Co-Designing User Personas and Risk Scenarios for Evaluating Adolescent Online Safety Interventions

Zainab Agha
Vanderbilt University
Nashville, USA
zainab.agha@vanderbilt.edu

Kelsey Miu
Vanderbilt University
Nashville, USA
kelsey.miu@vanderbilt.edu

Sophia Piper
Vanderbilt University
Nashville, USA
maria.s.piper@vanderbilt.edu

Jinkyung Park
Vanderbilt University
Nashville, USA
jinkyung.park@vanderbilt.edu

Pamela J. Wisniewski
Vanderbilt University
Nashville, USA
pamela.wisniewski@vanderbilt.edu

ABSTRACT

Adolescent online safety research has largely focused on designing interventions for teens, with few evaluations that provide effective online safety solutions. It is challenging to evaluate such solutions without simulating an environment that mimics teens online risks. To overcome this gap, we conducted focus groups with 14 teens to co-design realistic online risk scenarios and their associated user personas, which can be implemented for an ecologically valid evaluation of interventions. We found that teens considered the characteristics of the risky user to be important and designed personas to have traits that align with the risk type, were more believable and authentic, and attracted teens through materialistic content. Teens also redesigned the risky scenarios to be subtle in information breaching, harsher in cyberbullying, and convincing in tricking the teen. Overall, this work provides an in-depth understanding of the types of bad actors and risky scenarios teens design for realistic research experimentation.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

Adolescent online safety, teens, co-design, user persona, social media, simulation, nudges, interventions

ACM Reference Format:
Zainab Agha, Kelsey Miu, Sophia Piper, Jinkyung Park, and Pamela J. Wisniewski. 2023. Co-Designing User Personas and Risk Scenarios for Evaluating Adolescent Online Safety Interventions. In *Computer Supported Cooperative Work and Social Computing (CSCW '23 Companion), October 14–18, 2023, Minneapolis, MN, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3584931.3606964

CSCW '23 Companion, October 14–18, 2023, Minneapolis, MN, USA

ACM ISBN 979-8-4007-0129-0/23/10…$15.00
https://doi.org/10.1145/3584931.3606964

1 INTRODUCTION

In 2022, 97% of U.S. teens are reported to be online daily, and 46% of them are online almost constantly [18]. Although this constant connectivity can be beneficial to teens, it can also expose them to risks online, such as online harassment, sexual solicitations, privacy breaches, and exposure to explicit content [3, 4, 14, 15, 17]. Recently, co-design research with youth has been successful in including teen voices and their unique perspectives to design resilience-based approaches for online risks [2, 5, 16]. For instance, in a recent co-design effort by Agha et al. [1] which involved User Experience (UX) bootcamps with teens, "nudges" have been proposed as a 'just-in-time' intervention to support teens in the moment when they experience risks online [2]. The study provided valuable insights into the types of nudges teens design for commonly faced online risks (e.g., information breaches, cyberbullying, sexual risks). Meanwhile, in order for intervention designs to be truly beneficial, there is a need to implement and evaluate these nudges, in a way that accurately depicts teens' responses to these nudges when faced with a risk. Yet, a majority of the prior work within the online safety space has focused on *designing* interventions [7, 10], with few realistic evaluations of nudges that assess their effectiveness for online safety. This further exemplifies the need to build upon design work and move towards evaluations that can provide us with a holistic understanding of the effectiveness of adolescent online safety nudges.

One way for evaluating nudges is through simulation-based evaluations that mimic the environment and risks to understand how nudges lead to behavior change. Such simulation-based evaluation has been explored as a promising approach for evaluating adolescent online safety nudges in an ecologically valid manner in prior studies [12, 20]. Prior researchers have emphasized the need for ensuring experimental realism [8, 13] by simulating authentic social media experiences. To do this, DiFranzo et al. [8] conducted survey-based pilot studies with participants using Amazon Mechanical Turk, where they asked participants to rate the risky scenarios for believability. While DiFranzo et al.'s work was with general populations over the age of 18, teens have unique developmental needs and social media experiences that require further investigation to ensure experimental realism and ecological validity. It is challenging to design risky scenarios and bad actors for simulating online risks for realistic evaluation of interventions that are relevant to teens without their involvement through co-design. Moreover, Walker et

al. [19] encouraged researchers to include vulnerable populations such as youth throughout the research process, including while designing a research study, to ensure that the research meets their needs. To address this gap, we conducted meta-research with teens to obtain their feedback on the design of user personas and risks scenarios which will be later implemented in a social media simulation, for evaluating adolescent online safety interventions. We asked the following research questions:

- **RQ1:** *How would teens design realistic risky users they encounter on social media?*
- **RQ2:** *How would teens design risky scenarios they face online?*

To answer these questions, we conducted remote focus groups with 14 teens between the ages of 13-18, based in the United States, who had access to video-calling capabilities. During these sessions, we presented teens with 10 prepared user personas and 4 risky scenarios. The risk scenarios were based on prior research conducted with teens [1]. Teens redesigned various aspects of the personas and scenarios using an online whiteboard tool, FigJam. Through this work, we provide insights into the types of users and risks teens regularly encounter online. We contribute to the CSCW adolescent online safety and co-design communities by crowd-sourcing the experiences of youth and involving them in the design of an ecologically valid simulation of their social media experiences. In the process, we highlight teens' perspectives of simulated risky user accounts and scenarios on social media, rooted in their personal experiences.

2 METHODS

2.1 Study Overview

We conducted six focus groups with 14 youths (ages 13-18) virtually via Zoom, with 2-3 teens in each focus group, to have feedback on and design ecologically valid user personas and risk scenarios to be implemented later within a social media simulation. We build upon an open-source social media simulation developed by DiFranzo et al. [8], which allows researchers to change variables, actors, and the social media simulation environment to suit their needs. These personas and risk scenarios were based on findings from prior work with teens [1, 2]. All personas and risk scenarios were presented to each teen to provide feedback through design activities conducted on a virtual collaborative whiteboard, FigJam [9]. Participants were prompted to give verbal feedback as well as annotate on the virtual whiteboard with their design ideas. At the end of each activity, the researchers summarized the ideas shared by teens. This study was approved by the authors' Institutional Review Board (IRB) and parental consent was required for participants under the age of 18.

Participants were mainly recruited from personal contacts of Vanderbilt University students, universities, and schools across the U.S., and existing contacts with youth-serving organizations in the U.S. These organizations were contacted via email, call, distributing and/or posting on social media. The session lasted for about 2-3 hours and participants were compensated with $20 Amazon gift cards for participation.

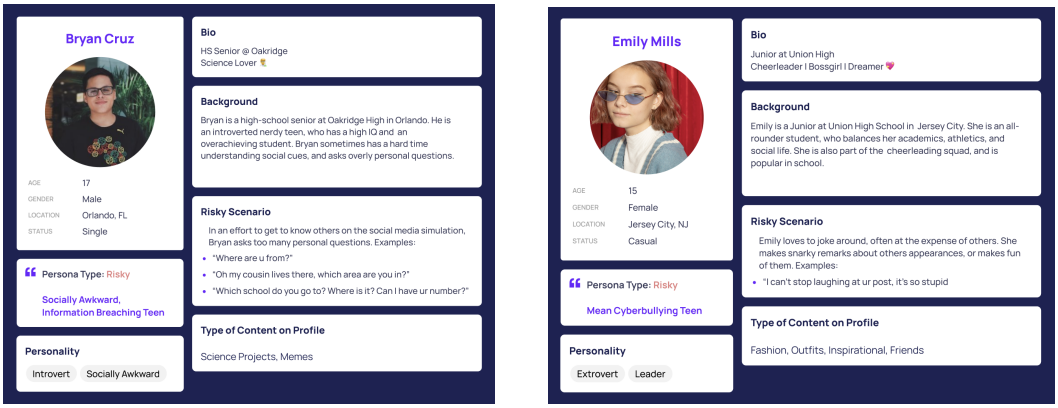
2.2 Design Prompt and Research Activities

2.2.1 User Personas and Risk Scenarios. We presented teens with 10 personas, which included 4 risky and 6 non-risky personas. The risky personas were based on prior online risks experiences shared by teens [1]. Broadly, the risky personas covered the following risk encounters, which are a combination of private and public risks; a) **Private Information Breaching** which focused on socially awkward introverted teen Bryan (Fig. 1a), who has a hard time understanding social cues, and asked overly personal questions, such as *"I don't see your location on your profile. Where do you live?"*, b) **Public Cyberbullying** which included Emily (Fig. 1b), a popular girl at school who loved to joke around, often at the expense of others and makes snarky remarks about others' posts publicly, such as *"I can't stop laughing at ur post, it's so stupid"*, c) **Private Predatory Messages** from Dave who was an adult who often sent inappropriate in private settings online, by first establishing trust with the teen and later sending creepy and predatory messages such as *"You look cute in that pic. I'd love to get to know u better.. Wanna Facetime?"*, d) **Private Spam & Explicit Content** from Kyle, who was a bot account that sent spam links to others with clickbait-y messages to entice teens to open the links, such as *"Check out this dope new game: www.gamez.com/nudepix"*

2.2.2 Research Activities. In groups of two or three teen participants, six remote focus group sessions were conducted over Zoom. The study included introductions of the researchers and participants, an introduction to adolescent online safety, after which nudges were then introduced with examples, and participants were asked a warm-up question on ways to evaluate interventions for implementation. Then, the social media simulation was explained, including user personas, interface design, and interventions. The participants were then asked to actively engage in a design activity focused on providing feedback on risk scenarios and user personas using an online interactive whiteboarding tool, FigJam. Each user persona included a user's age, location, relationship status, background, personality type, content on their profile, and their risky scenario including quotations. Teens were presented with the same personas and asked to choose at least one of the risky personas for detailed redesigning, while providing high-level feedback for the remaining personas. At the end of the session, all designed whiteboards were collected by downloading them from FigJam into a secured password-protected laptop. All sessions were video and audio recorded, and the recordings were fully transcribed by the researchers. After reviewing the recorded sessions, we conducted a preliminary thematic qualitative analysis by reviewing the dataand grouping recurring insights to identify major themes that emerge.

3 FINDINGS

We had an equal gender representation with 7 male (50%) and 7 female (50%) participants, with most participants between the ages of 16-17 (50%). Our participants identified themselves as White/Caucasian (7%), Black/African American (14%), Hispanic/Latino (21%), and Asian (57%). Below, we summarize our key takeaways regarding teens' perspectives of risky users and scenarios on social media.



(a) Private Information Breaching Persona **(b) Public Cyberbullying Persona**

Figure 1: Example Risky Personas

3.1 Teens designed risky user personas to have suitable personality traits, be more believable, care about their reputation, and entice teens through their content (RQ1)

Overall, teens were thoughtful about the user personas and how their personality plays a critical role in the type of interaction they would have. For instance, many teens wanted to change Bryan, the information-breaching persona's personality to match his actions more, as they considered that an introverted person would be less likely to ask such direct and invasive questions. Rather, teens considered that someone of an extroverted nature is more likely to perpetuate information breaching risks. A few teens were particularly critical of the awkward nature of Bryan's personality, as they thought that awkward individuals should not be confused with unsafe individuals. P1 explained, *"I don't believe that portraying the awkward character as maliciously harmful in a way comparable to bullying and spam is appropriate"* (P1, 18-year-old, Male).

Teens also considered whether the personas' background fit the environment in which they would perpetuate the risk. In this regard, teens were particularly critical of the cyberbullying persona, Emily, who portrayed a popular girl at school that cyberbullied others in public posts. Most teens thought that such popular people care about their reputation and would not cyberbully others publicly for fear of getting "canceled." Therefore, many teens redesigned this persona to bully in private settings or in person, while they pretended to be nice and supportive to others in public. P12 (13-year-old, Female) explained, *"I don't think I've seen popular people like say bad things on people's accounts online. But I have definitely seen them like bullying in person."* Few teens also designed this persona to be rich and show off their money, which implies that they look down on others and bully them about materialistic things.

Additionally, teens wanted the risky users to be believable, when it came to bot accounts. Teens found the spam bot persona to be too self-evident and obvious as it did not have a photo or bio and sent spam links that were clearly suspicious. With the increase of Artificial Intelligence, teens redesigned this persona to be believable and similar to regular social media accounts. Many of them redesigned the spam bot persona to be smarter and believable by including a photo, bio, and content on a profile. Some teens recommended adding more reposted content or memes on such spam bot accounts, as they often rely on existing content to populate their accounts.

Teens also questioned the Dave Fisher persona, who mimicked a 32-year-old doctor who sends creepy and predatory messages to teens in Direct Messages, as they found him to be unrealistic given his lifestyle and occupation. Teens thought that he would not fake his profession, and it seemed to them that a doctor would not have time for social media, making the persona suspicious. Additionally, based on their experiences with such users online, teens redesigned this persona to have more narcissistic traits, such as posting more photos of themselves and showing off money, cars, or materialistic things to attract teens. Being an adult, teens also imagined this persona to have a different texting style than teens, such as using too many emojis or not being familiar with slang.

3.2 Teens designed risky scenarios to be more subtle in information breaching, harsher in cyberbullying, and more convincing in tricking the teen (RQ2)

Subtle Information-Breaching Risk: Teens thought that the information-breaching risk was asking for information in very obvious and direct ways, and immediately jumped to asking about the teen's address. In contrast, in their experience, such risks were often perpetuated more ambiguously and happened over time. Therefore, many teens redesigned the information breaching persona (Bryan) to ask for personal information subtly, in less direct ways, for instance, revising the risk to, *"Hey, did you go to Oakridge, u look kinda familiar"* (P11, 18-year-old, Female). Moreover, some teens believed that such risk scenarios are often built up over time, where the risky user established rapport and shared context with the teen first, before asking for their personal information.

Harsher Cyberbullying: When we asked to redesign the cyberbullying scenario (Emily), most teens considered her remarks about a post being stupid to be too casual or did not consider it risky enough. Teens thought that making such remarks is often

common, especially in friend groups, and recommended that for a risk to be considered cyberbullying, the user should be meaner in their remarks. Therefore, most teens revised Emily's cyberbullying risky scenario to make more condescending remarks, specifically about others' appearances by body-shaming or giving backhanded compliments. For instance, one of the participants added a new cyberbullying remark, *"OMG that outfit would look so much better on me :)"* (P3, 17-year-old, Female). Similarly, another participant added a cyberbullying remark for Emily, which made hurtful comments about their body, *"You look so fat in these clothes, why do you even bother dressing up?"* (P5, 18-year-old, Female). Other teens believed that sometimes such cyberbullies make them feel unsafe by judging and backbiting about others, such as *"R u actually friends with (someone), aren't they annoying?"* (P9, 13-year-old, Female).

Persuasive Spam Links: Regarding the spam bot link risk, teens suggested that spam bots should first attempt to interact with users similar to real human accounts and then send malicious content, in order to increase believability. Moreover, they recommended that the bot should send personalized click-baits to match the type of spam they receive online and to make the link more deceiving such as *"Hey, is this you?"... "No? can you at least check this out."* Other ways in which teens suggested making the spam link more believable and enticing was by offering money, giftcards, or gaming points, such as *"Congrats, you've won our giveaway from Target! Click here to redeem..."* (P10, 15-year-old, Male). A few teens also changed the personality of the bot to be extroverted as it would want to initiate interaction with as many people as possible for spamming, which does not match an introverted personality. Additionally, a few teens commented that such spam links often come from hacked accounts of their friends, which often increased their chances of clicking the links as they came from someone they know.

Trustworthy Predatory Risk: Many teens redesigned the creepy predator risk to make personalized comments about the teen, related to a photo they uploaded, instead of generic remarks. Teens also thought that such risks often fall into two categories; a) stalkers who message you inappropriate comments out of the blue, or b) predators with a an ulterior motive who slowly build trust with the teen, and befriend them before making inappropriate comments. Many teens thought that for the purpose of our study, it would be realistic for Dave to build trust with the teen first, before sending risky messages. Other teens recommended that such risks are often accompanied by the user trying to share their problems and attempting to gain the teens' sympathy, and later revealing their risky motives such as requests to meet in-person. For instance, P14 added a quote for this risky user, *"I really enjoy talking to you and would love to get to know you better...want to meet up?"* (P14, 16-year-old, Male).

4 DISCUSSION

Our findings highlight key considerations for designing user personas and risk scenarios for conducting adolescent online safety research in a realistic, ecologically valid environment. We found that curating accurate and suitable personalities is a critical aspect when designing risk scenarios to be authentic and convincing for teens. Additionally, most teens wanted the personas to be believable, and considered the motivations of the personas to match their risk type carefully (e.g., taking into account a popular user's reputation who would not perpetuate a risk in public). Prior work in this space has largely focused on ensuring realism by selecting hypothetical risk scenarios with participants, through large-scale surveys [8, 11], with little emphasis on *who* the risk is coming from and the characteristics of the risky user. Overall, our findings demonstrate the importance of further improving experimental realism by co-designing user personas with teens and simulating teens' real-world social media experiences as much as possible. The importance of this is further amplified when working with teens, who have unique experiences and developmental needs, that cannot be imitated by researchers alone.

On the other hand, we found that it is equally important for the risk scenarios to be *nuanced* and *contextualized* to create a realistic setting for teens. For risk scenarios to be realistic, teens recommended subtlety in risks and for the conversation to build up before the risk is introduced. Therefore, building shared context matters as the risks teens face online are often not too sudden or direct. Additionally, the severity of the risks depends on the type of risk, where teens suggested that information breaching should be more subtle, whereas cyberbullying should be harsher. Yet, simulating these risks with teens as a vulnerable population comes with several challenges. The risk scenarios should imitate realistic risks, but at the same time, should not put teens at a risk higher than what they would encounter in their everyday interactions. Yet, prior work [6] emphasizes the need to include teens at every stage of the research, to ensure that they are benefitted and their needs are met in the research. Therefore, there is a need to further understand ethical considerations for conducting research that simulates online risks with teens to ensure that the research prioritizes teens' needs and well-being. Overall, this work contributes to a deeper and more nuanced understanding of the type of bad actors and risky scenarios teens encounter online, as well as their interplay. The contributions of this research are both tangible for the design of a realistic evaluation of online safety nudges while providing broader implications that inform researchers in our HCI and CSCW communities on how we can conduct realistic research on sensitive topics with teens like online risks. In summary, this research will enable researchers to more accurately assess the impact of different risk scenarios, and better understand how these scenarios play a role in evaluating various interventions.

5 CONCLUSION

The findings from this study emphasize the importance of designing simulations that are sensitive to the needs and perspectives of teens and that provide a nuanced and realistic environment for evaluating online safety interventions. Moving forward, we plan to implement the designs from this study in a between-subjects experimental design with teens to evaluate the effectiveness of the different types of nudges from our prior work. This will allow us to evaluate designed interventions within ecologically valid environments.

ACKNOWLEDGMENTS

This research was supported by the William T. Grant Foundation (#187941) and National Science Foundation under grants IIS-2333207. Any opinions, findings, conclusions, or recommendations

expressed in this study do not necessarily reflect the views of our sponsors.

REFERENCES

[1] Zainab Agha, Karla Badillo-Urquiola, and Pamela J Wisniewski. 2023. " Strike at the Root": Co-designing Real-Time Social Media Interventions for Adolescent Online Risk Prevention. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–32.

[2] Zainab Agha, Zinan Zhang, Oluwatomisin Obajemu, Luke Shirley, and Pamela J. Wisniewski. 2022. A Case Study on User Experience Bootcamps with Teens to Co-Design Real-Time Online Safety Interventions. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–8.

[3] Mamtaj Akter, Amy J Godfrey, Jess Kropczynski, Heather R Lipford, and Pamela J Wisniewski. 2022. From Parental Control to Joint Family Oversight: Can Parents and Teens Manage Mobile Online Safety and Privacy as Equals? *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–28.

[4] Ashwaq Alsoubai, Jihye Song, Afsaneh Razi, Nurun Naher, Munmun De Choudhury, and Pamela J. Wisniewski. 2022. From 'Friends with Benefits' to 'Sextortion:' A Nuanced Investigation of Adolescents' Online Sexual Risk Experiences. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 411 (nov 2022), 32 pages. https://doi.org/10.1145/3555136

[5] Zahra Ashktorab and Jessica Vitak. 2016. Designing cyberbullying mitigation and prevention solutions through participatory design with teenagers. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 3895–3905.

[6] Karla Badillo-Urquiola, Zachary Shea, Zainab Agha, Irina Lediaeva, and Pamela Wisniewski. 2021. Conducting risky research with teens: co-designing for the ethical treatment and protection of adolescents. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–46.

[7] Karla Badillo-Urquiola, Diva Smriti, Brenna McNally, Evan Golub, Elizabeth Bonsignore, and Pamela J Wisniewski. 2019. Stranger danger! social media app features co-designed with children to keep them safe online. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. 394–406.

[8] Dominic DiFranzo, Samuel Hardman Taylor, Franccesca Kazerooni, Olivia D Wherry, and Natalya N Bazarova. 2018. Upstanding by design: Bystander intervention in cyberbullying. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[9] Figma. 2023. FigJam Turn possibilities into plans. https://www.figma.com/figjam/

[10] Heidi Hartikainen, Netta Iivari, and Marianne Kinnula. 2019. Children's design recommendations for online safety education. *International Journal of Child-Computer Interaction* 22 (2019), 100146.

[11] Patricia Kearney, Timothy G Plax, Val R Smith, and Gail Sorensen. 1988. Effects of teacher immediacy and strategy type on college student resistance to on-task demands. *Communication Education* 37, 1 (1988), 54–67.

[12] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. 2010. Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT)* 10, 2 (2010), 1–31.

[13] Letitia Lew, Truc Nguyen, Solomon Messing, and Sean Westwood. 2011. Of course I wouldn't do that in real life: advancing the arguments for increasing realism in HCI experiments. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. 419–428.

[14] Jinkyung Park, Joshua Gracie, Ashwaq Alsoubai, Gianluca Stringhini, Vivek Singh, and Pamela Wisniewski. 2023. Towards Automated Detection of Risky Images Shared by Youth on Social Media. In *Companion Proceedings of the ACM Web Conference 2023*. 1348–1357.

[15] Jinkyung Park, Irina Lediaeva, Amy Godfrey, Maria Lopez, Kapil Chalil Madathil, Heidi Zinzow, and Pamela Wisniewski. 2023. How Affordances and Social Norms Shape the Discussion of Harmful Social Media Challenges on Reddit. *Human Factors in Healthcare* (2023), 100042.

[16] Afsaneh Razi, Karla Badillo-Urquiola, and Pamela J Wisniewski. 2020. Let's talk about sext: How adolescents seek support and advice about their online sexual experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[17] Emily A Vogels. 2022. Teens and cyberbullying 2022. https://www.pewresearch.org/internet/2022/12/15/teens-and-cyberbullying-2022/

[18] Emily A Vogels, Risa Gelles-Watnick, and Navid Massarat. 2022. Teens, social media and technology 2022. https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/

[19] Ashley Marie Walker, Yaxing Yao, Christine Geeng, Roberto Hoyle, and Pamela Wisniewski. 2019. Moving beyond'one size fits all' research considerations for working with vulnerable populations. *Interactions* 26, 6 (2019), 34–39.

[20] Maximilian Zinkus, Oliver Curry, Marina Moore, Zachary Peterson, and Zoë J Wood. 2019. Fakesbook: A social networking platform for teaching security and privacy concepts to secondary school students. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 892–898.