IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 September 2022, accepted 19 October 2022, date of publication 31 October 2022, date of current version 8 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3218415

SURVEY

Child Safety and Protection in the Online Gaming Ecosystem

ANUM FARAZ[1], (Member, IEEE), JINANE MOUNSEF[1], (Member, IEEE), ALI RAZA[2], (Member, IEEE), AND SANDRA WILLIS[3]

[1]Electrical Engineering Department, Rochester Institute of Technology, Dubai, UAE
[2]Computing Sciences Department, Rochester Institute of Technology, Dubai, UAE
[3]Global Mental Health Laboratory, Columbia University, New York, NY 10027, USA

Corresponding author: Jinane Mounsef (jmbcad@rit.edu)

This work was supported in part by the Grant from the Rochester Institute of Technology's Academic Research Committee (ARC).

ABSTRACT Online gaming no longer has limited access, as it has become available to a high percentage of children in recent years. Consequently, children are exposed to multifaceted threats, such as cyberbullying, grooming, and sexting. Although the online gaming industry is taking concerted measures to create a safe environment for children to play and interact with, such efforts remain inadequate and fragmented. There is a vital need to develop laws and policies to regulate and build minimum standards for the industry to safeguard and protect children online on the one hand, while promoting innovations in the gaming industry to preempt such threats. Many tools have been adapted to control threats against children in the form of content filtering and parental controls, thereby restricting contact with children to protect them from child predators. Different approaches utilizing machine learning (ML) techniques to detect child predatory behavior have been designed to provide potential detection and protection in this context. In this paper, we survey online threats to children in the gaming environment and present the limitations of existing solutions that address these threats. We also aimed to present the challenges that ML techniques face in protecting children against predatory behavior by presenting a systematic review of the available techniques in the literature. Therefore, this analysis provides not only recommendations to stakeholders to develop policies and practices that safeguard children when gaming, but also to the gaming industry to continue providing appropriate measures for a safe and entertaining gaming environment.

INDEX TERMS Artificial intelligence, chat logs, child protection, cyberbullying, machine learning, online gaming, predatory threats.

I. INTRODUCTION

The rapid growth of technology has remarkably transformed the way people connect with each other. The Internet is becoming a crucial source of information and entertainment. Social media, instant messaging, and audio/video calling platforms have become major sources of communication. Among the 4.95 billion Internet users, 1 in 3 are under 18 years of age and often use the Internet without the supervision of an adult [1], [2]. With the rapidly increasing use of the Internet among children and adolescents, it has become more important to provide them with a safe and secure environment. Children can face various threats while being involved in different online social activities that could involve exposure to violent content. Child harassment and pornographic content, cyberbullying, child victimization, abuse, grooming, sexting, and pedophilia are also among the common and serious threats children can face while socializing online with strangers or even with peers [3].

According to [4], more than 90% of children in the US play online games. This figure increases to 97% among children aged 12-17. Online gaming is considered a source of learning that aims to enhance children's cognitive abilities [5]. These games provide a useful means of building leadership qualities in children [5] and enhance teamwork skills [6]. Moreover, multiple games are developed to help children in educational

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Jiang.

fields, such as learning science and mathematics [7], [8], [9]. Online gaming is also found to be associated with positive outcomes, such as enhanced social relationships. However, problematic outcomes are also associated with excessive online gaming, such as negative emotions and attitudes, low self-esteem, loneliness, anxiety, poor academic performance, and maladaptive coping strategies [10]. Notably, mobile game addiction is associated with social anxiety, depression, and loneliness, with male adolescents reporting the highest social anxiety when gaming excessively [11]. Few studies have examined the relationship between game addiction and mental health outcomes, due to a lack of standardized instruments required to measure this new type of behavioral addiction. Gaming platforms provide public chat rooms, private chat rooms, group chat rooms and in-game chatting for online gamers to interact with each other. The same chat rooms on these platforms can also pose a considerable risk to children.

A survey of 10-17 years old children in the US showed that 56% of child abuse incidents occurred on social networking sites, while 11% occurred in online video chat rooms and 6% in game chat rooms or gaming sites [12]. Age-inappropriate games that include sexual, abusive and self-harm content are also a source of threat to children, knowing that the content could be a shared media over the chat. Gaming platforms provide parental control features to monitor the online content presented to children along with a variety of options to limit the contact with gamers that use the chatting features. Parental controls do not only provide an option to sway the children from being exposed to inappropriate content, but they also provide an opportunity for parents to restrict the socialization with strangers. However, the effectiveness of the parental controls remains debatable due to a deficiency of detailed knowledge and understanding of the control features. The lack of parental awareness about the control tools is a serious factor that hinders children's online safety [13], [14].

Policymakers and stakeholders are working along with governments to ensure effective vulnerable child protection online by placing protective measures and developing interactive user-friendly tools to enhance the knowledge of children and parents about online threats and the ways to combat them [15], [16], [17], [18]. Nevertheless, several challenges must be overcome to keep children safe online whilst enabling them to benefit from digital engagement opportunities.

Artificial Intelligence (AI) is widely used in gaming platforms through different applications. Most research in the literature is related to the development of AI agents to play games and compete with humans as opponents [19], [20], [21], [22], [23], [24]. The detection of predatory behavior on gaming chat platforms using AI tools is a growing field. Renowned AI techniques such as supervised learning require labelled datasets that are not widely available. However, raw chat logs are accessible and public. The authors in [25], [26], [27], and [28] picked chat logs from selected games, such as MovieStarPlanet [25], Online Battle Arena [26], World of Tanks [27], and Dota Ragnarok [28], to detect predatory behavior in online gaming. Nonetheless, most research [29], [52] has not been conducted in the context of online gaming but rather on chat logs related to social media websites, instant messaging applications, and public chat rooms. The distributed platform chatbots in [29], [30], [36], and [42] utilize natural language processing (NLP) and artificial intelligence markup language (AIML) to process the conversation for text content and emotion classification in addition to opinion classifiers to detect predatory behavior on chatting platforms. The most widely used technique for detecting child predators on chatting platforms is text classification, which utilizes the two available data sources reported in [53] and [54]. ML methods used to detect cyberbullying [27], [28], [31], [48] and sexual predatory behavior [25], [29], [30], [32], [40], [42], [47], [49], [52] remain the main focus of research over the last 10 years. In [55], a brief survey of ML algorithms was used to detect child grooming behavior on social media between 2007 and 2016. In [56], the automatic detection of cyberbullying published between 2008 and 2016 was surveyed. In [57], a human-centered review of computational approaches including ML to detect online sexual risk in children was presented. In [58], a survey of automated methods for detecting cyberbullying between 2008 and 2020 was provided. In [59], a survey of computational methods used to identify cyberpredators on different social media platforms was proposed. Despite the increasing number of cases of cyberbullying and child sexual grooming in the gaming environment, solutions to children's safety on gaming platforms using AI tools have not been adequately addressed in the literature.

The main contributions of our work are multifold and include:

- Examining the different aspects of child safety by highlighting the existing threats to children in gaming environments on the one hand, and the existing protection mechanisms provided by the industry, researchers, international and national law makers, and regulators.
- A substantial survey of the different AI tools applied in the gaming environment including the detection of child predatory behavior.
- Highlighting the need to leverage AI technology to identify the pervasiveness, type, and risks associated with predatory behavior in online gaming and its potential effects on children's and adolescents' mental health and overall protection.

Table 1 provides a comparison of the contributions of this work to existing surveys in the literature.

The rest of the paper is organized as follows. Section II describes the research methodology of the presented survey. Section III provides an overview of the threats to children, along with current international laws and policies to protect children on gaming platforms. Section IV summarizes the available solutions and their shortcomings, along with recommendations to stakeholders provided by UNICEF and the International Telecommunication Union (ITU). Section V presents the available AI applications on gaming platforms along with the available AI techniques used in the literature

TABLE 1. Comparison of our work with existing surveys.

Ref.	Threats		Summary of threats	Protection tools	ML solutions	Laws and policies
	Cyber-bullying	Sexual risks				
[55]	×	✓	×	×	✓	×
[56]	✓	×	×	×	✓	×
[57]	×	✓	×	×	✓	×
[58]	✓	×	×	×	✓	×
[59]	✓	✓	✓	×	✓	×
Our work	✓	✓	✓	✓	✓	✓

to protect children from predatory behavior. Section VI discusses the results of the survey and provides recommendations. Finally, Section VII concludes the paper.

II. RESEARCH METHODOLOGY

In this study, a survey was conducted to identify relevant research that addresses child safety in gaming environments or chatting platforms using ML algorithms and AI tools. The survey considers only the work published between 2017 and 2021 using academic portals that can access research databases, including IEEE Explorer, Elsevier, Springer, ACM, Cambridge, Wiley, ProQuest, and Sage. We use the following keywords to search the portals without filters: ''cyberbullying and AI'', ''cyberbullying and ML'', ''child predators, AI and games'', ''cyber threats'', ''cyber predators'', ''child paedophiles'', ''child pedophiles'', ''detect child predators'', ''predatory behavior and games'', ''child pedophilia and online games'', and ''child safety online games''.

The research focuses on the literature that describes the threats to online gaming children and presents a solution to detect and combat child predators on gaming platforms and chatting platforms available as part of social media.

Initially, 1073 references were found based on a keyword search. After filtering out newspaper, magazine articles and book chapters, 550 papers were selected, which were then filtered based on their titles and keywords, keeping 200 relevant papers. The abstracts of these papers were analyzed to filter out 100 papers; the remaining 100 papers were reviewed in detail, leaving 29 papers that fulfilled the required criteria. The second stage of the search included a comprehensive review of relevant cross-references that added five more papers to the previous ones, resulting in a total of 34 papers reviewed, as shown in Table 1. Table 2 provides a summary of the datasets and threats covered in the selected papers.

In addition to the review of ML and AI techniques for enhancing children's online safety, this work presents an overview of the existing laws, policies and regulations to protect children and recommendations provided by ITU and UNICEF to stakeholders, including children, parents, caretakers, educators, industry, and policymakers. This work also presents a review of tools and solutions proposed by the industry and organizations to enhance child online protection (COP).

III. ONLINE THREATS TO CHILDREN

A. THREAT TYPES

Online gaming is a popular leisure activity for children, but it also poses many threats. The digital environment provides great opportunities for children to learn in all disciplines, but at the same time, it poses a multitude of threats from organizations, adults, and peers. Online threats to children are broadly classified into three categories: content, contact, and conduct risks [60].

1) CONTENT RISK

Content risk includes exposure to inappropriate content such as adult, violent, extremist, and gory content. The assurance of content related to self-harm, self-abuse, destructive, and racist ideas is also considered a content risk. Exposure to incomplete and inaccurate information is another way to affect children's understanding of the world around them.

A research by UNICEF explored the consequences of exposure to game content and its potential effects on children's social relationships, education, physical activity, mental well-being, and psychological or developmental challenges such as depression, social anxiety, stress and excessive play [61]. Despite the volume of research, to date, the results from studies on the effects of online gaming on children's well-being, whether positive or negative, have been mixed [62], [63], [64], [65]. Evidence of the impact of online gaming on children has been oriented toward exploring issues or building initial theories, but it is not robust or reliable enough to inform policy decisions or best practice recommendations. This is the case not only for studies examining the influence of online gaming on children but also for research exploring the influence of digital technologies more broadly [66], [67], [68], [69].

2) CONTACT RISK

Children can face a broad range of contact threats from their adults and peers. Contact risk includes harassment, exclusion, defamation, victimization, child pedophilia, and grooming. A summary of the contact risks is presented below.

- Bullying is a common threat that children face online or in real life. The effects of bullying have been widely studied. Depression, anxiety, panic disorders, distributed personality, suicidality, criminality, and illicit drug misuse are common effects reported by people who face online or physical bullying during childhood [70]. In [71], the study showed a positive relationship between peer cyberbullying and suicidal ideation among young children and adolescents.
- Children with disabilities are more prone to experiencing victimization, including bullying, harassment exclusion, and discrimination.
- Defamation of a child, such as sharing images and/or videos or sharing altered images and/or videos of a child, can put him or her in complete devastation.
- Children can also be targeted, groomed and sexually abused by adults pretending to be someone they are not.

TABLE 2. Studies with datasets and identified threats.

Ref.	Year	Datasets					Covered Threats			
		PJ	PAN12	PAN13	Collected from gaming sources	Combination of datasets/other sources	Cyberbullying	Sexual threats	Cyberpredators	Other
[54]	2012		✓					✓		
[29]	2013	✓						✓		
[25]	2015				✓				✓	
[26]	2015				✓					✓
[30]	2017							✓		
[31]	2017						✓			
[32]	2017			✓				✓		
[33]	2017			✓				✓		
[27]	2018				✓		✓			
[34]	2019					✓		✓		
[35]	2019			✓				✓		
[36]	2019					✓			✓	
[28]	2019				✓		✓			
[37]	2019					✓		✓		
[38]	2019					✓		✓		
[39]	2019	✓						✓		
[40]	2019		✓						✓	
[41]	2019					✓	✓			
[42]	2020					✓		✓		
[43]	2020					✓		✓		
[44]	2020		✓					✓		
[45]	2021		✓				✓			
[46]	2021		✓					✓		
[47]	2021		✓					✓		
[48]	2021					✓				✓
[49]	2021		✓					✓		
[50]	2021		✓					✓		
[51]	2021		✓					✓		
[52]	2022		✓						✓	

In May 2019, in California, a man was sentenced to 14 years in prison to force an 11-year-old girl to produce child pornography [72]. He approached the girl through the Clash of Clans game. A similar case was reported in suburban Seattle, where a man was caught by the Law Enforcement Agencies (LEA) for blackmailing three boys and forcing them to share inappropriate photos. He was posing as a teenager and approached the victims through Minecraft and League of Legends [72]. In a recent event in May 2022, a 17-year-old boy committed suicide after being scammed by a man posing as a girl [73]. According to a CNN report, the scammer shared a nude photo and asked the victim to share his photo. After receiving the photo from the victim, the scammer started to extort him to send him money, but the victim was unable to arrange the money. He then committed suicide and his family came to know about the whole situation through the suicidal note.

3) CONDUCT RISK

Conduct risk includes the children behaving as perpetrators. Children can play a role in victimizing their peers. This includes harassment, bullying, sexting, exclusion, shaming, and the generation of inappropriate content by the child [74]. The different scenarios of conduct risk can be summarized as follows.

- Online bullying is more damaging than real-life bullying because it can spread in less time and the shared content or images are available for a longer period; hence, it is harder for the victim to overcome the embarrassing situation.
- Children are responsible for plagiarism, such as uploading pictures of others, without their consent.
- Children can use disrespectful names to harass or bully their peers.
- A very common behavior observed in adolescents is sexting, which involves sharing sexualized images and/or text via messages. The outcome of sexting is wide, ranging from positive and accepted to negative and unwanted [75]. The photos produced for sexting can be distributed to a wider audience, often leading to embarrassment, harassment, and placing adolescents in vulnerable positions [76].

B. CURRENT LAWS TO PROTECT CHILDREN

The electronic code of federal regulation (eCFR) is a compilation of the material published by the Code of Federal Regulations (CFR) and the Federal Register amendments produced by the National Archives and Records Administration's Office of the Federal Register (OFR) and the Government Publishing Office [77]. The regulations of eCFR related to children's usage of online platforms can be summarized as follows.

- It is unlawful for the operator of any platform to collect the personal information of a child without providing a written notice on the website.

- Parents should be informed of the collection of personal data and a verifiable parental consent is required prior to the collection, use, and disclosure of the data.
- The notice should clearly mention the information type that is collected and the way it is used by mentioning the disclosure practices.
- It is the responsibility of the operator to use the available technology to verify that consent has been provided by the parents.
- Once the verification process is complete, the information collected for parents identification should be deleted by the operator immediately from the company's record.
- Parental consent is mandatory for the approval of transactions made using the platform's online payment system.
- The operator is required to provide a reasonable platform for the parents to review the collected personal information of the child in order for them to allow or deny further use of the information.
- Parents should be given the opportunity to refuse or permit the operator to delete or use the provided information.
- The operator is not allowed to condition a child's participation in games by requiring the child to provide additional personal information necessary to participate in any activity.
- The operator is required to protect the confidentiality, security, and integrity of the children's personal information and to ensure that the information shared by a third party takes care of the integrity of the personal data.

IV. EXISTING PROTECTION MECHANISMS FOR CHILDREN's ONLINE GAMING

A. CHILDREN's RIGHTS AND GAMING ENVIRONMENT

The multitude of risks faced by children in the gaming environment also poses a threat to the infringement of their rights. UNICEF prioritizes engagement with the information and communication technology (ICT) industry and works in the following areas to increase children's safe usage of the Internet and the associated technologies by tackling different issues, such as the transmission of children's online sexual abuse images, exposure to inappropriate content or contact, and violation of the child's privacy [12]. UNICEF is also working with corporate partners that harness ICT to provide children with opportunities to become engaged digital citizens and use ICT platforms for learning, sharing, and communicating [15], [17].

UNICEF presented the children's rights related to the positive and negative impacts of online gaming in a discussion paper [12]. Children's rights related to online gaming should be taken care of by all stakeholders to protect the child from risk. Many of UNICEF children's rights can be associated with the gaming industry: acting in the best interest of the child (Article 3), a parental guidance consistent with the child's evolving capacities (Article 5), the right to leisure, play and culture (Article 31), the protection of a child from sexual abuse (Article 34), parents' primary responsibility for the upbringing and development of the child (Article 18), children's right to non-discrimination (Article 2) and freedom of association (Article 15), respect for the views of the child (Article 12), the children's right to freedom of expression (Article 13), the protection of privacy and personal information (Article 16), the protection of the child from all types of exploitations (Article 36), and the right to education (Articles 28 and 29). All the protection mechanisms developed for the children must uphold their rights while online.

B. STAKEHOLDERS AND THEIR ROLES

This section summarizes the guidelines and effective tools created by international and national organizations to enhance the child online protection (COP) for relevant stakeholders including children, parents, caretakers, educators, industry, and policymakers. In November 2008, the International Telecommunication Union (ITU) launched the COP initiative as a multi-stakeholder global initiative to create a safe and empowering online experience for children [15]. The COP guidelines have served national government entities, civil society organizations, industry, and many other stakeholders in their children's online protection efforts. Moreover, the COP's initiative attained further endorsement and validation during the 2018 Plenipotentiary Conference of the International Telecommunication Union held in Dubai. The multitude of stakeholders also framed the protection of children online within the framework of the United Nations Convention on the rights of children and other human rights treaties. To address the exploding and transforming threats to online children, the COP-updated guidelines were launched in June 2020.

1) CHILDREN

Children are key consumers in a gaming environment. Therefore, they need to be educated on their rights and the protection of their rights. The ITU launched an online safety course with Sangophone (Sango), where a child online protection mascot equips children with the knowledge they need to know about their rights and responsibilities when they are online [78]. The five episodes of this course have been launched to address the multiple issues a child can face online, such as inappropriate content, sharing of personal information, the vulnerability of different threats while using social media applications and gaming environment, downloading, and in-store purchase of games, etc. The ITU created three different resources to guide children in different age groups. A storybook with questions was developed for children under nine to provide them with an understanding of their rights and safety online. A workbook containing educational activities was designed for children aged 9-12 years. Through these activities, children can learn about their rights and online risks. A social media campaign was created for children aged 13-18 to help them learn how to manage risks online.

2) PARENTS/CARETAKERS & EDUCATORS

Parents, caretakers, and educators are responsible for the wellbeing of children. Therefore, they must play a positive role in protecting children online. The ITU provides recommendations to parents/caretakers and educators to understand children's vulnerabilities and the best protective measures to safeguard them [79]. These guidelines include all the major threats a child can face online and how parents/educators can help children by providing the right and complete information needed to protect them online.

The ITU recommends the parents to:

- Have a discussion with their children about the vulnerabilities and mechanisms available to protect against them by joining the children in their online activities.
- Monitor all devices used by their children including their mobile phones, laptops, tablets, gaming consoles, fitness trackers, smart televisions, and applications used on any of these devices.
- Install firewall and antivirus software on all the devices. Parental controls and filtering are useful tools; however, children's privacy should also be considered.
- Control their children's access to age-appropriate websites, set rules such as screen time, and teach their children about their privacy issues.
- Create a positive environment so that children can express their problems and opinions. Many websites may not ask parents' permission for their children to join a website or a platform.
- Be aware of the minimum age requirements for their children to use these platforms.
- Be aware of the unauthorized access to the debit or credit cards through their children's account by controlling the use of cards and other payment mechanisms.
- Be aware of reporting a person or inappropriate content on any platform their children use.
- Talk to their children about any advertisement that might be misleading and inappropriate.
- Educate their children about the threats related to their relationship with strangers.
- Be aware of the people their children are chatting with or meeting with online.
- Teach the children about the privacy issues and managing their personal information online.
- Inform their children that photos can reveal a lot of personal information and thus, explain the risks associated with uploading photos or any other confidential content.
- Tell the children about obtaining their parents' consent before sharing any information or photos of their family or friends.

Educators were also provided with guidelines to protect children from online threats. The ITU recommends that teachers and other relevant school staff:

- Ensure that all devices are password-protected and that the antivirus and firewalls are updated.
- Communicate a clear policy about how technology can be used to students and their parents.
- Acquire parents' consent when taking photos of children and sharing them on social media platforms.
- Ensure that inappropriate content is filtered and monitored via the Internet network provided by the school.
- Raise awareness of the importance of the digital footprint and online reputation.
- Understand the importance of professional online communication with students, parents, and other stakeholders.
- Have knowledge of risks and vulnerabilities students can be exposed to when they are online.

3) INDUSTRY

ITU guidelines are helpful in creating a connected framework for the COP to create harmony among all stakeholders. To accomplish this, the industry is also provided with guidelines to play a role in the COP. The industry includes Internet service providers, social networks, messaging and gaming platforms, hardware and software manufacturers, companies providing digital media and several services, such as streaming, digital file storage and cloud-based. The ITU recommendations to the industry require the following actions in collaboration with international and national governments and law enforcement organizations:

- Identify, prevent, and mitigate the adverse impacts of ICT on children's and adolescents' rights by developing child protection, safeguarding policies, and integrating risks and opportunities into company-wide policy commitments [80].
- Play a role in combatting Child Sexual Abuse Material (CSAM) and prohibiting the uploading or sharing of content that violates the rights of any party.
- Provide users with a comprehensive way to report any inappropriate content and take prompt actions against them in accordance with the international and national government and law enforcement organizations.
- Be responsible for actively monitoring any content hosted on the company's server on a regular basis using tools, such as hash scanning of known children's abuse images, image identification software and URL blocking to oversee CSAM.
- Make sure to provide a safe and enjoyable digital environment for children and adolescents by providing age-appropriate content to children of different age groups, enhancing the existing parental control features and developing new tools using technological advances to help the COP.
- Create an efficient framework for the awareness of customers related to spam, data theft and inappropriate contact, such as bullying and grooming, and educate them on the procedures to combat them.
- Invest in research and develop tools or educating material to enhance children's, parents', caretakers', and educators' knowledge about the children's rights and protection mechanisms provided by the industry itself,

international and national governments, policymakers, and other law enforcement agencies.

The European Commission suggested that the industry be self-regulatory, as self-regulations allow the latter to create their own system by which they can deal with the challenges a child can face online [81].

4) POLICYMAKERS

The ITU highlights the need for policy frameworks to address all harms against children in the digital environment but at the same time, this should not unduly restrict children's rights. The national-level recommendation provided by the ITU is summarized as follows [82].

- It is important to note that any illegal act against children in the real world is illegal online. Therefore, framing legal regulations for online data protection and privacy rules for children is necessary.
- Self-regulatory or co-regulatory policy development is required along with the full regulatory framework.
- A mechanism should be established and promoted to report any illegal content as well as reporting user issues or concerns.
- Research is required to engage all stakeholders to determine their opinions, ideas, experiences, difficulties, and opportunities for COP.
- Digital literacy features should be a part of the national school curriculum that is applicable to children of different age groups.
- Educational resources should be developed to reflect cultural norms and laws and to enhance the COP.
- National awareness campaigns are needed to highlight the COP's related issues.
- The understanding of tools, applications and settings that help COP should be evaluated and improved.

C. TOOLS FOR AWARENESS OF END CONSUMERS

The ITU in collaboration with the National Cybersecurity Authority (NCA) of the Kingdom of Saudi Arabia launched an ITU global program on children's online protection in December 2020 [83]. The work stream of this project was divided into two stages. In the first stage, cyber-skill development is provided to train children, adolescents, parents, and educators. It also intends to develop a game and an application for children of different age groups to understand the COP guidelines by the end of 2022. The translation of the COP guidelines into national languages and the use of the Sango tool are also included in the scope of this first stage [78]. The second phase of this project aims to provide national strategy development on COP and capacity building for ICT professionals and government international and national stakeholders. This phase will be completed by 2024.

UNICEF and the Global Partnership to End Violence Against Children launched an AI-based game, which is a social-emotional learning tool that teaches children the skills needed to stay safe and protected online [17]. This tool was designed for 5-10 years old children to learn the skills necessary to protect themselves. It also provides parents and teachers with guidelines on how to teach their children about online safety.

In the UAE, the Ministry of Community Development, in collaboration with the Telecommunication Regulatory Authority, launched a digital platform kidX that provides an interactive environment using games and virtual reality technologies to raise children's and adolescents' knowledge about online safety [16].

Finally, the European Union (EU) developed an interactive platform to enhance parents' and children's awareness by creating different scenarios related to online grooming, sexual exploitation, and domestic violence [18].

D. TOOLS FOR SAFETY OF END CONSUMERS

1) AGE-APPROPRIATE RATING

Age-appropriate ratings provide guidelines to consumers including children, parents/caretakers and the industry regarding the age group for which a game is appropriate. Established in 2013, the International Age Rating Coalition (IARC) provides a globally standardized age classification process for digital games and mobile applications [84]. IARC is administered by games rating authorities including the Australian Classification Board from Australia [85], the Classificacao Indicativa (Classind) from Brazil [86], the Game Rating and Administration Committee (GRAC) from the Republic of Korea [87], the Entertainment Software Rating Board (ESRB) from North America [88], the Pan European Game Information (PEGI) from Europe [89] and the Unterhaltungssoftware Selbstkontrolle (USK) from Germany [90]. The IARC system was established in close collaboration with rating authorities, game developers, and game retailers. The rating contains a three-part categorization that suggests age appropriateness, content descriptors that indicate the content type that may have triggered a particular rating, and interactive elements that advise about several risks, such as sharing the user's location with other users, or the fact that personal information may be shared with third parties. A free of cost application can be downloaded from Google Play or the Apple Store to check the rating of any video game and obtain insight into the game content.

A developer must submit the deployed game along with the questionnaire available on the IARC website. Based on the information provided, the IARC assigns age ratings and content descriptors in accordance with regional rating authorities and the interactive elements are assigned universally. The IARC rating authorities check the ratings assigned to the game to ensure the accuracy of the age rating if needed, and corrections are implemented by the storefronts. Developers need to submit an IARC report when the game is submitted to a publisher. The rating system is not only used by game publishers and retailers but also by parents who are the key consumers. The rating system provides guidelines to the parents by describing age ratings and content descriptors. The parental controls also used the age rating to filter games for the different age groups of children.

The Entertainment Software Association (ESA) conducted a survey on the video game industry in July 2021, where approximately 4000 participants from the US took part [91]. According to the survey, 86% of parents were aware of the ESRB rating and 76% were using it to protect their children online. The survey, which only involved the US population, should be extended to other countries to provide more comprehensive insight into the usage of age-appropriate ratings.

2) GAMING PLATFORMS AND PARENTAL CONTROLS

In this section, selected gaming platforms are presented along with the parental control features provided by these platforms. Table 3 summarizes the gaming platforms with their chat features and parental controls. Each of these platforms, along with the features provided, are described below.

Roblox can be accessed through its website on desktop computers [92]. For mobile users, the application is available in play stores. It provides a large variety of games from different genres. There is no restriction on age to make an account or play a game, but for children aged under 13 years, the game limits the user's account to a restricted view of the website/application. It also provides basic filtration of games, which is not suitable for players under 13. An age verification process is also introduced in which users need to upload any identity card along with their photo, but this is just an optional requirement. *Roblox* provides multiple chat features including chats with friends, group chat rooms, and chats with unknown people while playing a game. Chatting communication is written only and does not include voice messaging. As there is no age restriction for playing a game, children as young as five years can have a chat with an unknown person. Game developers claim to have a combination of chat filters that are both human-controlled and machine-moderated to filter inappropriate content, words associated with bullying and harassment, and personal information shared in the chat. *Roblox* provides a wide variety of parental control options. To use them, parents must generate a pin to secure the settings of the child's account. Parents/caregivers can link their account to the child's account and control or restrict the platform's available features. For example, a parent can place restrictions on persons who can message their children, chat with them in the application and chat with them in the games. Multiple options are provided to a parent including ''no one can chat'', ''only friends can chat'', and ''anyone can chat''. A purchase notification can also be enabled, and parents can receive a message whenever their children buy an item. A monthly spending limit can also restrict children from purchasing items from the game store. *Roblox* also provides an option to enable ''Account Restriction'' that blocks all the chats and prevent anyone from searching the account. Only a suitable pre-approved list of games and content is available for the restricted account.

Steam provides a wide variety of games that a user can play using the application on a desktop as well as on mobile devices [93]. An account is required to play a game and a minimum age of 13 years is required to create an account. However, parents can create a ''Family view'' account to enable any family member to access the game. The ''Family view'' option in the *Steam* platform provides options to restrict the accessible features that enable parents to choose the games that would be visible in the ''Family View'' for up to 10 different accounts. A parent can also disable access to the *Steam* store, chat rooms, friends list and the online profile of the primary user (parent). Content sharing is also available such as screenshots, video clips, game play broadcasts, and the user-generated data. *Steam* also provides users with the capability to chat with friends/group, voice message, and share content including photos and videos. All these features can be blocked in the ''Family view'' mode. In addition, *Steam* provides a wide variety of content filtering options for blocking mature content, frequent violence, nudity, sexual content, and adult content using the settings. Any related content game can be blocked or made inaccessible. *Steam* also blocks strong profanity and slurs in chatting platforms. Additional words or tags associated with games that need to be blocked in chats can be specified in the settings. These content and chat filtering options are applied to the primary and secondary accounts used in the ''Family View''.

Play Station 5 (PS5) is a popular console developed by Sony for playing games [94]. To use the console, the user must create an account with a minimum required age of 18. Parents can link children under 18 years old to their own account. *PS5* provides multiple chat features including chats with friends, group chat rooms, chats with unknown people while playing games, voice messaging, and video chats. The web browsing option is also available for children who can access any content online. Content sharing, such as screenshots, video clips, game play broadcasts, and user-generated data, is another feature of *PS5*. *PS5* provides a wide variety of parental control features ranging from content filtering to restricted communication. In content filtering, the application requires the user to enter the age of the child in account settings. Content and games are filtered according to age. Manual filtering is also permitted such that parents can customize and create a list of games that the children can access. Multiple chat options are provided to a parent to choose from, ranging from no chat features, chats limited to only friends, or open chats, whereby anyone can chat. In addition, *PS5* provides a feature that blocks the content created by other users. The daily screen time, monthly spending limit, and web browsing can be restricted by the parents. After the time or spending limit has been reached, parents receive notifications via their registered email.

Xbox is a gaming console developed by Microsoft [95]. Like *PS5*, *Xbox* provides a wide variety of games for children and adults. An account is required to play games at a minimum age of 18 years. Younger children can link their accounts to their parents to play games in *Xbox*. A wide range of chat features is available, such as chats with friends, group chat rooms, chats with unknown people while playing games, voice messaging, and video chats. Content-sharing features are also available. *Xbox* provides a Family Setting application

TABLE 3. Popular gaming platforms and their features.

Platforms	Platform type	Age restriction	Voice messages	Video chat	Content sharing	Chat filtering	Chat restriction	Monthly spend restriction	Customized blocking	Screen time limit	Additional parental control features
Roblox	App/ Website	No Restriction	Yes	No	No	Yes	Yes	Yes	Yes	No	N/A
Steam	App	13	Yes	No	Yes	Yes	Yes	No	Yes	No	N/A
PS5	Console/ App	18, under 18 can use PS through family account	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Xbox	Console/ App	18, under 18 can use *Xbox* through family account	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nintendo	Console/ Website	13, under 13 can use *Nintendo* through parent account	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

to manage parental controls, which can be installed on the mobile phone. This application is available in the Apple and Google Play stores. Thank to this application, the parents can set up accounts for their children. Several options are available, such as screen time limits, content filtering, and spending limits. Parents can approve or block any user who sends a friend request. Contacts, chatting with friends, group chats or chats with unknown people, and any item purchased can be restricted by parents. Moreover, *Xbox* generates a weekly or monthly report for all activities performed by the children, which is sent to the parents. *Xbox* also provides an opportunity to locate family members through the Family Setting application.

Nintendo provides a series of switches with a variety of games for all age groups [96]. The most appealing feature of the *Nintendo* Switch is the ease of carrying the device that can be used in travel because of its light weight. Switches can also be connected to TVs or computers to enjoy playing on large screens. The minimum age required to create an account for this platform is 13 years. However, parents can link their accounts to their children's younger than 13 years. *Nintendo* Switch provides multiple chat features including chats with friends, chats with unknown people in games, group chat rooms, and voice chats. *Nintendo* provides a *Nintendo* Switch Parental Controls mobile application for parents to restrict and monitor the activities of linked accounts. Many parental control options are available, such as setting screen time limits, content filtering, and spending limits for the *Nintendo* store. Chats with friends, group chats, chats with unknown people, item purchases, Internet browsing, and friends' request approval can all be restricted by parents.

3) INDUSTRY's PROTECTION MECHANISMS

The industry has also initiated actions to improve existing child protection mechanisms. For instance, Millicom, a leading provider of cable and mobile services dedicated to emerging markets in Latin America and Africa, partnered with UNICEF to map out the risks and opportunities faced by the telecommunications sector with respect to children's rights [97]. The partnership aimed to develop guidelines and tools for telecommunications companies to assess how their policies and processes might affect children's rights. Microsoft's *photoDNA* is a software devoted to tracking child sexual content by assigning a DNA to each photo. Since its launch, billions of photos have been analyzed [98]. Microsoft also developed a program to support national governments in establishing initiatives and action plans for the COP [99]. Safaricom has built upon children's rights in business principles by developing their own children's rights and business policy [100]. They highlight the importance of respecting children's rights and introducing business cases to do so. In partnership with Chicos.net, Disney's Amigos Conectados Project offers teachers, parents, and children in Latin America the digital literacy and citizenship skills necessary to fully engage in the digital future [101]. Thorn, a nonprofit entity that drives technological innovation to fight the sexual exploitation of children, developed a solution to help companies identify tools and practices that can help prevent their platforms from being used for child sexual exploitation [102]. LEGO collaborated with a key supplier in India to develop and implement training on child rights as a part of the LEGO Academy [103]. Lego's supplier guidelines help suppliers, including those providing digital marketing or product development services, to apply these guidelines to everyday business operations.

The gaming industry is looking into finding appropriate ways to control different kinds of risks, as discussed above, by improving parental control features and content filtering tools. However, these tools are insufficient to rectify multifaceted threats. In terms of content filtering, gaming platforms cannot completely filter inappropriate content. Available tools are limited in providing safety for gamers and result in disabling chat features, which are essential for online multiplayer games. Meanwhile, there is no enforcement from a regularity body to provide safety to children on the gaming platforms and invoke regulations to be followed by the industry.

Lately, 48% of the parents have used parental control features in the US [13] and 46% in the UK [14] to control the exposure of a child to inappropriate content, which indicates that more than half of the population in the US and UK are

not even aware of the usage of parental controls. Therefore, more than 50% of the children's population in the US and the UK is vulnerable to a multitude of risks, as discussed in Section III. Although international and national governments, policymakers, and the gaming industry are trying to create a safe environment for children, the threats related to online child victimization cases are increasing day by day, which shows the lack of research and policies that ensure the safety of a child online and specifically, in the gaming ecosystem. More advanced and efficient solutions are needed to enable children to play online while being safeguarded against predators' threats without sacrificing the ability to communicate with other players.

V. GAMING AND AI

A. AI APPLICATIONS IN GAMES

AI has become an integral part of the gaming environment covering various aspects of games. The most popular use of AI in games is the development of intelligent machines that can play games to entertain humans. On May 11, 1997, the IBM Deep Blue computer defeated the world chess champion in a six-game match [104]. In March 2016, Alpha Go developed by Google defeated the 18-time champion Lee Sedol with a 4-1 score [105]. In a recent AI research progression, many researchers developed AI agents to play a game independently or support players in a game [19], [20], [21], [22], [23], [24]. Another application of AI in games is player profiling, which is responsible for enhancing players' experience and is helpful in increasing the revenue of gaming platforms [106], [107]. AI algorithms have also been utilized to detect cheating in online games [108] and to test AI algorithms and bots [109].

Despite exhaustive research on the applications of AI in non-educational and educational games, the application of AI tools and ML algorithms for child protection on gaming platforms are used solely to detect child predatory behaviors in the context of chat logs. The authors in [25] present a text classification approach to detect sexual predators using real chat data provided by the game company MovieStarPlanet. Most research on the detection of child predatory behavior uses pseudo-victim data developed from the Perverted Justice (PJ) website [53] and the PAN12 dataset [54]. The data provided by MovieStarPlanet are divided into subsets for testing different approaches: bag of words (BOW), sentiment features, and rule-breaking features. During data preprocessing, all spelling mistakes were removed, which improved the classification accuracy. Different classification methods were deployed, namely naïve Bayes (NB), decision tree (DT), multilayer perceptron (MLP), k-nearest neighbor (k-NN) and support vector machine (SVM). A maximum accuracy of 92.51% was achieved when MLP was applied to all features. Testing was also performed on the PAN12 dataset and an accuracy of 93% was achieved.

The highly competitive environment of the participants in the Multiplayer Online Battle Arena (MOBA) games might lead to the emergence of undesirable toxic behavior. In [26], the authors examined whether it was possible to predict toxicity in MOBA games by developing a method for classifying toxic remarks. For this purpose, they used an NLP framework to detect profanity in the MOBA chat logs. An SVM classifier was trained on the feature-set based on the term frequency inverse document frequency (TF-IDF) of each word to predict the winning team. The accuracy of the classifier was used to demonstrate how toxicity is non-trivially linked to game success. In [27], the authors generated a dataset from the chat log of the World of Tanks. Their work presented a method of collecting data from the World of Tanks gaming environment and included the classification of the collected data using the NB, twin support sentiment analysis classifiers, and Microsoft Azure sentiment analysis. In [28], the dataset was developed using the data built in [110] and additional data collected from the Dota Ragnarok game. After preprocessing the data, a convolution neural network (CNN) was used to classify the chats and identify cyberbullying.

B. AVAILABLE DATASETS TO DETECT CHILD PREDATORS IN CHAT LOGS

It is important to note that public datasets containing chat logs of gaming platforms are not available, which limit the research related to child predatory behavior on gaming platforms [25], [26], [27], [28]. As more researchers enter this field, future research should attempt to be more proactive in collecting, labelling, and distributing gaming chat logs datasets. This would help to develop classifiers that efficiently identify predatory behavior on gaming platforms. Currently, two sources provide data related to real victims and child pedophiles in chat rooms of social media: the Perverted Justice website (PJ) [53] and the Sexual Predator Identification dataset [54].

1) PERVERTED JUSTICE

The Perverted Justice (PJ) is a non-profitable American organization that initiated an operation to detect child pedophiles in different chat room applications [53]. Police officers and volunteer adults were trained to pretend being minors to attract child pedophiles. When the user is confirmed to be a child pedophile, the LEA are alerted and informed, which helps the PJ team to convict 623 child pedophiles. Chat logs of trained adults and child pedophiles are available on the website. Photos of the convicted are also shared on the website. PJ is the only resource available for public chat logs of verified child pedophiles. Many researchers have used it in their work.

2) PAN12 SEXUAL PREDATOR IDENTIFICATION DATASET

The sexual predator identification competition was launched in 2012 by the Conference and Labs of the Evaluation Forum (CLEF) [54]. This dataset was collected using four different resources to cover three possible chat-feature scenarios: (1) normal non-sexual chats, (2) sexual chats among adults with mutual consent, and (3) child predators chats

with victims. Normal chats were collected from two sources [111], [112]. Sexual chats among adults were collected from *Omegle* data collection [113]. Child pedophile chats were collected from PJ. The data were designed to mimic an actual situation in which the ratio of child pedophiles chats is exceedingly less than that of normal chats. There were a total of 66,927 conversations in the training dataset, consisting of 2,723 predatory chats and 64,911 non-predatory chats. In the test dataset, the conversations are 15,5128 out of which 5,321 are predatory chats and the remaining are non-predatory chats.

C. AI PROTECTION MECHANISMS AGAINST CHILD PREDATORS

The literature that detects child pedophiles on chatting platforms can be subdivided into two main categories: conversational models that detect child predators in a chat environment, and classification models that detect child predators in chat logs. Table 4 provides a summary of the chatbots proposed in the literature along with the models and results. Negobot was the first prominent contribution in developing an AI agent to detect a child pedophile behavior on chatting platforms [29]. Negobot was trained on a dataset consisting of 377 chats taken from the PJ website [53]. When the conversation is started, the user sends a text to Negobot, which processes the data using a conversational unit by removing all emotions and slangs, translating when needed, creating a meaningful sentence, and feeding it into a high-performance information retrieval (IR) tool, which evaluates the similarity index of the conversation with conversations taken from the PJ. Negobot uses a structure of seven chatter-bots specifically designed to perform in different scenarios depending on the conversation topic. AIML is used to provide Negobot with the ability to converse with users. Negobot collects as much information as possible by applying the game theory. Finally, the chat was classified into levels assigned by the system from -1 to $+3$ such that -1 is assigned to the chat demonstrating the least predatorial content while $+3$ indicates a high likelihood of predatorial content from a suspect. The system was tested by using two chats, one of which was aggressive and the other was passive. System testing is not extensive enough to claim the performance of Negobot.

A conceptual platform called BotHook was proposed in [30]. BotHook is a chatbot containing three major modules. The first module is capture, classification and analysis of cybercriminals and cyberpedophiles module (CCAM), which attracts and analyzes attacks on the system. The second module is bot module (BOTM), which is responsible for an interactive chat with suspects without showing its identity as a bot. The third module is the pedophile trend characterization module (PTCM), which assigns the value of a pedophile trend to users based on their chats. This work is a theoretical proposal, whereby practical implementation and testing were not performed. In [36], a chatbot was designed to collect data from the website Omegle [113]. An SVM classifier was used as an emotional classifier while multinomial naïve Bayes (MNB) was used as an opinion classifier. The users who were not interested in child predatory behavior, were labelled as ''indifferent''. The people who showed an interest in predatory behavior without committing an offense were labelled as ''interested''. Finally, the users who actually behaved as child predators were labelled as ''perverts''. Later, this work was extended to [42] and the results were improved by deploying the chatbot for 50 days and collecting information from 7,199 users.

In 2012, a competition was conducted to identify sexual predators in online chats. A dataset PAN12 was developed to perform two tasks: the first task was to identify the predator among all users in different conversations and the second task was to identify the lines or parts of the conversation that were distinctive to predatory behavior [54]. The authors of [54] provided a summary of all submissions to the competition. This study is a prominent contribution to the field of identifying a predatory behavior in chats, as it provides a publicly available dataset containing predatory and non-predatory chats.

In the literature, the detection of child predators was modeled as a text classification task. Table 5 summarizes the classification models proposed in the literature to detect the child predatory behavior. The pipeline used for text classification is illustrated in Fig. 1. In [32], a classifier was built to label chats as grooming or non-grooming. The dataset was compiled using 105 grooming chats collected from the PJ website and 45 non-grooming chats collected from the website in [114]. Seventy grooming chats and 30 non-grooming chats were used for training purposes. The test data consisted of 35 grooming and 15 non-grooming chats. Seventeen characteristics of grooming chats were identified in the training and test datasets. It was deducted that chats can be classified using the number of grooming characteristics found in a conversation. The proposed classifier labelled a chat as grooming if 11 out of 17 characteristics were found in the conversation. Otherwise, the chat was labeled as non-grooming. The proposed classifier achieved an accuracy of 96.8% using the built database. For the same dataset, the accuracies of the SVM and KNN were 98.6% and 97.8%, respectively.

The authors of [33] used the PAN13 dataset [115] to train a classifier using a five-step process to detect a child predatory behavior in chat logs. As the PAN13 dataset was generated for age and gender detection in chat logs, it was included in the first step of the preprocessing of data, which eliminated all meta data except the words included in the conversation and conversation ID. In the next step, the feature extraction of the BOW and TF-IDF were applied using the fuzzy rough feature selection (FRFS) method, which was used to identify the most important features that describe the dataset. The data were then classified using Gaussian naïve Bayes, random forest (RF), logistic regression (LR), and AdaBoost. Using different normalization techniques, such as l_1, l_2, and power normalization (PN), the highest accuracy was achieved for the LR classifier using BOW, PN-l_2 normalization, and LR

TABLE 4. Proposed chatbots to detect child predators.

Ref.	Conversational model	Classification model	Dataset Source	Results
[29]	AIML	Game Theory	PJ	Classified two chats
[30]	BOTM	PTCM	N/A	N/A
[36]	AIML, LSTM-NN	Emotion classifier (SVM), Opinion classifier (MNB)	Omegle	Classified 35 chats as indifferent, suspected and perverts
[42]	AIML, LSTM-NN	Emotion classifier (SVM), Opinion classifier (MNB)	Omegle	Classified 7,199 users as indifferent, interested and perverts

classification. This study was extended to [34], by developing a dataset using the two sources in [53] and [112]. Additional classifiers, that is, linear and RBF coordinate descent fuzzy twin support vector machines (CDFTSVM), were tested. The authors of [35], [38], [39], [40], and [45] used the text classification approach to detect predatory behavior in chat logs using a wide variety of well-known classifiers such as SVM, CNN, deep artificial networks (DAN), RF, and NB. In [41], the authors used the text classification method with three types of features: textual, behavioral, and demographic features. The SVM and Bernoulli NB classifiers were used to classify the chats into two categories. The dataset used in this study was collected from PJ [53] and from [116]. In [37], three different approaches to detecting child predators were based on the message, author, and conversation using a wide variety of classifiers, such as LR, ridge, NB, SVM and NN.

In [43], a two-stage classifier was used whereby the messages were classified in the first stage and the whole conversation was classified in the second stage to detect predatory behavior in chat logs. In [44], a two-stage classification was implemented, such that the chat logs were classified in the first stage using a wide variety of well-known classifiers and the results from the first stage were used for classification in the second stage using a soft voting-based ensemble. In [46], an approach was presented to address class imbalance using hybrid sampling and class re-distribution to build an augmented dataset, and then classify it by using histogram gradient boosted decision trees (HGBDT). In [49], the authors also used a two-stage classifier whereby the chat logs were classified in the first stage using different classifiers that labelled the chats as non-predatory and predatory. In the second stage, false positive events were minimized.

In [47], the authors explained how the task in the digital forensic investigation process could be mapped to ML methods. After proposing a mapping between digital forensic and ML methods, the classification of chat logs was performed for predatory and non-predatory chats using LR, XGBoost, MLP and long short-term memory (LSTM). In [48], the authors presented the detection of child grooming behavior in chat logs using an SVM. An age detection mechanism using deep neural networks (DNN) was introduced to reduce the false positives, as only chats with children were classified, whereas others were discarded. The dataset was developed using two sources [53], [117]. In [50], two types of features were extracted: vocabulary-based and emotional-based, which were fed into a variety of classifiers, such as DT, SVM, and RF to classify the conversation as predatory or non-predatory.

FIGURE 1. Pipeline for detecting child predatory behavior in chat logs using text classification model.

In [51], the authors presented an approach for classifying a child grooming conversation using a two-step approach. In the first step, the features were extracted using a CNN, and classification was performed using MLP. This approach used sentiment analysis in connection with lexical features. In [52], the detection of predatory behavior was proposed using bidirectional encoder representations from transformers and feedforward neural networks.

The protection mechanisms proposed in the literature can be broadly divided into two categories: chatbots to detect predators in online chat logs and text classification methods using traditional ML classification algorithms. The most common preprocessing methods used in the literature are tokenization, stemming, and lemmatization in addition to manual parsing and the removal of slangs and stop words. The widely used feature extraction methods are BOW and TF-IDF, while the extensively used classification methods are SVM and CNN. It is worth noting that the developed chatbots were not tested on balanced and large datasets because of the limited availability of the collected data. Moreover, classification methods used to detect predatory behavior have not been applied to gaming platforms to avert real-time threats while children are playing and interacting. Better solutions using chatbots/ML tools should be devised to seamlessly integrate the AI model into the gaming platform, thereby allowing immediate online detection and prevention of different predatory threats.

VI. DISCUSSION AND RECOMMENDATIONS

The explosion of information and communication technology (ICT) has created unprecedented opportunities for both children and young people. Both benefits and undeniable risks exist. Children are vulnerable to many kinds of threats in online gaming environments, such as inappropriate content and predatory behavior. Inappropriate content includes adult, violent or gory content. The influence of violent games on children's mental health varies. Factors such as physical

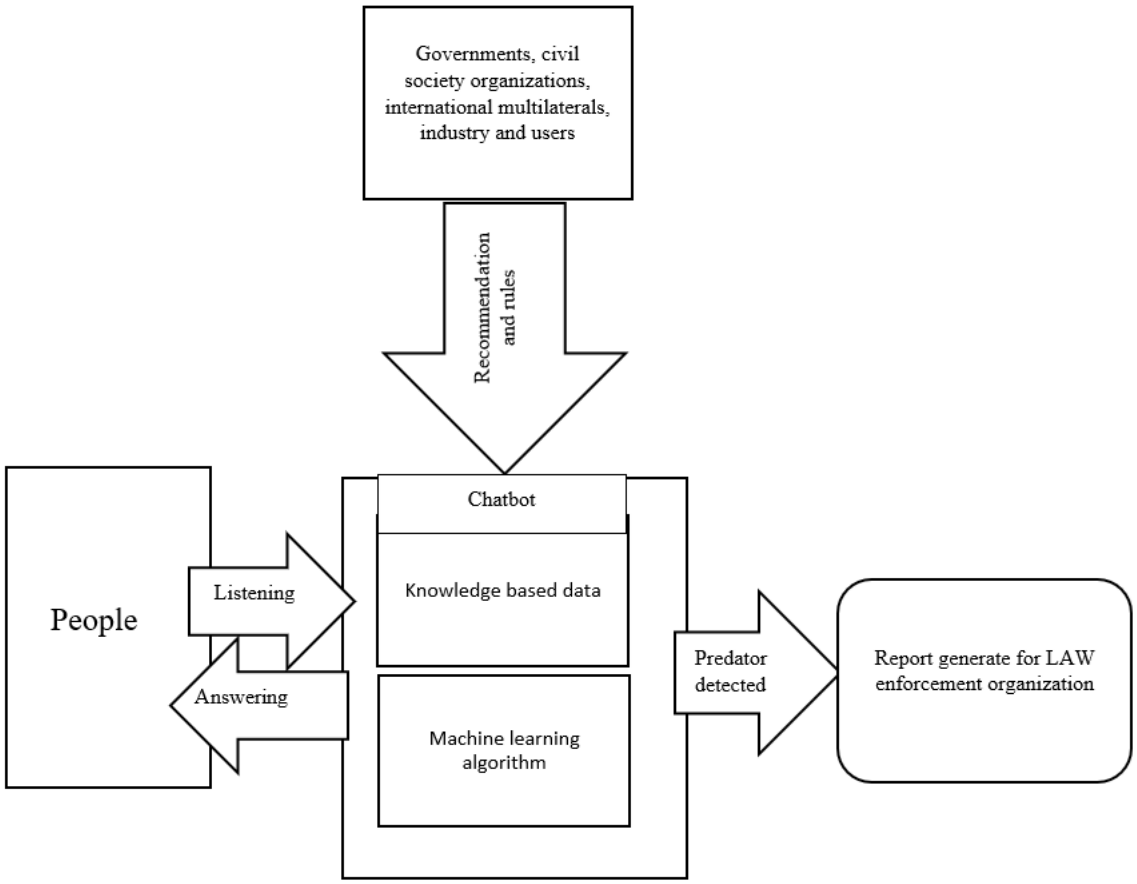

FIGURE 2. Framework for child protection on a gaming platform.

abuse, divided family, toxic environment at home, and predatory behavior from peers or family members are considered strong catalysts that influence the mental health of children. Predatory behavior, which includes child bullying, harassment, pedophilia, and grooming, can adversely affect children's mental health and traumatize them. The most common issues reported by victims of child predators are depression, fear, panic attacks, lack of trust in their relations, anxiety, self-harm, and suicides [118]. Child predators not only adversely influence children but also devastate their families.

The industry has made a fair contribution to enhancing the COP by adding parental controls and age rating guidance to the gaming environment. However, the effectiveness of parental controls remains controversial. Paid tools, such as NetAlert [119], NetNanny [120], and Bark [121], provide options to parents but are limited to content filtering, and thereby, they only address the content threat. It is observed that many children do not abide by the age limitation requirements placed on games registration platforms [12], [82]. The mechanisms used to determine the age of a user are not sufficiently robust to detect the actual age of the consumer [122]. Children can easily access age-inappropriate content by entering a fake age into the sign-up process. An AI-enabled facial recognition system to identify the age of the user could be a solution to this problem but it may violate the children's right to privacy. A better solution is that the parents would always be aware of the content used by the children and would adopt appropriate parental control options to safeguard their children from online threats. The lack of awareness of parents is another important issue yet to be addressed. As indicated in Section V, surveys show that more than half of the parents in the US and UK are not aware of the usage of parental controls [13], [14].

On the other hand, the available control tools have not proven to be a robust solution to overcome multitude threats to children. Algorithmically filtered content can significantly influence child development, opinions, values, and habits. The filter creates an isolation bubble, which restricts children from exploring a wide variety of opinions and ideas [82]. To combat contact and conduct threats, parental controls can be used to restrict the children's account, so that they cannot participate in chats with unknown people. However, in many cases, peers are responsible for cyberbullying, defamation, or the exclusion of victims. So, by allowing children to chat with friends only does not provide a protection to children against these threats. Moreover, if the chat features are fully blocked the actual essence of most of games vanishes. More robust solutions are needed to protect against risks

TABLE 5. Proposed classification models to detect child predators.

Ref.	Classification	Preprocessing	Feature Extraction Method	Feature Selection Method	Classifier
[25]	Binary	Yes	BOW, TF-IDF	NA	DT, MLP, LR, KNN, SVM
[26]	Binary	Yes	TF-IDF	NA	SVM
[32]	Binary	Yes	TF-IDF	17 defined characteristics	SVM, KNN
[33]	Binary & Multi-label	Yes	BOW, TF-IDF	FRFS	GNB, RF, AdaBoost, LR
[34]	Binary & Multi-label	Yes	BOW, TF-IDF	FRFS	GNB, RF, LR AdaBoost, LCDFTSVM, RCDFTSVM
[35]	Binary& Multi-label	NA	BOW, TF-IDF	FRFS	Backpropagation neural network
[28]	Binary	Yes	None	NA	CNN
[37]	Binary	Yes	BOW, TF-IDF	NA	LR, Ridge, NB, SVM, NN
[38]	Binary	Yes	Term frequency inverse document frequency inverse class space density frequency (TF.IDF.ICSDF)	FRFS	SVM
[39]	Multi-label	Yes	NA	NA	DAN, CNN
[40]	Binary	Yes	BOW, TF-IDF	NA	SVM linear, SVM non-linear, RF, NB
[41]	Binary	Yes	Manual textual, behavioral and demographical feature extraction	NA	SVM, Bernoulli NB
[43]	Binary	NA	NA	NA	Recurrent neural network
[44]	Binary	Yes	BOW, TF, TF-IDF	NA	MNB, Bernoulli NB, SVM, NN, KNN, LR, RF, DT
[45]	Binary	Yes	Continuous BOW, skip gram	NA	NLP, CNN
[46]	Binary	Yes	NA	NA	HGBDT
[47]	Binary	Yes	NA	NA	LR, XGBoost, MLP, LSTM
[48]	Binary	Yes	LIWC	NA	RF, NB, SVM, KNN, AdaBoost-DNN
[49]	Binary	Yes	Word embedding aggression	NA	Linear Discrimination Analysis, SVM, RF, LASSO, Generalized boosting machine
[50]	Binary	Yes	BOW, MoodBook	NA	DT, SVM, RF
[51]	Binary	Yes	CNN	NA	MLP CNN
[52]	Binary	Yes	Pre-trained bidirectional encoder representations from transformers (BERT)	NA	BERT_frozen, BERT_tuned

without compromising the quality of playing and socializing on online platforms.

Parents/caretakers' and educators' awareness of threats to children is another important aspect. To address this, governments, civil society organizations, and international multilaterals are playing their roles by developing interactive user-friendly tools to enhance the knowledge of children and parents about online threats and ways to combat them. The Sango developed by ITU [78] and the social-emotional learning tool developed by UNICEF [17] are great initiatives at the international level. The interactive platform developed by the EU also helps enhance the knowledge of parents and children [18]. At the national level, governments are taking actions to better protect children. For instance, the UAE government developed the kidX tool to provide children with awareness about their rights and to increase their knowledge about functions and services provided by the government to protect them online [16].

A concerted and collaborative effort to reduce the risks of the digital world, particularly the risks targeting children and adolescents, is needed among multiple stakeholders, such as governments, civil society organizations, international multilaterals, industry, and users. Policymakers and regulators must continue striving for higher safety and protective measures to keep children safe online. We live in an ever-connected, digitized world, with children increasingly using the Internet and digital technologies for a multitude of purposes, including their learning, gaming, and social connection. Protecting children online and in digital technologies such as gaming is a global issue; however, a coordinated global response is lacking to protect the increased number of children connected to the Internet through web browsing, education, social media, gaming, and entertainment websites and applications. In recent years, the COVID-19 pandemic has perpetuated a great surge in the number of children and young people using the Internet and digital technologies. Certainly, the Internet provides opportunities for children's learning and growth; however, it also exposes them to many types of risks. This hyper-connectivity exacerbates the exposure of children to a multitude of risks, which is becoming a global phenomenon. Policymakers need to map out urgent strategies and plans aimed at tackling challenges for the protection and safety of children online. Therefore, it is imperative to establish global capacity-building programs, launch collaborative and multi-stakeholder initiatives, strategize transnational legislation and laws, and develop standards and regulations with

frameworks and programs to protect children's welfare and well-being online.

The following are recommendations that are, in part, in response to systemic, industrial, regulatory, and protection gaps:

- Establish a global authority or governance body to ensure children's rights are protected from online harm.
- Develop a global strategy aligning existing international normative frameworks for children's rights and provisions with the requisite multi-stakeholder policy frameworks coordinating industry with intergovernmental bodies such as UNICEF.
- Publish global standards, regulations, and guidelines on child online protection and safety to foster safer Internet/digital technologies/gaming for children at the industry, global, regional, national, familial, and individual levels needed to enable children's online safety and protection.
- Improve coordination among multiple stakeholders including international bodies, governments, law enforcement agencies, industry, policymakers, academics, and civil society organizations.
- Increase awareness, legislative and regulatory measures, and mechanisms at strategic, tactical, and operational levels.
- Provide opportunities to improve innovations and build capacity and capability in child online protection and safety literacy, training, re-skilling, and upskilling, as well as children's ability to protect themselves.

Over the past two decades, research to explore the use of AI tools and ML algorithms for children's online safety has increased. However, the main focus of this research remains restricted to the detection of predatory behavior in chat logs. This solution is not enough to protect the child from predators, as it is only a step toward building a mechanism to protect children online.

It is also observed that the focus of research on detecting predatory behavior is on chat logs pertaining to social media's chatting platforms [123], [124], [125], [126], [127]. However, gaming platforms have rarely been explored. The main reason for this research gap is the unavailability of public datasets of online game players chat logs. Owing to privacy issues, the data are not stored or saved, which makes it difficult to build a dataset. We recommend that the industry and researchers collect data on gaming platforms after obtaining the consent of the parents and children to share their chat logs to build the required datasets.

Distributed platforms chatbots have been developed to detect child predators on chatting platforms [29], [36], [42]. However, owing to the limited availability of relevant datasets, bots have not been tested on balanced and large datasets. We recommend generating a robust mechanism that utilizes an AI-based chatbot to combat child predators in gaming environments by addressing different threats through the collaboration of governments, civil society organizations, international multilaterals, industry, and users. The framework of a proposed AI-based protection mechanism is shown in Fig. 2. The chatbot should be designed to consider the recommendations by stakeholders to combat various threats including cyberbullying and sexual predation without compromising the rights of children to play, leisure, and culture (Article 31). Advanced age detection methods from chat logs can be utilized to detect if children are behaving as adults to access mature content, and if adults are pretending to be children to illegally bond with real children. The following are our proposed recommendations for an efficient chatbot that uses AI tools:

- Combat the content threat by providing a robust filtering of games content and broadcasts shared by other players in group chats.
- Combat the contact and conduct threats by providing a robust solution to effectively detect child predators on a live chatting platform and report the predators before they could cause any harm.

VII. CONCLUSION

In this study, we conducted a systematic survey of online child protection mechanisms covering the multifaceted threats and efforts made to enhance child protection. The motivation behind this work is to highlight the existing gaps in research and present available solutions.

In our survey, we observed the following limitations: 1) A collaborative global response is lacking to protect the increased number of children connected to the Internet through web browsing, education, social media, gaming, and entertainment websites and applications. This is presumably in part due to the systemic, industrial, regulatory, and protection gaps accentuated by the absence of a global authority that ensures children's protection, a global strategy to protect children online, global clear standards, regulations and guidelines on the COP and coordination among multiple stakeholders; 2) in the last two decades, researchers were attempted to develop AI tools to enhance children protection online, but there is no robust system or framework deployed in research or in the industry; and 3) AI tools and ML algorithms are used to detect child predatory behavior in a limited context of chat logs that only pertain to chatting platforms. Thus, the use of AI for child protection is limited, which does not provide a robust solution to the threats that children face in the gaming environment. As child protection in this particular environment has not been sufficiently investigated, intelligent and robust mechanisms are needed to address child protection in gaming by detecting child predators and blocking them before they can approach children and harm them emotionally or physically.

We recommend developing an AI-based robust chatbot to be integrated in gaming platforms for efficiently detecting and blocking child predators under the guidelines provided by the different stakeholders, in particular, engineers, psychologists, sociologists, law enforcement, and data analysts who need to provide new insight into understanding, detecting, and stopping predatory behavior in the gaming environment.

REFERENCES

[1] S. Kemp. (Jan. 2021). Digital 2021. Kepios Pte. Ltd, Singapore. Accessed: Jan. 5, 2022. [Online]. Available: https://datareportal.com/reports/digital-2021-global-overview-report

[2] P. Stalker, S. Livingstone, D. Kardefelt-Winther, and M. Saeed, "Growing up in a connected world," UNICEF, Innocenti, Florence, Itlay, Tech. Rep. inorer1060, Nov. 2019.

[3] T. Weru, J. Sevilla, J. Olukuru, L. Mutegi, and T. Mberi, "Cyber-smart children, cyber-safe teenagers: Enhancing internet safety for children," in *Proc. IST-Africa Week Conf. (IST-Africa)*, Windhoek, Namibia, May 2017, pp. 1–8, doi: 10.23919/ISTAFRICA.2017.8102292.

[4] (Oct. 2019). *Resolution Violent Video Games*. American Psychological Association, Washington, DC, USA. Accessed: Dec. 8, 2021. [Online]. Available: http://www.apa.org/about/policy/violent-video-games.aspx

[5] T. Nuangjumnong, "The effects of gameplay on leadership behaviors: An empirical study on leadership behaviors and roles in multiplayer online battle arena games," in *Proc. Int. Conf. Cyberworlds*, Santander, Spain, Oct. 2014, pp. 300–307, doi: 10.1109/CW.2014.48.

[6] N. Pobiedina, J. Neidhardt, M. D. C. Calatrava Moreno, L. Grad-Gyenge, and H. Werthner, "On successful team formation: Statistical analysis of a multiplayer online game," in *Proc. IEEE 15th Conf. Bus. Informat.*, Vienna, Austria, Jul. 2013, pp. 55–62, doi: 10.1109/CBI.2013.17.

[7] M. H. Hussein, S. H. Ow, L. S. Cheong, M.-K. Thong, and N. A. Ebrahim, "Effects of digital game-based learning on elementary science learning: A systematic review," *IEEE Access*, vol. 7, pp. 62465–62478, 2019, doi: 10.1109/access.2019.2916324.

[8] J. S. Kinnebrew, S. S. Killingsworth, D. B. Clark, G. Biswas, P. Sengupta, J. Minstrell, M. Martinez-Garza, and K. Krinks, "Contextual markup and mining in digital games for science learning: Connecting player behaviors to learning goals," *IEEE Trans. Learn. Technol.*, vol. 10, no. 1, pp. 93–103, Jan. 2017, doi: 10.1109/TLT.2016.2521372.

[9] O. Dele-Ajayi, J. Sanderson, R. Strachan, and A. Pickard, "Learning mathematics through serious games: An engagement framework," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Erie, PA, USA, Oct. 2016, pp. 1–5, doi: 10.1109/FIE.2016.7757401.

[10] C.-H. Ko, J.-Y. Yen, C.-S. Chen, Y.-C. Yeh, and C.-F. Yen, "Predictive values of psychiatric symptoms for internet addiction in adolescents: A 2-year prospective study," *Archives Pediatrics Adolescent Med.*, vol. 163, no. 10, pp. 937–943, Oct. 2009, doi: 10.1001/archpediatrics.2009.159.

[11] J.-L. Wang, J.-R. Sheng, and H.-Z. Wang, "The association between mobile game addiction and depression, social anxiety, and loneliness," *Frontiers Public Health*, vol. 7, p. 247, Sep. 2019, doi: 10.3389/fpubh.2019.00247.

[12] *Child Rights and Online Gaming: Opportunities & Challenges for Children and the Industry*, UNICEF, Innocenti, Florence, Itlay, Aug. 2019.

[13] Statista Research Department. *Percentage Parents Placing Limits Children's Media Consumption United States 2019 2020*. Accessed: Dec. 5, 2021. [Online]. Available: https://www.statista.com/statistics/232345/parental-control-over-childrens-media-consumption-in-the-us/#:~:text=Parental%20control%20over%20children%27s%20media%20consumption%20U.S.%202019%2D2020&text=A%202020%20study%20revealed%20that,slightly%20from%20the%20previous%20year

[14] I. Taylor. *UKIE: Only 19% Parents Set Enforce Screen Time Limits for Their Children*. Accessed: Dec. 5, 2021. [Online]. Available: https://www.gamesindustry.biz/articles/2018-09-14-digital-school-house-only-19-percent-of-parents-set-and-enforce-screen-time-limits-for-their-children

[15] International Telecommunication Union. *Keeping Children Safe Online*. Accessed: May 9, 2022. [Online]. Available: https://www.itu-cop-guidelines.com/

[16] The United Arab Emirates' Government portal. *Gamification*. Accessed: May 16, 2022. [Online]. Available: https://u.ae/en/about-the-uae/digital-uae/gamification

[17] *Tilli: Gamified Social-Emotional Learning for Child Online Safety*. Accessed: May 16, 2022. [Online]. Available: https://gdc.unicef.org/resource/tilli-gamified-social-emotional-learning-child-online-safety

[18] European Commission. *Enhancing Professionals' Capacity to Deal With Child Victims*. Accessed: May 16, 2022. [Online]. Available: http://childprotect.eu/

[19] E. Lebedeva and J. A. Brown, "Companion AI for starbound game using utility theory," in *Proc. Int. Conf. Nonlinearity, Inf. Robot. (NIR)*, Innopolis, Russia, Dec. 2020, pp. 1–5, doi: 10.1109/NIR50484.2020.9290164.

[20] I. Oh, S. Rho, S. Moon, S. Son, H. Lee, and J. Chung, "Creating pro-level AI for a real-time fighting game using deep reinforcement learning," *IEEE Trans. Games*, vol. 14, no. 2, pp. 212–220, Jun. 2022, doi: 10.1109/TG.2021.3049539.

[21] H. Baier, A. Sattaur, E. J. Powley, S. Devlin, J. Rollason, and P. I. Cowling, "Emulating human play in a leading mobile card game," *IEEE Trans. Games*, vol. 11, no. 4, pp. 386–395, Dec. 2019, doi: 10.1109/TG.2018.2835764.

[22] S. Ariyurek, A. Betin-Can, and E. Surer, "Automated video game testing using synthetic and humanlike agents," *IEEE Trans. Games*, vol. 13, no. 1, pp. 50–67, Mar. 2021, doi: 10.1109/TG.2019.2947597.

[23] M. Ishihara, S. Ito, R. Ishii, T. Harada, and R. Thawonmas, "Monte-Carlo tree search for implementation of dynamic difficulty adjustment fighting game AIs having believable behaviors," in *Proc. IEEE Conf. Comput. Intell. Games (CIG)*, Maastricht, The Netherlands, Aug. 2018, pp. 1–8, doi: 10.1109/CIG.2018.8490376.

[24] W. Konen, "General board game playing for education and research in generic AI game learning," in *Proc. IEEE Conf. Games (CoG)*, London, U.K., Aug. 2019, pp. 1–8, doi: 10.1109/CIG.2019.8848070.

[25] Y.-G. Cheong, A. K. Jensen, E. R. Gudnadottir, B.-C. Bae, and J. Togelius, "Detecting predatory behavior in game chats," *IEEE Trans. Comput. Intell. AI Games*, vol. 7, no. 3, pp. 220–232, Sep. 2015, doi: 10.1109/TCIAIG.2015.2424932.

[26] M. Martens, S. Shen, A. Iosup, and F. Kuipers, "Toxicity detection in multiplayer online games," in *Proc. Int. Workshop Netw. Syst. Support for Games (NetGames)*, Zagreb, Croatia, Dec. 2015, pp. 1–6, doi: 10.1109/NetGames.2015.7382991.

[27] S. Murnion, W. J. Buchanan, A. Smales, and G. Russell, "Machine learning and semantic analysis of in-game chat for cyberbullying," *Comput. Secur.*, vol. 76, pp. 197–213, Jul. 2018.

[28] J. A. Cornel et al., "Cyberbullying detection for online games chat logs using deep learning," in *Proc. IEEE 11th Int. Conf. Humanoid, Nanotechnol., Inf. Technol., Commun. Control, Environ., Manage. (HNICEM)*, Laoag, Philippines, Nov. 2019, pp. 1–5, doi: 10.1109/HNICEM48295.2019.9072811.

[29] C. Laorden, P. Galán-García, I. Santos, B. Sanz, J. M. Hidalgo, and P. G. Bringas, "Negobot: A conversational agent based on game theory for the detection of paedophile behaviour," in *Proc. Int. Joint Conf.*, Berlin, vol. 189, Jan. 2013, pp. 261–270.

[30] P. Zambrano, M. Sanchez, J. Torres, and W. Fuertes, "BotHook: An option against cyberpedophilia," in *Proc. 1st Cyber Secur. Netw. Conf. (CSNet)*, Rio de Janeiro, Brazil, Oct. 2017, pp. 1–3, doi: 10.1109/CSNET.2017.8241994.

[31] A. H. Alduailej and M. B. Khan, "The challenge of cyberbullying and its automatic detection in Arabic text," in *Proc. Int. Conf. Comput. Appl. (ICCA)*, Sep. 2017, pp. 389–394, doi: 10.1109/COMAPP.2017.8079791.

[32] F. E. Gunawan, L. Ashianti, and N. Sekishita, "A simple classifier for detecting online child grooming conversation," *Telecommun. Comput. Electron. Control*, vol. 16, no. 3, pp. 1239–1248, Jun. 2018.

[33] Z. Zuo, J. Li, P. Anderson, L. Yang, and N. Naik, "Grooming detection using fuzzy-rough feature selection and text classification," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Rio de Janeiro, Brazil, Jul. 2018, pp. 1–8, doi: 10.1109/FUZZ-IEEE.2018.8491591.

[34] P. Anderson, Z. Zuo, L. Yang, and Y. Qu, "An intelligent online grooming detection system using AI technologies," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, New Orleans, LA, USA, Jun. 2019, pp. 1–6, doi: 10.1109/FUZZ-IEEE.2019.8858973.

[35] Z. Zuo, J. Li, B. Wei, L. Yang, F. Chao, and N. Naik, "Adaptive activation function generation for artificial neural networks through fuzzy inference with application in grooming text categorisation," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, New Orleans, LA, USA, Jun. 2019, pp. 1–6, doi: 10.1109/FUZZ-IEEE.2019.8858838.

[36] J. Murcia Trivino, S. Moreno Rodriguez, D. O. Diaz Lopez, and F. Gomez Marmol, "C3-sex: A chatbot to chase cyber perverts," in *Proc. IEEE Int. Conf. Dependable, Autonomic Secure Comput., Intl. Conf. Pervasive Intell. Comput., Intl. Conf. Cloud and Big Data Comput., Intl. Conf. Cyber Sci. Technol. Congr. (DASC/PiCom/CBDCom/CyberSciTech)*, Fukuoka, Japan, Aug. 2019, pp. 50–57, doi: 10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00024.

[37] P. Bours and H. Kulsrud, "Detection of cyber grooming in online conversation," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Delft, The Netherlands, Dec. 2019, pp. 1–6, doi: 10.1109/WIFS47025.2019.9035090.

[38] N. R. Sulaiman and M. Md. Siraj, "Classification of online grooming on chat logs using two term weighting schemes," *Int. J. Innov. Comput.*, vol. 9, no. 2, pp. 43–50, Nov. 2019.

[39] T. R. Ringenberg, K. Misra, and J. T. Rayz, "Not so cute but fuzzy: Estimating risk of sexual predation in online conversations," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Bari, Italy, Oct. 2019, pp. 2946–2951, doi: 10.1109/SMC.2019.8914528.

[40] P. R. Borj and P. Bours, "Predatory conversation detection," in *Proc. Int. Conf. Cyber Secur. Emerg. Technol. (CSET)*, Doha, Qatar, Oct. 2019, pp. 1–6, doi: 10.1109/CSET.2019.8904885.

[41] S. Andleeb, R. Ahmed, Z. Ahmed, and M. Kanwal, "Identification and classification of cybercrimes using text mining technique," in *Proc. Int. Conf. Frontiers Inf. Technol. (FIT)*, Doha, Qatar, Dec. 2019, pp. 227–2275, doi: 10.1109/FIT47737.2019.00050.

[42] J. I. Rodríguez, S. R. Durán, D. Díaz-López, J. Pastor-Galindo, and F. G. Mármol, "C3-sex: A conversational agent to detect online sex offenders," *Electronics*, vol. 9, no. 11, p. 1779, Oct. 2020, doi: 10.3390/electronics9111779.

[43] J. Kim, Y. J. Kim, M. Behzadi, and I. G. Harris, "Analysis of online conversations to detect cyberpredators using recurrent neural networks," in *Proc. 1st Int. Workshop Social Threats Online Conversations, Understand. Manage.*, Marseille, France, May 2020, pp. 15–20.

[44] M. A. Fauzi and P. Bours, "Ensemble method for sexual predators identification in online chats," in *Proc. 8th Int. Workshop Biometrics Forensics (IWBF)*, Apr. 2020, pp. 1–6, doi: 10.1109/Iwbf49977.2020.9107945.

[45] G. Isaza, F. Muñoz, L. Castillo, and F. Buitrago, "Classifying cybergrooming for child online protection using hybrid machine learning model," *Neurocomputing*, vol. 484, pp. 250–259, May 2022, doi: 10.1016/j.neucom.2021.08.148.

[46] P. R. Borj, K. Raja, and P. Bours, "Detecting sexual predatory chats by perturbed data and balanced ensembles," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, Sep. 2021, pp. 1–5, doi: 10.1109/BIOSIG52210.2021.9548303.

[47] C. H. Ngejane, J. H. P. Eloff, T. J. Sefara, and V. N. Marivate, "Digital forensics supported by machine learning for the detection of online sexual predatory chats," *Forensic Sci. Int., Digit. Invest.*, vol. 36, Mar. 2021, Art. no. 301109, doi: 10.1016/j.fsidi.2021.301109.

[48] K. S Kirupalini, A. Baskar, A. Ramesh, G. Rengarajan, S. Gowri, S. Swetha, and D. Sangeetha, "Prevention of emotional entrapment of children on social media," in *Proc. Int. Conf. Emerg. Techn. Comput. Intell. (ICETCI)*, Hyderabad, India, Aug. 2021, pp. 95–100, doi: 10.1109/ICETCI51973.2021.9574068.

[49] Y. Singla, "Detecting sexually predatory behavior on open-access online forums," in *Proc. Res. Appl. Adv. Intell. Syst. Comput.*, Kolkata, India, vol. 1355, Jun. 2021, pp. 27–40, doi: 10.1007/978-981-16-1543-6_3.

[50] M. A. Wani, N. Agarwal, and P. Bours, "Sexual-predator detection system based on social behavior biometric (SSB) features," *Proc. Comput. Sci.*, vol. 189, pp. 116–127, May 2021.

[51] S. Preub, "Automatically identifying online grooming chats using CNN-based feature extraction," in *Pro. 17th Conf. Natural Lang. Process. (KONVENS)*, Dusseldorf, Germany, Sep. 2021, pp. 137–146.

[52] N. Agarwal, T. Unlu, M. A. Wani, and P. Bours, "Predatory conversation detection using transfer learning approach," in *Proc. 7th Int. Conf. Mach. Learn., Optim., Data Sci.*, Grasmere, U.K., vol. 13163, Oct. 2022, pp. 488–499, doi: 10.1007/978-3-030-95467-3_35.

[53] Perverted Justice Foundation. *Perverted-Justice.com Archives*. Accessed: Dec. 5, 2021. [Online]. Available: http://www.perverted-justice.com/

[54] G. Inches and F. Crestani, "Overview of the international sexual predator identification competition at PAN-2012," in *Proc. CLEF (Online Work. Notes/Labs/Workshop)*, vol. 30, Sep. 2012.

[55] C. H. Ngejane, G. Mabuza-Hocquet, J. H. P. Eloff, and S. Lefophane, "Mitigating online sexual grooming cybercrime on social media using machine learning: A desktop survey," in *Proc. Int. Conf. Adv. Big Data, Comput. Data Commun. Syst. (icABCD)*, Durban, South Africa, Aug. 2018, pp. 1–6, doi: 10.1109/ICABCD.2018.8465413.

[56] S. Salawu, Y. He, and J. Lumsden, "Approaches to automated detection of cyberbullying: A survey," *IEEE Trans. Affect. Comput.*, vol. 11, no. 1, pp. 3–24, Jan. 2020, doi: 10.1109/TAFFC.2017.2761757.

[57] A. Razi, S. Kim, A. Alsoubai, G. Stringhini, T. Solorio, M. De Choudhury, and P. J. Wisniewski, "A human-centered systematic literature review of the computational approaches for online sexual risk detection," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW2, pp. 1–38, Oct. 2021, doi: 10.1145/3479609.

[58] F. Elsafoury, S. Katsigiannis, Z. Pervez, and N. Ramzan, "When the timeline meets the pipeline: A survey on automated cyberbullying detection," *IEEE Access*, vol. 9, pp. 103541–103563, 2021, doi: 10.1109/ACCESS.2021.3098979.

[59] M. Mladenovic, V. Osmjanski, and S. V. Stankovic, "Cyber-aggression, cyberbullying and cyber grooming: A survey and research challenges," *ACM Comput. Surv.*, vol. 54, no. 1, pp. 1–42, Jan. 2022, doi: 10.1145/3424246.

[60] E. Staksrud and S. Livingstone, "Children and online risk: Powerless victims or resourceful participants?" *Inf., Commun. Soc.*, vol. 12, no. 3, pp. 364–387, Apr. 2009, doi: 10.1080/13691180802635455.

[61] M. Stoilova, S. Livingstone, R. K. Stoilova, and M., Livingstone, "Investigating risks and opportunities for children in a digital world: A rapid review of the evidence on children's internet use and outcomes," UNICEF, Innocenti, Florence, Itlay, Innocenti Discuss. Paper 2020-03, Feb. 2021.

[62] P. A. Chan and T. Rabinowitz, "A cross-sectional analysis of video games and attention deficit hyperactivity disorder symptoms in adolescents," *Ann. Gen. Psychiatry*, vol. 5, no. 1, pp. 1–10, Oct. 2006, doi: 10.1186/1744-859X-5-16.

[63] M. B. Mathur and T. J. VanderWeele, "Finding common ground in meta-analysis wars on violent video games," *Perspect. Psychol. Sci.*, vol. 14, no. 4, pp. 705–708, Jun. 2019, doi: 10.1177/1745691619850104.

[64] Z. Hussain and M. D. Griffiths, "Excessive use of massively multiplayer online role-playing games: A pilot study," *Int. J. Mental Health Addiction*, vol. 7, no. 4, pp. 563–571, Feb. 2009, doi: 10.1007/s11469-009-9202-8.

[65] C. J. Ferguson, "Do angry birds make for angry children? A meta-analysis of video game influences on children's and Adolescents' aggression, mental health, prosocial behavior, and academic performance," *Perspect. Psychol. Sci.*, vol. 10, no. 5, pp. 646–666, Sep. 2015, doi: 10.1177/1745691615592234.

[66] S. Domingues-Montanari, "Clinical and psychological effects of excessive screen time on children," *J. Paediatrics Child Health*, vol. 53, no. 4, pp. 333–338, Feb. 2017, doi: 10.1111/jpc.13462.

[67] D. Richards, P. H. Caldwell, and H. Go, "Impact of social media on the health of children and young people," *J. Paediatrics Child Health*, vol. 51, no. 12, pp. 1152–1157, Nov. 2015, doi: 10.1111/jpc.13023.

[68] V. Bell, "Online information, extreme communities and internet therapy: Is the internet good for our mental health?" *J. Mental Health*, vol. 16, no. 4, pp. 445–457, Jul. 2009, doi: 10.1080/09638230701482378.

[69] V. Suchert, R. Hanewinkel, and B. Isensee, "Sedentary behavior and indicators of mental health in school-aged children and adolescents: A systematic review," *Preventive Med.*, vol. 76, pp. 48–57, Jul. 2015, doi: 10.1016/j.ypmed.2015.03.026.

[70] R. Armitage, "Bullying in children: Impact on child health," *BMJ Paediatrics*, vol. 5, no. 1, Mar. 2021, Art. no. e000939, doi: 10.1136/bmjpo-2020-000939.

[71] M. V. Geel, P. Vedder, and J. Tanilon, "Relationship between peer victimization, cyberbullying, and suicide in children and adolescents: A meta-analysis," *JAMA Pediatrics*, vol. 168, no. 5, pp. 435–442, May 2014, doi: 10.1001/jamapediatrics.2013.4143.

[72] The New York Times. *Video Games Online Chats are Hunting Grounds for Sexual Predators*. Accessed: May 9, 2022. [Online]. Available: https://www.nytimes.com/interactive/2019/12/07/us/video-games-child-sex-abuse.html

[73] J. Campbell and J. Kravarik. *A 17-Year-Old Boy Died by Suicide Hours After Being Scammed. The FBI Says It's Part a Troubling Increase Sextortion Cases*. Accessed: May 9, 2022. [Online]. Available: https://edition.cnn.com/2022/05/20/us/ryan-last-suicide-sextortion-california/index.html

[74] M. K. Khan, O. Bamasag, A. A. Algarni, and M. Alqarni, "Policy brief: Fostering a safer cyberspace for children," Think 20 Engagement Group, ON, Canada, Tech. Rep., Dec. 2020.

[75] X. Zhang, "Charging children with child pornography–using the legal system to handle the problem of sexting," *Comput. Law Secur. Rev.*, vol. 26, no. 3, pp. 251–259, May 2010, doi: 10.1016/j.clsr.2010.03.005.

[76] C. Doyle, E. Douglas, and G. O'Reilly, "The outcomes of sexting for children and adolescents: A systematic review of the literature," *J. Adolescence*, vol. 92, pp. 86–113, Oct. 2021.

[77] Code of Federal Regulations. *Part 312–Children's Online Privacy Protection Rule*. Accessed: Jun. 9, 2022. [Online]. Available: https://www.ecfr.gov/current/title-16/part-312

[78] International Telecommunication Union. *Guidelines for Children*. Accessed: May 9, 2022. [Online]. Available: https://www.itu-cop-guidelines.com/children
[79] *Guidelines for Parents and Educators on Child Online Protection 2020*, International Telecommunication Union (ITU), Place des Nations, Geneva, Switzerland, 2020.
[80] *Guidelines for Industry on Child Online Protection 2020*, International Telecommunication Union (ITU), Place des Nations, Geneva, Switzerland, 2020.
[81] European Commission. *Self-Regulation for a Better Internet for Kids*. Accessed: Jun. 9, 2022. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/self-regulation-better-internet-kids
[82] *Guidelines for Policy-Makers on Child Online Protection 2020*, International Telecommunication Union (ITU), Place des Nations, Geneva, Switzerland, 2020.
[83] W. M. Taibah, H. K. Khalifa, and A. M. Alshebaiki, "Strengthening the convention on the rights of the child (CRC) governing children's digital world," Think 20 Engagement Group, ON, Canada, Tech. Rep., Dec. 2020.
[84] International Age Rating Coalition. (Apr. 4, 2022). *How IARC Works*. [Online]. Available: https://www.globalratings.com/how-iarc-works.aspx
[85] Australian Classification. *Helping You Choose What to Watch Play*. (Accessed: Apr. 5, 2022). [Online]. Available: https://www.classification.gov.au/
[86] Ministry of Justice and Public Security. *What is Rating System?*. Accessed: Apr. 5, 2022. [Online]. Available: https://www.gov.br/mj/pt-br/assuntos/seus-direitos/classificacao-1
[87] Gaming Rating & Administration Committee. *Age Rating Symbol*. Accessed: Apr. 6, 2022. [Online]. Available: https://www.grac.or.kr/english/
[88] Entertainment Software Rating Board. *Tools for Parents*. Accessed: Apr. 6, 2022. [Online]. Available: https://www.esrb.org/
[89] Pan European Game Information. *PEGI Helps Parents to Make Informed Decision When Buying Video Games*. Accessed: Apr. 6, 2022. [Online]. Available: https://pegi.info/
[90] Unterhaltungssoftware Selbstkontrolle. *Classification Games Apps*. Accessed: Apr. 8, 2022. [Online]. Available: https://usk.de/
[91] *2021 Essential Facts About Video Game Industry*, Entertainment Software Association, Washington, DC, USA, Jul. 2021.
[92] *For Parents*. Accessed: Jan. 3, 2022. [Online]. Available: https://corp.roblox.com/parents/
[93] *Family View*. Accessed: Jan. 3, 2022. [Online]. Available: https://help.steampowered.com/en/
[94] *How to Set Parental Controls on PS4 Consoles*. Accessed: Jan. 10, 2022. [Online]. Available: https://www.playstation.com/en-ae/support/account/ps4-parental-controls-and-spending-limits/
[95] *Gaming That is Safer for All*. Accessed: Jan. 10, 2022. [Online]. Available: https://www.xbox.com/en-U.S./community/for-everyone/responsible-gaming
[96] *Nintendo Switch Parental Controls Mobile App*. Accessed: Jan. 10, 2022. [Online]. Available: https://www.nintendo.com/switch/parental-controls/
[97] UNICEF. *Case Study: Millicom's Impact Assessment*. Accessed: Jun. 5, 2022. [Online]. Available: https://sites.unicef.org/csr/files/MILLICOM_casestudy.pdf
[98] UNICEF. *Case Study: Microsoft's Photo DNA*. Accessed: Jun. 5, 2022. [Online]. Available: https://sites.unicef.org/csr/files/MICROSOFT_casestudy.pdf
[99] UNICEF. *Case Study: Supporting National Governments to Develop Child Online Protection-Related National Action Plans*. Accessed: Jun. 5, 2022. [Online]. Available: https://sites.unicef.org/csr/files/Case_study_Microsoft.pdf
[100] UNICEF. *Case Study: Safaricom: Integrating Children's Rights into Core Business*. Accessed: Jun. 5, 2022. [Online]. Available: https://sites.unicef.org/csr/files/Case_study_Safaricom.pdf
[101] UNICEF. *Case Study: Amigos Conectados Project by the Walt Disney Company Latin America and chicos.net*. Accessed: Jun. 5, 2022. [Online]. Available: https://sites.unicef.org/csr/files/ase_study_Disney.pdf
[102] UNICEF. *Case Study: Strengthening Technology Companies Practices' to Fight Child Sexual Exploitation on Their Platforms*. Accessed: Jun. 5, 2022. [Online]. Available: https://sites.unicef.org/csr/files/Case_study_Thorn_Sound_Practices_GuideC.pdf
[103] UNICEF. *Case Study: LEGO Supplier Training Through the LEGO Academy in India*. Accessed: Jun. 5, 2022. [Online]. Available: https://sites.unicef.org/csr/files/LEGO_supplier_training.pdf
[104] *Deep Blue*. Accessed: Apr. 9, 2022. [Online]. Available: https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/
[105] C. Sang-Hun. *Google's Computer Program Beats Lee Se-Dol in Go Tournament*. Accessed: Apr. 9, 2022. [Online]. Available: http://www.nytimes.com/2016/03/16/world/asia/korea-alphago-vs-lee-sedol-go.html?_r=0
[106] S. H. Siddiqi, V. K. Shukla, A. B. Bhardwaj, and D. Gaur, "Analyzing psychological gamers profile through progressive gaming and artificial intelligence," in *Proc. 9th Int. Conf. Rel., Infocom Technol. Optim. Trends Future Directions (ICRITO)*, Noida, India, Sep. 2021, pp. 1–5, doi: 10.1109/ICRITO51393.2021.9596185.
[107] S. Zhao, R. Wu, J. Tao, M. Qu, H. Li, and C. Fan, "Multi-source data multi-task learning for profiling players in online games," in *Proc. IEEE Conf. Games (CoG)*, Osaka, Japan, Aug. 2020, pp. 104–111, doi: 10.1109/CoG47356.2020.9231585.
[108] J. Tao, Y. Xiong, S. Zhao, Y. Xu, J. Lin, R. Wu, and C. Fan, "XAI-driven explainable multi-view game cheating detection," in *Proc. IEEE Conf. Games (CoG)*, Osaka, Japan, Aug. 2020, pp. 144–151, doi: 10.1109/CoG47356.2020.9231843.
[109] D. Perez-Liebana, J. Liu, A. Khalifa, R. D. Gaina, J. Togelius, and S. M. Lucas, "General video game AI: A multitrack framework for evaluating agents, games, and content generation algorithms," *IEEE Trans. Games*, vol. 11, no. 3, pp. 195–214, Sep. 2019, doi: 10.1109/TG.2019.2901021.
[110] N. Oco, L. R. Syliongka, T. Allman, and R. E. Roxas, "Resources for Philippine languages: Collection, annotation, and modeling," in *Proc. 30th Pacific Asia Conf. Lang., Inf. Comput. (PACLIC)*, Seoul, South Korea, Oct. 2016, pp. 131–138.
[111] *Internet Relay Chat*. Accessed: Jan. 5, 2022. [Online]. Available: http://netsplit.de
[112] *Kick Ass Open Web Technologies IRC Logs*. Accessed: Jan. 5, 2022. [Online]. Available: http://krijnhoetmer.nl/irc-logs/
[113] *Inportb.com*. Accessed: Jan. 5, 2022. [Online]. Available: http://inportb.com
[114] *Literotica.com*. Accessed: May 5, 2022. [Online]. Available: http://literotica.com
[115] F. Rangel, P. Rosso, M. Koppel, E. Stamatatos, and G. Inches, "Overview of the author profiling task at PAN 2013," in *Proc. CLEF Eval. Labs Workshop Work. Notes Papers*, Valencia, Spain, 2013, pp. 352–365.
[116] *Myspace.com*. Accessed: Jan. 5, 2022. [Online]. Available: https://myspace.com/
[117] *Abusive Sexual Conversations Between Adults*. Accessed: May 9, 2022. [Online]. Available: https://www.fugly.com/victims/
[118] C. M. Arata, J. Langhinrichsen-Rohling, D. Bowers, and L. O'Farrill-Swails, "Single versus multi-type maltreatment," *J. Aggression, Maltreatment Trauma*, vol. 11, no. 4, pp. 29–52, Aug. 2005, doi: 10.1300/J146v11n04_02.
[119] *Net Alert*. Accessed: Jan. 5, 2022. [Online]. Available: https://netalert.me/
[120] *Net Nanny*. Accessed: Jan. 5, 2022. [Online]. Available: https://www.netnanny.com/
[121] *Bark*. Accessed: Jan. 5, 2022. [Online]. Available: https://www.bark.us/
[122] M. Pietikainen, "Recommendations for the online gaming industry on assessing impact on children gaming & the rights of the child," UNICEF, New York, NY, USA, Tech. Rep., Apr. 2020.
[123] M. A. Al-garadi, M. R. Hussain, N. Khan, G. Murtaza, H. F. Nweke, I. Ali, G. Mujtaba, H. Chiroma, H. A. Khattak, and A. Gani, "Predicting cyberbullying on social media in the big data era using machine learning algorithms: Review of literature and open challenges," *IEEE Access*, vol. 7, pp. 70701–70718, 2019.
[124] B. A. H. Murshed, J. Abawajy, S. Mallappa, M. A. N. Saif, and H. D. E. Al-Ariki, "DEA-RNN: A hybrid deep learning approach for cyberbullying detection in Twitter social media platform," *IEEE Access*, vol. 10, pp. 25857–25871, 2022, doi: 10.1109/ACCESS.2022.3153675.
[125] A. Jevremovic, M. Veinovic, M. Cabarkapa, M. Krstic, I. Chorbev, I. Dimitrovski, N. Garcia, N. Pombo, and M. Stojmenovic, "Keeping children safe online with limited resources: Analyzing what is seen and heard," *IEEE Access*, vol. 9, pp. 132723–132732, 2021, doi: 10.1109/ACCESS.2021.3114389.

[126] J.-I. Martinez-de-Morentin, A. Lareki, and J. Altuna, "Risks associated with posting content on the social media," *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, vol. 16, no. 1, pp. 77–83, Feb. 2021, doi: 10.1109/RITA.2021.3052655.
[127] M. Fire, R. Goldschmidt, and Y. Elovici, "Online social networks: Threats and solutions," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2019–2036, 4th Quart., 2014, doi: 10.1109/COMST.2014.2321628.

ANUM FARAZ (Member, IEEE) received the bachelor's degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2011. She is currently pursuing the master's degree with the Department of Electrical Engineering, Rochester Institute of Technology, Dubai, UAE. She has working experience in teaching topics related to digital signal processing, data algorithms, and communication systems. Her research interest includes applications of machine learning in communications systems.

JINANE MOUNSEF (Member, IEEE) received the Ph.D. degree in electrical engineering with a focus on signal processing and communications from Arizona State University, USA. She is currently an Assistant Professor with the Department of Electrical Engineering, Rochester Institute of Technology, Dubai, UAE. She has been in the research field for over more than ten years in machine learning, computer vision, and image processing, during which she worked with multidisciplinary teams on a long track of research articles. Her methodological and theoretical research as well as considerable portion of her applied and collaborative work address novel techniques in computer vision, in addition to designing and implementing smart systems. She is an Advisory Board Member at the New York Institute and Laboratory for Artificial Intelligence (NYILAI).

ALI RAZA (Member, IEEE) received the Ph.D. degree in telecommunications from the Queen Mary University of London, U.K. He is currently an Associate Professor with the Department of Computing Sciences, Rochester Institute of Technology, Dubai, UAE. He has been engaged in professional network and security consulting engagements with a number of government entities in the Middle East. These have included UAE, Qatar, Saudi Arabia, and Oman. His technical expertise has been welcomed in areas related to the design, deployment, and operations of secure infrastructure for mission-critical public safety and intelligence services. With the growing climate of geo-political instability in the region fueled by the sophisticated and covert means used by non-friendly intelligence groups, the security of such infrastructure is beyond the amount of investment. He has been active in the development of Chatbots for covert intelligence operations, since 2016 using platforms, such as AWS and GCP. Most of his work focuses on areas associated with Foreign Intelligence Agencies (FIAs) intent on cyber espionage through covert means. His research work in this area is under strict NDA.

SANDRA WILLIS received the Ph.D. degree in experimental social psychology from Tulane University, New Orleans, USA. She is a Psychologist by profession with 30 years of experience in academic, government, and non-government organizations engaging in teaching, mentoring, research, social policy, and social development activities leading to leadership roles that emphasize turnaround social development strategies. A versatile professional consistently achieving goals and moving from vision, research, policy, and strategy to implementation, she is adept at designing academic, social development, and capacity-building plans. She is the Director of Policy and Advocacy at the Global Mental Health Laboratory, Columbia University's Teachers College, and the Director of Learning and Development in World Enabled, an affiliate of the University of California at Berkeley.