisms continue to expose significant gaps in the regulatory response. It is therefore necessary to examine the regulatory challenges involved in combating AI-driven cyberbullying against children in India.

4. Regulatory Challenges in Combating AI-Driven Cyberbullying Against Children in India

The present structure in India that addresses AI-based bullying of minors online is incomplete. Under Section 67C of the IT Act 2000, intermediaries must preserve and retain data for purposes of investigation. However, the Act imposes no specific requirements on them to retain data about AIs. Section 79 provides safe-harbour protection after due diligence has been exercised, although the principal statute remains largely neutral regarding AIs. MeitY has initiated the process of closing the gap through recent advisories and proposed 2025 regulations for "synthetically generated information", including requirements for labelling, metadata, and verification of the existence of synthetic data, such as the identity of the source of that data, by certain intermediaries. Currently, India's status can best be described as partially evolving rather than completely silent; however, there remains considerable uncertainty regarding enforcement, evidence preservation, and platform responsibility regarding deepfakes, impersonation, and AI-generated abuse. Jurisdiction is similarly problematic; while Section 75 of the IT Act provides for the extraterritorial application of the Act when an offence involving a computer resource located in India has occurred, enforcement faces major challenges in addressing offending content hosted outside India or on foreign or encrypted services (Information Technology Act, 2000).

Compared with newer platform-governance models abroad, the Indian framework remains less developed. The EU's Digital Services Act and its 2025 minors guidelines require a risk-based approach to children's safety, including measures to address grooming, cyberbullying, harmful design, age assurance, and safeguards for AI chatbots. The UK's Online Safety Act similarly imposes illegal-content and child-safety duties, and from 7 April 2026 requires in-scope user-to-user services to report detected and unreported CSEAM content to the

National Crime Agency. Australia's eSafety regime, established in 2015 and strengthened by the Online Safety Act 2021, now operates alongside a social-media minimum-age framework for certain age-restricted platforms. The United States has also moved beyond preliminary discussion through the TAKE IT DOWN Act, signed into law on 19 May 2025. However, it still lacks a comprehensive federal framework specifically focused on AI-driven child cyberbullying.

(**EU:** Regulation (EU) 2022/2065, 2022; European Commission, 2025.

UK: Online Safety Act 2023, 2023; Ofcom, 2026.

Australia: Online Safety Act 2021, 2021; Online Safety Amendment (Social Media Minimum Age) Act 2024, 2024.

US: TAKE IT DOWN Act, 2025)

In addition, socio-cultural aspects can affect how strongly a legal framework works in India as many people deeply entrenched societal stigma associated with AI-generated indecent visual representations of a child and this leads to the vast underreporting of crimes by victims and families based on fear or their feeling ashamed about reporting their experience due to fear of society's reaction (Prabhu et al., 2023; Singh et al., 2022). This is also made worse by both children and parents having a lack of digital literacy, which makes them more likely to be victims of AI-generated child sexual abuse material and other types of online exploitation (Agnihotri, 2024). Thus, offenders can exploit differences in access to information among people and many people's lack of technological savvy when using digital technology, leading to increased risk in the digital space.

Question	India	United States	United Kingdom	European Union	Australia

What are the key AI-driven cyberbullying issues that children are facing?	Deepfakes, identity theft, AI-generated CSAM, extortion, impersonation	AI-enhanced cyberstalking, deepfake pornography, and AI-driven harassment	AI-generated revenge pornography, cyber harassment, and AI-powered online abuse	AI-facilitated cyberbullying, deepfake child exploitation, and AI-generated misinformation	AI-driven cyberstalking, manipulated content, deepfake harassment
Are there legal frameworks for AI-driven cyberbullying against children?	IT Act, 2000**;** Intermediary guidelines 2021, POCSO Act 2016, BNS 2023	Children's Online Privacy Protection Act (COPPA), state-level laws (California, New York)	Online Safety Act, 2023**;** Communications Act, 2003	EU AI Act**,** Digital Services Act (DSA)**,** 2022, General Data Protection Regulation (GDPR), 2016	Online Safety Act, 2021**;** Cybercrime Act, 2001
What challenges exist in tackling AI-driven cyberbullying against children?	Lack of AI-specific laws, jurisdictional challenges, and slow enforcement	Free speech concerns, decentralised legal framework, and evolving AI threats	Content moderation loopholes and enforcement delays	Balancing AI innovation through regulation and enforcement consistency across member states	Delayed removal of AI-generated content, lack of legal clarity
How are technology platforms regulated to prevent AI-facilitate	The IT Rules (2021) impose responsibilities on intermediaries, but	FTC regulations and self-regulation by platforms such as Meta and Google	Ofcom regulates platforms and mandates proactive AI moderation	DSA requires large platforms to mitigate AI-related harm.	The e-Safety Commissioner enforced strict laws regarding the removal

d cyberbull ying?	enforceme nt is weak.				of AI content.
Are there initiatives to educate children about AI risks associate d with cyberbull ying?	Limited AI-specific Child Education Initiatives and National Cybersecur ity Awareness programs	K-12 AI literacy programs, school-based digital safety campaigns	Online Media Literacy Strategy and AI Safety Curricula in Schools	EU-funded AI education programs and safer internet campaigns	e-Safety programs and AI risk education in schools

Table 2. Comparison of the gap in combating AI-driven cyberbullying of children in India and other countries

5. Recommendations

5.1 Policy-Focused Recommendations

Legal reforms that have an immediate effect are necessary to deal with the risks that Gen-AI causes. Such changes to the IT (Intermediary Guidelines) Rules, 2021, should also explicitly address AI-generated activities, such as deepfakes, impersonation, and cyberbullying of children. From a medium-term perspective, it would be beneficial to amend Section 67C of the IT Act, 2000, to include an AI-specific log-retention requirement, which would significantly assist law enforcement in identifying and prosecuting offenders. Reforms in the distant future should aim to establish a centralised AI regulatory authority to enforce standards to protect children, oversee technology providers, and coordinate with national and international stakeholders. In addition, this body will issue ethical AI guidelines and monitor their application across sectors. India should also consider and implement the proposed reforms to intermediary liability laws, establish AI-specific due diligence requirements, strengthen enforcement mechanisms, and cooperate legally internationally to combat AI-enabled child cyberbullying.

Currently, the Indian Information Technology Act 2000 does not require the reporting of AI-generated cyberbullying against children. To provide additional protection to Children, it is essential to put in place a mechanism requiring

Credit Card Companies and Internet Service Providers to report AI-generated indecent visual representations of a child to law enforcement agencies. In such a case, the POCSO Act, 2012, and the IT Act, 2000, would be the subsequent steps that are properly coordinated. In contrast, Section 19 of the POCSO imposes an obligation on the general public, including children, to report offences to a law enforcement agency. However, it is limited to offences enumerated in the Act and does not cover the entire range of AI-generated cyberbullying against children. There is a recommendation that both the IT Act and the POCSO Act be amended to enhance the current law regarding AI-generated cyberbullying against children.

Due diligence obligations should also be framed more precisely. Under the 2021 Rules, intermediaries must make reasonable efforts to prevent users from hosting or sharing unlawful content. They must acknowledge complaints within 24 hours and ordinarily resolve them within 15 days, while specified takedown complaints must be acted upon expeditiously and resolved within 72 hours. Where unlawful information is identified through a court order or government notification, removal or disabling of access must occur as early as possible and no later than 36 hours. However, due diligence is a difficult task and is usually subject to a court order or the Secretary of Information Technology's consent; thus, there is a time lag between performance and actual enforcement. In order to enhance the enforcement laws that are in place at present, the following proposals have been put forward: changes in the IT Act, 2000, and the Intermediaries Guidelines, 2021, that would specify detailed due diligence requirements and also give proactive measures by providing methods for the pre-filtering of illegal content on digital platforms. These measures make safety and accountability accessible to users while allowing ISPs to comply with their primary legal obligations.

Children should be treated as rights-holders, not merely as victims. In line with the Convention on the Rights of the Child, 1989, and General Comment No. 12 on the right of the child to be heard, India should ensure that children can express views on digital governance and online safety measures that affect them. General Comment No. 25 on children's rights in relation to the digital environment likewise stresses that States should adopt legislative and policy measures for children's rights online, involve civil society and child-led groups, and ensure effective remedies and support mechanisms. India should incorporate these principles more directly into its AI, intermediary, and child-protection frameworks (UN General Assembly, 1989; Committee on the Rights of the Child, 2021).

5.2 Practice-Focused Recommendations

From a practical perspective, tech businesses should incorporate Safety by Design principles throughout the entire AI development process. Automated moderation, therefore, aims to stop and eliminate dangerous material, including deepfakes, impersonation profiles, and CSEAM content. Gen-AI tools should, by definition, have abuse filters, parental restrictions, and emergency flagging mechanisms. Platforms supporting or ignoring AI child abuse should face sanctions, suspensions, penalties, or even criminal prosecution.

To strengthen its collaboration with international law enforcement agencies, India should accede to the Budapest Convention on Cybercrime. Currently, as an unsigned country, India poses obstacles to international law enforcement agencies due to its inability to share information, trace IP addresses, and arrest criminals operating in other countries. In addition, the United Nations cybercrime treaty, adopted in December 2024, includes provisions to address the exploitation of children online. Therefore, as international law evolves to address future challenges in digital offences, India should carefully examine how to leverage international treaties to strengthen partnerships and combat the growing problem.

Efforts to establish preventive measures, such as age verification, responsible platform design, and real-time monitoring of children's digital activities, have not been properly implemented in India. The French government has set strict age verification, i.e., restricting social media access for children under 15, requiring parental authorisation below that age and other regulatory standards for digital platforms to restrict children's access to harmful content; therefore, it focuses on preventive measures. Australia also sets the age limit for social media access, such as gaming, search engines, and app stores, at under 16 years, but this does not necessarily extend to all gaming platforms, search engines, and app stores. Digital ID verification, facial age estimation from selfies, and Third-Party verification are methods used globally to verify an individual's age. India can address these international best practices by developing a comprehensive strategy to implement age-verification measures across all digital platforms, increasing accountability on digital platforms, and using technology-based monitoring tools to pre-emptively protect children in digital environments.

According to the 2024 Organisation for Economic Co-operation and Development (OECD) Guidelines on Children in the Digital World, the Government of India must establish real-time reporting systems with national helplines to enable timely responses to threats and abuse experienced by children online (OECD, 2025). The government must implement an AI literacy programme at a national level for children, parents and teachers to understand how to identify misinformation, exploitative behaviour and AI-related risks. Schools should require the incorporation of AI safety and digital ethics education in the curriculum, as well as training for teachers on digital safety/AI,

and completion of peer-led monitoring systems. Together, these efforts will be a foundation for creating a secure, inclusive and human rights-centric digital environment that conforms to the trustworthy AI standard.

As a practice-based framework, the Technological, Organisational, Environmental (TOE) model helps organise recommendations for addressing Gen-AI-enabled cyberbullying of children in India (Tornatzky & Fleischer, 1990). At the technological level, it points to the need for detection systems, rapid takedown tools, metadata preservation, and safer platform design. At the organisational level, it highlights the roles of schools, child-protection bodies, law-enforcement agencies, and digital intermediaries in creating trained, coordinated response systems. At the environment level, it highlights the need for more precise legal definitions; harmonised regulations; increased accountability among intermediaries; and the existence of child-centred policies to protect children from risks associated with new technologies. Essentially, the TOE model provides an effective multi-level strategy to address the growing threats posed by Gen- AI to the environment.

6. Conclusion

The rise of new Gen- AI methods has created children's digital spaces that may provide benefits, but also expose them to more sophisticated forms of harm. While digital technologies provide several significant benefits, they also create a highly vulnerable environment for children where offenders can take advantage of weaknesses in the platform to perpetrate abuses against children, including through the use of AI-generated indecent visual representations of children as well as other synthetic forms of abuse. The seriousness of the risk is compounded by the fact that children already face structural disadvantages in accessing, understanding, and safely navigating digital technology. Reliance on piecemeal changes to current laws and policies in this area may be neither timely nor efficient in addressing these risks. Instead of anticipating that general statutes will encompass all areas of abuse towards children that has been perpetrated through AI, India should look to develop specific child-centred statutes or legal structures to specifically address AI-driven sexualised images, deep fakes, impersonation, and other forms of AI-based cyberbullying as a way of developing clearer definitions, more targeted remedies, and more clearly defined responses to technology-based harms which are evolving rapidly.

However, legal remedies alone are not sufficient to establish effective protection for children on the Internet; effective child protection will require an integrative strategy that combines legal accountability, platform governance, reporting mechanisms, and public awareness, and is achieved through collaboration among multiple stakeholders. One example of such collaboration within the United Kingdom is the IWF, an independent reporting centre in the UK for reporting and helping to combat child sexual abuse images online, as

well as working together with both the public and industry. In the United States, the Centre for Missing & Exploited Children's CyberTipline provides one centralised source for the reporting of child sexual exploitation via the Internet. It provides support to law enforcement in responding to incidents of child sexual exploitation or missing children through the reporting of child sexual exploitation through the CyberTipline. In India, taking a similar approach would require stronger collaboration among parents, children, teachers, legal representatives, non-governmental organisations, Legislators, technology businesses, and law enforcement. By creating Public-Private Partnerships, stakeholders can work together to strengthen the monitoring and removal of illegal content from the Internet, improve digital citizenship and platform accountability, and increase the awareness of emerging harms caused by AI in the community.

References

Agnihotri, A. (2024). *Parenting guide: 10 effective strategies for parents to foster digital literacy in adolescents. Hindustan Times.* Retrieved May 7, 2026, from https://www.hindustantimes.com/lifestyle/relationships/parenting-guide-10-effective-strategies-for-parents-to-foster-digital-literacy-in-adolescents-101716814723512.html

Badawy, W. (2025). The ethical implications of using children's photographs in artificial intelligence: Challenges and recommendations. *AI and Ethics*, 1–12.

Bastian, M. (2024, November 13). *Generative AI reduces demand for some freelance jobs in writing, coding and design, study says. The Decoder.* Retrieved May 7, 2026, from https://the-decoder.com/generative-ai-reduces-demand-for-some-freelance-jobs-in-writing-coding-and-design-study-says/

Beaumont, R. (2022, March 31). *LAION-5B: A new era of open large-scale multimodal datasets. LAION*. Retrieved May 7, 2026, from https://laion.ai/blog/laion-5b/

Bharatiya Nyaya Sanhita, (2023). No. 45 of 2023 (India). Retrieved May 7, 2026, from https://www.indiacode.nic.in/handle/123456789/20062?view_type=browseB

Bickert, M. (2020, February 17). *Charting a way forward on online content regulation*. *Facebook.* Retrieved May 7, 2026, from https://about.fb.com/news/2020/02/online-content-regulation/

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). *On the opportunities and risks of foundation models*. arXiv. https://doi.org/10.48550/arXiv.2108.07258

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901. https://doi.org/10.48550/arXiv.2005.14165

Chakrabarty, S. P., & Chattopadhyay, S. (2024). *Criminal liability of artificial intelligence-generated deepfakes in India.* In *Proceedings of the 5th International Ethical Hacking Conference* (pp. 233–250). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-1234-5_20

Chowdhury, R., & Lakshmi, D. (2023). *Your opinion doesn't matter, anyway: Exposing technology-facilitated gender-based violence in an era of generative AI*. United Nations Educational, Scientific and Cultural Organization.

Committee on the Rights of the Child. (2021). *General comment No. 25 (2021) on children's rights in relation to the digital environment* (CRC/C/GC/25). United Nations.

DePaolis, K. J., & Williford, A. (2019). Pathways from cyberbullying victimization to adverse health outcomes among elementary school students: A longitudinal investigation. *Journal of Child and Family Studies, 28*(9), 2390–2403. https://doi.org/10.1007/s10826-018-1104-6

European Commission. (2025, July 14). *Commission publishes guidelines on the protection of minors*. Retrieved May 7, 2026, from https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-protection-minors

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press

Goel, H., & Chaudhary, G. (2024). Securing the digital footprints of minors: Privacy implications of AI. *Balkan Social Science Review, 23*(23), 235–?. https://doi.org/10.46763/BSSR242323235G

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative adversarial networks*. arXiv. https://doi.org/10.48550/arXiv.1406.2661

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society, 7*(1), 2053951719897945. https://doi.org/10.1177/2053951719897945

Goyat, A. (2025, August 26). *Panjab University students' elections: Fake Instagram page circulates private photos of candidate, probe on. The Indian Express. Retrieved* May 8, 2026, from https://indianexpress.com/article/cities/chandigarh/panjab-university-students-elections-fake-instagram-page-circulates-private-photos-of-candidate-probe-on-10211305/

Grassini, S. (2024). Computational power and subjective quality of AI-generated outputs: The case of aesthetic judgement and positive

emotions in AI-generated art. *International Journal of Human–Computer Interaction*. https://doi.org/10.1080/10447318.2024.2422755

Gupta, R., Tiwari, S., and Chaudhary, P. (2025). Applications of generative AI models. In *Generative AI: Techniques, models and applications* (Lecture Notes on Data Engineering and Communications Technologies, Vol. 241). https://doi.org/10.1007/978-3-031-82062-5_8

He, S., & Lu, Y. (2024). Effectiveness of Gen AI in assisting students' knowledge construction in humanities and social sciences courses: Learning behaviour analysis. *Interactive Learning Environments, 32*(10), 7041–7062. https://doi.org/10.1080/10494820.2024.2415444

Hinduja, S. & Patchin, J. W. (2024). *Bullying Beyond the Schoolyard: Preventing and Responding to Cyberbullying* (3rd Ed.). Thousand Oaks, CA: Sage Publications

Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (India). Retrieved May 7, 2026, from https://www.meity.gov.in/static/uploads/2024/02/Information-Technology-Intermediary-Guidelines-and-Digital-Media-Ethics-Code-Rules-2021-updated-06.04.2023-1-2.pdf

Information Technology Act, 2000, No. 21 of 2000 (India). Retrieved May 7, 2026, from https://www.indiacode.nic.in/bitstream/123456789/1999/1/A2000-21%20(1).pdf

INHOPE. (2022, November 2). *What is CSAM?* INHOPE. Retrieved May 8, 2026, from https://www.inhope.org/EN/articles/what-is-csam

Internet Watch Foundation. (2023). *How AI is being abused to create child sexual abuse imagery*. *IWF*. Retrieved May 8, 2026, from https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/

Just Rights for Children Alliance v. S. Harish, 2024 SCC OnLine SC 2611 : 2024 INSC 716, Criminal Appeal Nos. 2161-2162 of 2024, decided on 23 September 2024. Retrieved May 8, 2026, from linkinglaws.com

JustAI. (2024, September 29). *School students in UP booked for posting AI-generated obscene image of teacher, stressing on urgent need for cyber awareness.* Retrieved May 8, 2026, from https://justai.in/school-students-in-up-booked-for-posting-ai-generated-obscene-image-of-teacher-stressing-on-urgent-need-for-cyber-awareness-29-09-24/

Kanders, K., Stupple-Harris, L., Smith, L., & Gibson, J. L. (2024). *Perspectives on the impact of generative AI on early-childhood development and education. Infant and Child Development, 33*(4), e2514. https://doi.org/10.1002/icd.2514

Kowalski, R. M., & McCord, A. (2020). Perspectives on cyberbullying and traditional bullying: Are they the same or different? In *The Routledge companion to digital media and children* (pp. 460–468). London: Routledge.

Lakatos, S. (2023, December 8). *A revealing picture: AI-generated 'undressing' images move from niche pornography discussion forums to*

a scaled and monetised online business. *Graphika*. Retrieved May 8, 2026, from https://graphika.com/reports/a-revealing-picture

Levine, A. S. (2024, May 20). *'I want that sweet baby': AI-generated kids draw predators on TikTok and Instagram. Forbes.* Retrieved May 8, 2026, from https://www.forbes.com/sites/alexandralevine/2024/05/20/ai-generated-kids-tiktok-instagram-social-media-child-safety-predators/

Malsia, E., & Loku, A. (2024). Generative artificial intelligence in health system management: Transformative insights. *Journal of Service Science and Management, 17*(2), 107–117. https://doi.org/10.4236/jssm.2024.172005

McStay, A., & Rosner, G. (2021). Emotional artificial intelligence in children's toys and devices: Ethics, governance, and practical remedies. *Big Data & Society, 8*(1), 2053951721994877. https://doi.org/10.1177/2053951721994877

Milosevic, T., Van Royen, K., & Davis, B. (2022). Artificial intelligence to address cyberbullying, harassment, and abuse: New directions in the midst of complexity. *International Journal of Bullying Prevention, 4*, 1–5. https://doi.org/10.1007/s42380-022-00117-x

Milosevic, T., Verma, K., Carter, M., Vigil, S., Laffan, D., Davis, B., & O'Higgins Norman, J. (2023). Effectiveness of artificial intelligence–based cyberbullying interventions from youth perspective. *Social Media + Society, 9*(1). https://doi.org/10.1177/20563051221147325

Mishna, F., Birze, A., Greenblatt, A., & Khoury-Kassabri, M. (2021). Benchmarks and bellwethers in cyberbullying: The relational process of telling. *International Journal of Bullying Prevention, 3*(4), 241–252. https://doi.org/10.1007/s42380-020-00082-3

Mishna, F., McInroy, L. B., Lacombe-Duncan, A., Bhole, P., Van Wert, M., Schwan, K., & Johnston, D. (2016). Prevalence, motivations, and social, mental health, and health consequences of cyberbullying among school-aged children and youth: Protocol of a longitudinal and multi-perspective mixed method study. *JMIR Research Protocols, 5*(2), e83. https://doi.org/10.2196/resprot.5292

NDTV. (2024, September 28). *Case against UP school students for posting AI-generated obscene image of teacher*. Retrieved May 8, 2026, from https://www.ndtv.com/india-news/case-against-up-school-students-for-posting-ai-generated-obscene-image-of-teacher-6671522

O'Higgins Norman, J. (2020). Tackling bullying from the inside out: Shifting paradigms in bullying research and interventions. *International Journal of Bullying Prevention, 2*(3), 161–169. https://doi.org/10.1007/s42380-020-00076-1

OECD. (2025). *How's life for children in the digital age?* OECD Publishing. https://doi.org/10.1787/0854b900-en

Ofcom. (2026, March). *Online safety industry bulletin - March 2026*. Retrieved May 8, 2026, from https://www.ofcom.org.uk/cy/online-safety/illegal-and-harmful-content/online-safety-industry-bulletins/online-safety-industry-bulletin-march-2026

Online Safety Act 2021 (Cth). (2021). Federal Register of Legislation. Retrieved May 8, 2026, from https://www.legislation.gov.au/Series/C2021A00076

Online Safety Act 2023 (UK). (2023). UK Legislation. Retrieved May 8, 2026, from https://www.legislation.gov.uk/ukpga/2023/50/contents

Online Safety Amendment (Social Media Minimum Age) Act 2024 (Cth). (2024). Federal Register of Legislation. Retrieved May 8, 2026, from https://www.legislation.gov.au/C2024A00127

Otis, N., Clarke, R., Delecourt, S., Holtz, D., & Koning, R. (2024). The uneven impact of generative AI on entrepreneurial performance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4671369

Prabhu, S., Prabhu, S., & Noronha, F. (2023). Knowledge of child sexual abuse and attitudes towards reporting it among teachers and parents of children studying in selected primary schools of Udupi Taluk, India. Egyptian Journal of Food Science, 13 (1). https://doi.org/10.1186/s41935-023-00365-y

Protection of Children from Sexual Offences Act, 2012 (Act No. 32 of 2012). Retrieved May 8, 2026, from https://www.indiacode.nic.in/bitstream/123456789/2079/1/AA2012-32.pdf

Ranjith, P. J., Vranda, M. N., & Kishore, M. T. (2023). Predictors, prevalence, and patterns of cyberbullying among school-going children and adolescents. *Indian Journal of Psychiatry, 65*(7), 720–728.

Reeder, B., & Lee, K. (2024). Evaluating and incorporating generative AI in nursing informatics and data science graduate courses. *Studies in Health Technology and Informatics.* 315, 205–209. https://doi.org/10.3233/SHTI240135

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a single market for digital services and amending Directive 2000/31/EC (Digital Services Act). (2022). EUR-Lex. Retrieved May 8, 2026, from https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng

Reuters. (2025, August 29). *Meta created flirty chatbots of Taylor Swift, other celebrities without permission.* Retrieved May 8, 2026, from https://www.reuters.com/business/meta-created-flirty-chatbots-taylor-swift-other-celebrities-without-permission-2025-08-29/

Reznikov, R. (2024). Leveraging generative AI: Strategic adoption patterns for enterprises. *Modelling the Development of the Economic Systems, 1*, 201–207. https://doi.org/10.31891/mdes/2024-11-29

Singh, S., Saini, R., & Sagar, R. (2022). Quality of online news media reports of child sexual abuse in India. Industrial Psychiatry Journal, 31(2), 336. https://doi.org/ 10.4103/ipj.ipj_238_21

Smith, P. K. (2016). Bullying: Definition, types, causes, consequences, and intervention. *Social and Personality Psychology Compass, 10*(9), 519–532. https://doi.org/10.1111/spc3.12266

Sood, Y. (2023). *Addressing algorithmic bias in India: Ethical implications and pitfalls. SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.4466681

TAKE IT DOWN Act, Pub. L. No. 119-12, 139 Stat. 55 (2025). Retrieved May 8, 2026, from https://www.congress.gov/119/plaws/publ12/PLAW-119publ12.pdf

Tamboer, S. L., Kleemans, M., & Daalmans, S. (2022). *'We are a neeeew generation': Early adolescents' views on news and news literacy. Journalism, 23*(4), 806–822. https://doi.org/10.1177/1464884920924527

Times of India. (2025, August 28).*Classmates circulate AI pics of Hwh school girls on social media. Times of India*. (2025, August 28). Retrieved May 8, 2026, from https://timesofindia.indiatimes.com/city/kolkata/classmates-circulate-ai-pics-of-hwh-school-girls-on-social-media/articleshow/123569994.cms

Thiel, D. (2023). *Identifying and eliminating CSAM in generative ML training data and models.* Stanford Internet Observatory. Retrieved May 8, 2026, from https://cyber.fsi.stanford.edu/io/news/csam-generative-ai

TIME. (2024, January 26). *Taylor Swift deepfakes highlight the need for new legal protections.* Retrieved May 8, 2026, from https://time.com/6589263/taylor-swift-deepfakes-legal-protections/

Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation.* Lexington, Mass.: Lexington Books.

Tyagi, R. (2025, May 29). *Virtual scars, real harm: India's legal shift on child abuse in the digital space* [Opinion]. NDTV. Retrieved May 8, 2026, from https://www.ndtv.com/opinion/virtual-scars-real-harm-indias-legal-shift-on-child-abuse-in-the-digital-space-8538786

UNICEF Innocenti. (2025). *Childhood in a digital world: Screen time, digital skills and mental health.* Retrieved May 8, 2026, from https://www.unicef.org/innocenti/reports/childhood-digital-world

United Nations. (1989). *Convention on the Rights of the Child*. Office of the United Nations High Commissioner for Human Rights. Retrieved May 8, 2026, from https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child

Yang, K.-C., Singh, D., & Menczer, F. (2024). *Characteristics and prevalence of fake social media profiles with AI-generated faces. Journal of Online Trust and Safety, 2*(4). https://doi.org/10.54501/jots.v2i4.197

Yu, Y., Sharma, T., Hu, M., Wang, J., & Wang, Y. (2025). Exploring parent-child perceptions on safety in generative AI: Concerns, mitigation strategies, and design implications. In *Proceedings of the 46th IEEE Symposium on Security and Privacy (SP 2025)* (pp. 2735–2752). https://doi.org/10.1109/SP61157.2025.00090

Zhou, C., & Hou, F. (2024). Can AI empower L2 education? Exploring its influence on EFL teachers' and language learners' behavioural, cognitive, and emotional engagement. *European Journal of Education,*

59(4).
https://doi.org/10.1111/ejed.12750https://doi.org/10.1111/ejed.12750

Zsila, Á., & Reyes, M. E. S. (2023). *Pros & cons: Impacts of social media on mental health. BMC Psychology, 11*, Article 201. https://doi.org/10.1186/s40359-023-01243-x