Ethical Considerations for Adolescent Online Risk Detection AI Systems

Afsaneh Razi
University of Central Florida, USA
afsaneh.razi@knight.ucf.edu

Seunghyun Kim
Georgia Tech, USA
skim888@gatech.edu

Munmun De Choudhury
Georgia Tech, USA
munmund@gatech.edu

Pamela Wisniewski
University of Central Florida, USA
pamwis@ucf.edu

ABSTRACT
We seek to develop future Artificial Intelligent (AI) risk detection algorithms to address keeping adolescents safe by providing accurate and customized services to teens and their parents. Training such accurate algorithms needs data of minor which raise ethical challenges. Also, the proper use of such systems is another issue. In this workshop, we hope to gain more insights about possible approaches to address these ethical challenges.

Author Keywords
Ethical AI; Online risk detection; Adolescent Online Safety.

CSS Concepts
• Human-centered computing~Human computer interaction (HCI)

GRAND ETHICAL CHALLENGES FOR AI ADOLESCENT ONLINE SAFETY
Internet and social media use is increasingly deeply intertwined in teens' lives [11]. Although it provides a great opportunity for teens to learn, it also exposes teens to various online risks [6]. Research has shown that solutions for adolescent online safety are relying more on parental control through device-based restrictions and direct monitoring [3]. The current approaches overwhelm parents with teens' information which they do not find it useful, and also are privacy-invasive to teens [3, 4]. The reason is current risk detection algorithms do not take the context of online interactions into consideration (e.g. teens using curse words when joking in a group chat is different than using such words directly targeted to someone). Thus, machine learning risk detection algorithms should be optimized to the actual content that parents and teens find it risky. In the future, these optimized systems with the use of AI will help adolescent and their parent have a safe online experience. We have an NSF funded project [7] to improve automated risk detection algorithms using human-centered principles. We plan to commercialize the final solution for an easy-to-use and accessible service for adolescent online risk detection.

Using Teen Data to Create ML Algorithms
In our project, we are collecting social media data of teens which includes both their private and public data. Since training with representative data is a big part of making effective classifiers, we need a contextual and in-depth knowledge of adolescent risk behaviors [9, 10]. Current risk detection systems are not taking into consideration the nuance in risk classification when it comes to risky social media content of teens [1, 5]. So, we are using human-centered approaches and qualitative analysis to develop risks concepts. We aim to label teens social media data for ground truth before developing advanced and automated algorithms for risk detection.

However, the data collected from teens might include sensitive and possibly illegal artifacts, such as sexually explicit images that could be classified as child pornography [8]. These sensitive data pose serious ethical issues for the human-centered approach to building AI risk detection systems. Thus, we need to ensure that our data collection and algorithmic analysis of teen social media data is ethical. We have considered some ways to address these ethical issues. In collecting teens data, not only we obtain parental consent, but also, we get teen assent. We believe it gives teens more sense of authority and control over their data. There is also an intrinsic challenge in the process of collecting the data, since asking the teens to self-identify harassments might result in a recap of those experiences. We make sure to not inflict a time constraint to the participants when collecting data to not possibly overwhelm them with such experiences. For preserving the privacy of the users, we make sure to not publish any of the personally identifiable information or any quotes of their messages that can be retrieved through online search. Also, we convert usernames to randomly generated IDs to protect privacy of teens. Ethics is essential when developing data sharing policies, terms of appropriate use, and licensing agreements for the solutions that result from this project.

Deploying Adolescent Online Risk Detection Algorithms for Good (Not Evil)
After the AI system is built, it is important to protect it from the wrong hands. For instance, if there exists a system that can detect minors sexually explicit images, some people can misuse the system and use it as a tool to find sexual images of youth to save it in a porn website or in darknet. Also, the data detected as risky by these systems can be hacked or misused. To combat against the potential reverse engineering of our trained adolescent online risk algorithms [2], the trained algorithms will not be released as open-source. Access privileges to the system should be monitored in order to ensure that only the correct people have the right type of access to the system. Thus, safeguards should be devised to protect the security and privacy of such system and more research should address these challenges.

CONCLUSION

We found approaches to address some of the ethical challenges for improving adolescent risk detection systems. As we want to make sure to maintain the confidentiality and privacy of the data and security of the AI system that we design. We hope to gain more insights from participating in the Good Systems: Ethical AI for CSCW to tackle challenges for designing AI systems that can promote online safety for teens.

ACKNOWLEDGMENTS

This research is supported by the U.S. National Science Foundation under grant number #IIP-1827700. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the research sponsors.

REFERENCES

[1] Adnan, A. and Nawaz, M. 2016. RGB and Hue Color in Pornography Detection. *Information Technology: New Generations* (2016), 1041–1050.

[2] Barreno, M. et al. 2006. Can Machine Learning Be Secure? *Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security* (New York, NY, USA, 2006), 16–25.

[3] Ghosh, A.K. et al. 2018. A Matter of Control or Safety?: Examining Parental Use of Technical Monitoring Apps on Teens' Mobile Devices. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), 194:1–194:14.

[4] Ghosh, A.K. et al. 2018. Safety vs. Surveillance: What Children Have to Say about Mobile Apps for Parental Control. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (Montreal QC, Canada, 2018), 1–14.

[5] Hosseinmardi, H. et al. 2015. Analyzing Labeled Cyberbullying Incidents on the Instagram Social Network. *Social Informatics* (Dec. 2015), 49–66.

[6] Livingstone, S. and Helsper, E. 2010. Balancing opportunities and risks in teenagers' use of the internet: the role of online skills and internet self-efficacy. *New Media & Society*. 12, 2 (Mar. 2010), 309–329. DOI:https://doi.org/10.1177/1461444809342697.

[7] NSF Award Search: Award#1827700 - PFI-RP: A Multi-Disciplinary Approach to Detecting Adolescent Online Risks.: *https://nsf.gov/awardsearch/showAward?AWD_ID=1827700&HistoricalAwards=false*. Accessed: 2018-12-21.

[8] Poole, E.S. and Peyton, T. 2013. Interaction Design Research with Adolescents: Methodological Challenges and Best Practices. *Proceedings of the 12th International Conference on Interaction Design and Children* (New York, NY, USA, 2013), 211–217.

[9] Wisniewski, P. et al. 2016. Dear Diary: Teens Reflect on Their Weekly Online Risk Experiences. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2016), 3919–3930.

[10] Wisniewski, P. et al. 2015. Resilience Mitigates the Negative Effects of Adolescent Internet Addiction and Online Risk Exposure. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (New York, NY, USA, 2015), 4029–4038.

[11] 2018. Teens, Social Media & Technology 2018 | Pew Research Center http://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/.