J. Internet Comput. Serv.
ISSN 1598-0170 (Print) / ISSN 2287-1136 (Online)
http://www.jics.or.kr

An AI-Based Prevention Program to Protect Youth from Cybergrooming☆

김 기 정[1*] 리푸 후앙[2] 조 진 희[2]
Kee Jeong Kim Lifu Huang Jin-Hee Cho

ABSTRACT

The Digital Age calls for improvement of information literacy particularly among children and youth who are vulnerable to cybergrooming. Taking an interdisciplinary approach by leveraging our team's expertise including child and adolescent development, data analytics, and cybersecurity, this study proposes an interactive artificial intelligence (AI)-based preventive simulation program that raises youth knowledge and awareness about the risk of cybergrooming as well as increases resilient self-efficacy in their cybersecurity-relevant skills. The primary purpose of this project is to evaluate the effectiveness of the simulation program on preventing cybergrooming. More specifically, this study is designed to examine developmental changes in self-efficacy of cybersecurity-relevant skills among youth participants as a function of the preventive simulation program. Further, this study will identify risk and protective factors that explain interindividual differences in the ability of children and youth either to fall victim to advances from a cyber predator or to recognize and deter such threats. The preliminary data will help improve the effectiveness of the preventive simulation program as well as the methods of implementation to large groups of youth. The findings from the proposed study will contribute to making specific recommendations to parents, educators, practitioners, and policy makers for the prevention of cybergrooming.

☞ keyword : Artificial intelligence (AI)-based simulation, Cybergrooming, Cybersecurity, Children and Youth

1. Introduction

The Digital Age has enriched our lives in many exciting, prolific, and constructive ways. Access to an endless supply of knowledge and entertainment, remote education systems, online financial services, and social platforms to share opinions and ideas within a global community are only a few examples of the opportunities and benefits that we have received from living in the Digital Age. No opportunity comes without risk. It is estimated that one in ten children and youth growing up in the United States receive unwanted sexual solicitations through information and communication technologies [1]. The volume of online incidents reported to the National Center for Missing and Exploited Children is staggering. During the fall of 2020, the weekly average reported incidents of child pornography was more than 300,000 and over 500 incidents of online enticement of children for sexual acts were reported [2]. During the fall of 2020, the weekly average reported incidents of child pornography was more than 300,000 and over 500 incidents of online enticement of children for sexual acts were reported.

Traditionally, child grooming occurred when a predator acquired a physical proximity with children on playgrounds, at sports events, or other youth-oriented venues. In these situations, a risk for the predator to get caught was relatively high as someone could be suspicious of the predator's inappropriate yet intentional approaches to a victim—usually a young girl [3]. In the Digital Age, however, the cybergrooming process happens without predators revealing their true identity. These predators are skilled at concealing their identity and also at gradually progressing the intensity of the interaction with a victim. Behind the cloak of anonymity, the cyber predators approach multiple victims as if they were one of the victims' online friends struggling with academic work, family issues, peer pressure, and/or

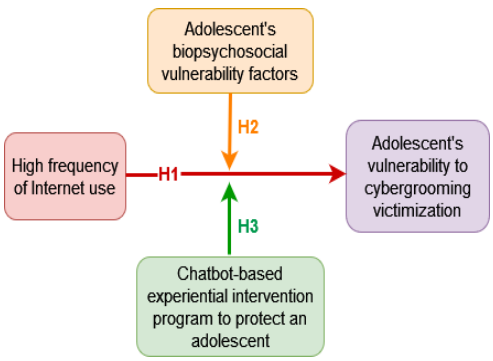
1 Department of Human Development and Family Science, Virginia Tech, Blacksburg Virginia 24061, USA
2 Department of Computer Science, Virginia Tech, Blacksburg Virginia 24061, USA
* Corresponding author (keekim@vt.edu)
[Received 10 April 2023, Reviewed 14 May 2023(R2 17 July 2023), Accepted 02 August 2023]
☆ The authors are grateful to the Institute of Creativity, Arts, and Technology at Virginia Tech for supporting this research project through its major SEAD research grant program.

emotional problems. It may take days, weeks, and even months for a cyber predator to build an emotional bonding and trust with a victim, which in turn, leads to the victim's desensitization to sexually explicit information and later direct exploitation.

The point of departure for this proposed transdisciplinary research collaboration is that the foundation of the cybergrooming process is established in the online chat rooms where youth are misled by a cyber predator to believe that they engage in a conversation with a peer. Parents are educated about protecting their children from toxic messages and doing their best to provide monitoring and supervision. Yet, it is clear that the Information Age calls for improvement of information literacy particularly among youth who are vulnerably exposed to cybergrooming.

2. The Primary Goal of Research

Combining our research team's expertise in child and adolescent development, data analytics, and cybersecurity, we plan to develop a preventive simulation program that raises children's knowledge and awareness about the danger of cybergrooming and also increases self-efficacy in their cybersecurity-relevant skills. The program is called *Chat-SARA* (see the next Methods section below for more information) that will be developed by leveraging Artificial Intelligence (AI) technologies. The simulation will be achieved by developing a chatbot based on a natural language processing, namely NLP [4] and game theoretical threat approaches [5, 6]. as the main AI techniques used.



(Figure 1) A Conceptual Model

It will safely but effectively inform youth about how to respond to advances from a cyber predator. The primary goal of this research is to present a conceptual model evaluating the effectiveness of *Chat-SARA* on preventing cybergrooming (Figure 1).

- To achieve the goal, the following two specific aims are developed:
- To determine developmental changes in self-efficacy in cybersecurity-relevant skills among participants as a function of the preventive simulation program.
- To identify risk and protective factors that impact the ability of a youth participant either to fall victim to advances from a cyber predator or to recognize and deter such threats.

The following hypotheses are derived from the conceptual model:

H1: Adolescents' frequent Internet use will directly affect increasing their vulnerability to cybergrooming.

H2: Adolescents' vulnerability to cybergrooming will substantially increase if they have risk factors, such as sensation seeking and impulsive traits, low self-esteem, and adjustment problems at home and school.

H3: Vulnerability to cybergrooming will not be observed among adolescents whose resilience is enhanced through the proposed experiential intervention program.

3. Methods

3.1. Development and Evaluation of ChatBot-basedResilienceandRiskIdentificationtoCybergrooming(Chat-SARA)

This research project begins with developing a natural language processing (NLP)-based chatbot that can identify and measure the levels of resilience and risk in a participant's response to cybergrooming. We will name the chatbot Chat-SARA, representing the ChatBot-based resilience and risk identification to cybergrooming. The

Chat-SARA will be developed with real-time chatting capability with youth participants where it is trained with the language and behaviors of cyber perpetrators (Figure 2).

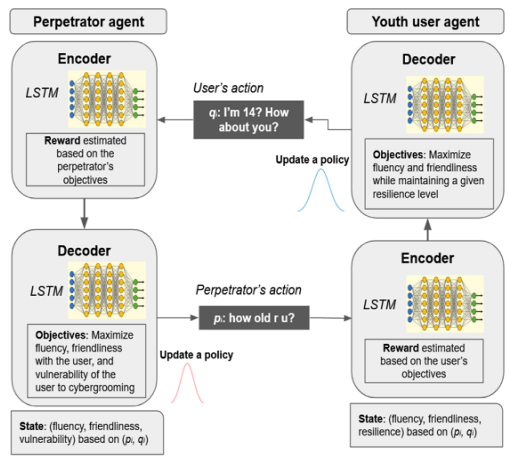
To develop a baseline model and to train the Chat-SARA (playing the role of a perpetrator), we propose to use the linguistic inquiry and word count (LIWC) framework [8, 9] to capture the key psycho-social features of potential victims by analyzing publicly available datasets including Perverted Justice datasets [10] and PAN12 datasets [11] for victim youth and chatting datasets for normal youth [12]. Note that Perverted Justice datasets [11] include chatting texts between perpetrators (450 perpetrators) and victims (309 victims) and PAN12 datasets (140 perpetrators) [12] includes chatting texts of pedophiles. We have focused on identifying verbal communication cues and will explore their relationship to risk and resilience factors of victims of cybercrimes as part of this research project [13].

	Red Flag	Harm -less
They give you lots of attention and go out of their way to tell you how special you are.		
They avoid telling you their real age.		
They share a specific mutual interest.		
They leave nice comments on your social media images.		
They want to keep your interactions private or avoid talking in public forums or spaces.		
They tell you not to tell your friends or parents about your interactions.		
You have never met them in person, but they say they're a friend of a friend.		
They say things like "I know we were supposed to FaceTime today, but my camera is broken."		

(Figure 2) Youth participants must learn to identify and mitigate "red flag" versus harmless statements while using social media and interacting with potential cyber predators online [7].

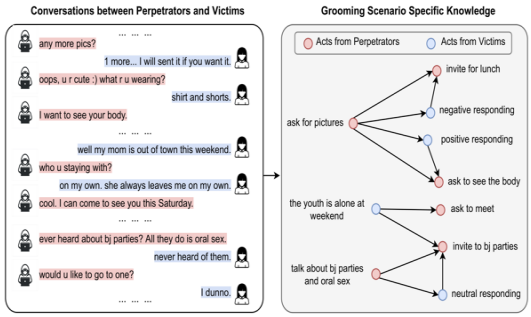
We speculate that cybergrooming process takes place by following a set of progressive stages. The first stage would be greetings and casual talks to establish a connection between an online perpetrator and a victim. The second stage is driven primarily by the perpetrator in his/her attempt to collect personal and identifiable information such as name, age, gender, location, interests, family, school, or daily routines. The third stage is to be centered on laying out sexual conversations and asking questions related to sexual behaviors. The last stage would be when a victimization has been established by the victim's sending sexual photos and video images.

The Chat-SARA is designed to progress to achieve its goal (i.e., reaching the last and fourth stage illustrated above) by maintaining its natural flow of conversational dialogues with a user. More specifically, a variation in a user agent (i.e., a potential victim)'s resistance to cybergrooming dialogues will be allowed in the chatbot to increase the adaptive capability of a perpetrator agent to handle a different type of users. The chatbot user's actions will be a set of utterances belonging to the user's resilience stage where we consider the three levels of resistance: low, medium, and high. This will allow the perpetrator agent to adaptively respond to a human user depending on the user's resistance level. The chatbot will consider the dialogue history that is transferred into a Long Short-Term Memory (LSTM) encoder system [14] (Figure 3).



(Figure 3) A Dialogue Mechanism between a Perpetrator and a Victim in Chat-SARA

One of the main objectives of this project is to identify verbal communication cues and also to explore their relationship to risk and protective factors of victims to cybercrimes. To determine the realism of the chatbot discussion with real human subjects, rigorous analyses such as a Turing test are expected to be conducted by human judges to compare a conversation between two humans and between *Chat-SARA* and human subjects and select the ones that is most likely to be human-to-human conversations [15, 16].



(Figure 4) Simulated Conversations with a Perpetrator and Grooming-in-the-Making

The texts that human participants use during the simulation session should be thoroughly analyzed. As depicted in Figure 4, the simulation dialogues between a victim and a cyber sex offender reveal that ongoing and reciprocal messages are characteristic of cybergrooming acts. For instance, if a victim responses to a request of sending pictures without resistance, the offender is highly likely to request additional private and sensitive information upon the receipt of the pictures. In contrast, if the victim changes to other subject matters without reciprocating the request of sending some pictures, the offender is also likely to be adaptive the conversion topic change and then ask if the victim is alone at home or suggest meeting on a weekend. This temporal order of conversations and cybergrooming act knowledge significantly benefits the chatbots to mimic the way the offenders approach to victims and develop the cybergrooming process between perpetrators and victims. However, the temporal order of cybergrooming act knowledge is likely to be covertly embedded in the unstructured conversations without any structured knowledge resources available.

3.2. Protection of Human Subjects

We will ensure that the research outcome from this kind of research work will not be used in a negative fashion. For example, the identified vulnerability factors to cybergrooming should not be accessed by real perpetrators who can leverage the research findings to attack potential victims effectively. Therefore, we will selectively make our research outcome and datasets available to avoid any malicious application of this kind of research.

We will do our due diligence to protect the youth research participants' anonymity and identity from cybersecurity threat by executing the following action plans: (a) Have educational parties can use the outcome and findings of this kind of research to develop curricula to educate adolescents about how to respond to abusive online messages and cope with cybergrooming, (b) Share datasets (privacy-anonymized) selectively with research groups wanting to use it to protect adolescents from online risks, (c) Let parents use the outcomes and findings of this study to learn grooming conversations and use them to educate their children to build resistance and resilience against the potential risk of cybergrooming, (d) Mitigate any possible misuse of sensitive or inappropriate languages used by this kind of research and eliminate them using linguistic resources, such as the profane lexicons [17] to replace offensive/profane words in the real datasets with moderate ones when the cybergrooming situations are described to human subjects to avoid any potential ethical issues, (e) Make our source code and analysis model accessible to parties providing a clear research goal and identity, (f) Store securely all the conversational data collected should be securely stored under the regulations and standards stated in the legal frameworks, such as the General Data Protection Regulation [18], and (g) Follow best practices established based on the prior related works, such as [19, 20, 21] and Institutional Review Boards (IRB) guidelines when handling potentially privacy-invasive, illegal, risky, and/or otherwise sensitive topics among adolescents, especially for all human subjects research, particularly research with minors.

4. Conclusions

Children and teenagers are heavily engaged in online activities with the Internet access being widespread. According to a study by the Pew Research Center, approximately 95% of teenagers (ages 13-17) in the US have access to a smartphone and 45% of them say they are online almost constantly [22]. Furthermore, the COVID-19 pandemic had a significant impact on the increase in teenagers' internet use for remote learning, social connection, information and news consumption, and mental health and support. Increased Internet use among teenagers, especially those who seek social connection and support, might have made them more vulnerable to cybergrooming due to various personal, familial, and social factors.

The current study aims to develop an effective prevention program that empowers children and teenagers in the *Information Age* by helping raise their knowledge and awareness about the dangers and risks of cybergrooming and also increase self-efficacy in their cybersecurity relevant skills. To understand the roles of risk and resilient factors in increasing susceptibility to victimization as well as interacting with the proposed simulation program's effectiveness, a survey must be administered to children and adolescents. The evaluation design should include a pre- and post-test to measure whether the expected changes take place in the human subjects' understanding and awareness of the victimization processes [23].

It should be noted that a survey study examining the effectiveness of *Chat-SARA* on preventing cybergrooming involves children who are defined as "persons who have not attained the legal age for consent to treatments or procedures involved in the research, under the applicable law of the jurisdiction in which the research will be conducted" [24]. Additional protections for children participating in human subjects research should be provided. Having human subjects participate in experiments online raises new ethical concerns. Our research team will do our due diligence to protect the children's anonymity and identity from cybersecurity threat by ensuring electronic data security [25]. A successful development of the AI-based preventive program and execution of it with a community-based representative group of children and adolescents is highly likely to suggest a new proactive model for protecting children from online predators. Results from the developmental assessment of the effectiveness of the preventive program will offer parents, educators, practitioners, and policy makers insights on how to help children navigate the Internet safely.

References

[1] L.K. Jones, K. Mitchell, and D. Finkelhor, "Trends in youth internet victimization: Findings from three youth internet safety surveys 2000-2010, *Journal of adolescent health*, vol 50, no.2, pp. 179-186, 2012. https://www.unh.edu/ccrc/sites/default/files/media/2022-03/trends-in-youth-internet-victimization.pdf

[2] National Center for Missing & Exploited Children, 2021. Retrieved from https://www.missingkids.org/HOME

[3] S. van der Hof and B. J. Koops, "Adolescents and cybercrime: Navigating between freedom and control," *Policy & Internet*, vol. 3, pp. 1-28, 2011. https://doi.org/10.2202/1944-2866.1121

[4] E. N. Forsyth and C. H. Martell, "Lexical and discourse analysis of online chat dialog" *Proc. of the First IEEE International Conference on Semantic Computing (ICSC)*, pp. 19-26, 2007. https://ieeexplore.ieee.org/document/4338328/;jsessionid=aHb_zqkvMXbwf4BRQFbHbsqfsvlUKUcZy_uVyvytKNYkL7hs81qd!-1543151615

[5] C. Laorden, D. Carlos, P. Galán-García et al., "Negobot: A conversational agent based on game theory for the detection of pedophilic behavior," *Proc. of International Joint Conference CISIS'12-ICEUTE' 12-SOCO' 12 Special Sessions,* 2013. https://doi.org/10.1007/978-3-642-33018-6_27

[6] P. Zambrano, J. Torres, L. Tello et al., "Technical mapping of the grooming anatomy using machine learning paradigms: An information security approach," *IEEE Access*, vol 7, pp. 142129-142146, 2019.

[7] CYBER S.W.A.T., "Predatory behavior online. SRO Learning module," *National White Collar Crime Center and the Safe Surfin' Foundation*, 2019. Retrieved from

https://www.teamcyberswat.org/students/

[8] LIWC2015., "Linguistic inquiry and word count (LIWC)," *Pennebaker Conglomerates*, 2020. Retrieved from https://liwc.wpengine.com/

[9] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of Language and Social Psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[10] Perverted Justice Foundation Inc., "Perverted Justice," 2021. Retrieved from http://www.perverted-justice.com/?archive=byUserVotes

[11] I. Giacomo and C. Fabio, "*PAN12 Deception Detection: Sexual Predator Identification [Data set]*," CLEF 2012 Labs and Workshops, Notebook Papers. Zenodo, 2012. http://doi.org/10.5281/zenodo.3713280

[12] A. S. Masten, "Risk and resilience in development," *The Oxford handbook of developmental psychology*, vol. 2. pp. 579–607, 2011.

[13] P. Zambrano, J. Torres, L. Tello-Oquendo, J. Ruben, *et al.*, "Technical mapping of the grooming anatomy using machine learning paradigms: An information security approach," *IEEE Access*, vol. 7, pp. 142129-142146. 2019. https://10.1109/ACCESS.2019.2942805

[14] H. Cuáyahuitl, D. Lee, S. Ryu, S. Choi, I. Hwang, and J. Kim, "Deep reinforcement learning for chatbots using clustered actions and human-likeness rewards," , pp.1-8, 2019. htpps://10.1109/IJCNN.2019.8852376

[15] N. M. Radziwill and M.C. Benton, "Evaluating quality of chatbots and intelligent conversational agents," *Computers and Society*, 2017 https://doi.org/10.48550/arXiv.1704.04579

[16] A. M. Turing, "Computing Machinery and Intelligence," *Parsing the Turing Test*. Springer, Dordrecht. 2009.

[17] L. von Ahn. *Useful resources from Luis von Ahn's Research Group*. 2021. Retrieved from https: //www.cs.cmu.edu/~biglou/resources/

[18] Intersoft Consulting, "General data protection regulation," 2020. Retrieved from https://gdprinfo.eu/

[19] K. Badillo-Urquiola, Z. Shea, Z. Agha, I. Lediaeva, and P. Wisniewski, "Conducting risky research with teens: Co-designing for the ethical treatment and protection of adolescents," *Proc. ACM Hum.-Comput. Interact*, vol. 4, pp. 46-49, 2021. https://doi.org/10.1145/3432930

[20] N. McDonald, K. Badillo-Urquiola, M. G. Ames, N. Dell, E. Keneski, M. Sleeper, and P. Wisniewski, "Privacy and power: Acknowledging the importance of privacy research and design for vulnerable populations," *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1-8, 2020. https://doi.org/10.1145/3334480.337517

[21] A. M. Walker, Y. Yao, C. Geeng, R. Hoyle, and P. Wisniewski, "Moving beyond 'One Size Fits All': Research considerations for working with vulnerable populations. *Interactions*, vol. 26, no. 6, pp. 34–39, 2019. https://doi.org/10.1145/3358904

[22] B. Auxier and M. Anderson, "Social media use in 2021," *Pew Research Center*, pp. 1-18, 2021.

[23] G. Alessandri, A. Zuffianò, and E. Perinell, "Evaluating intervention programs with a Pretest-Posttest Design: A structural equation modeling approach," *Frontiers in Psychology,* vol 2, no. 8, pp. 223-227, 2017.

[24] Code of Federal Regulations, 45 C.F.R. § 46. *U.S. Department of Health and Human Services*, 2018.

[25] G. Kelly and B. McKenzie, "Security, privacy, and confidentiality issues on the Internet," *Journal of Medical Internet Research,* Vol 4, No. 2, 2002.

◐ 저 자 소 개 ◑

김 기 정(Kee Jeong Kim)
1996년 아이오와 주립대학교 인간발달학과 (석사)
1998년 아이오와 주립대학교 인간발달학과 (박사)
1999년~2001년 아이오와 주립대학교 포스트닥터
2002년~2004년 캘리포니아 주립대학교-데이비스 인간발달학과 연구과학자
2004년~현재 버지니아 공대 인간발달학과 교수
관심분야 : 빅데이터 통계분석, 아동과 청소년의 행동 분석 및 미래 행동 예측, 부모와 자녀 관계 분석
E-mail : keekim@vt.edu

리푸 후앙 (Lifu Huang)
2014년 베이징 대학교 컴퓨터공학과 (석사)
2019년 일리노이 주립대학교-어바나 샴페인 컴퓨터공학과 (박사)
2020년~현재 버지니아 공대 컴퓨터공학과 교수
관심분야 :데이터 분석, 머신 러닝, 자연어 처리
E-mail : lifuh@vt.edu

조 진 희 (Jin-Hee Cho)
2004년 버지니아 공대 컴퓨터공학과 (석사)
2008년 버지니아 공대 컴퓨터공학과 (박사)
2020년~현재 버지니아 공대 컴퓨터공학과 교수
관심분야 : 데이터 분석, 머신 러닝, 보안
E-mail : jicho@vt.edu