The version of record of this article, first published in FICC 2024: Advances in Information and Communication, is available online at Publisher's website:
http://dx.doi.org/10.1007/978-3-031-54053-0_40

A Stakeholders' Analysis of the Sociotechnical Approaches for Protecting Youth Online

Xavier Caddle[1*], Jinkyung Katie Park[2] and Pamela J. Wisniewski[2]

[1*]Computer Science, University of Central Florida, 4328 Scorpius St, Orlando, 32816, FL, USA.
[2]Computer Science, Vanderbilt University, 2201 West End Ave, Nashville, 37235, TN, USA.

*Corresponding author(s). E-mail(s): xavier.caddle@ucf.edu;
Contributing authors: jinkyung.park@vanderbilt.edu;
pamela.wisniewski@vanderbilt.edu;

Abstract

Feasible and developmentally appropriate sociotechnical approaches for protecting youth from online risks have become a paramount concern among human-computer interaction research communities. Therefore, we conducted 38 interviews with entrepreneurs, IT professionals, clinicians, educators, and researchers who currently work in the space of youth online safety to understand the different sociotechnical approaches they proposed to keep youth safe online, while overcoming key challenges associated with these approaches. We identified three approaches taken among these stakeholders, which included 1) leveraging artificial intelligence (AI)/machine learning to detect risks, 2) building security/safety tools, and 3) developing new forms of parental control software. The trade-offs between privacy and protection, as well as other tensions among different stakeholders (e.g., tensions toward the big-tech companies) arose as major challenges, followed by the subjective nature of risk, lack of necessary but proprietary data, and costs to develop these technical solutions. To overcome the challenges, solutions such as building centralized and multi-disciplinary collaborations, creating sustainable business plans, prioritizing human-centered approaches, and leveraging state-of-art AI were suggested. Our contribution to the body of literature is providing evidence-based implications for the design of sociotechnical solutions to keep youth safe online.

Keywords: Youth online safety, Sociotechnical approach, Stakeholder interviews

1 Introduction

Internet and social media have become a significant part of teenagers' lives. According to Pew Research, 97% of teens (ages 13-17) in the U.S. go online daily, 95% have access to smartphones, and 95% engage on one or more social media platforms [1]. The prevalent use of mobile smartphones and social media facilitated a wide array of benefits such as social connections and learning [2], but it also amplified online risks [3]. In 2022, almost half of U.S. teens reported having experienced online risks ranging from harassment, sexual messages, and threats [4]. Given the adverse impacts of such risks on youth [5], the phenomena of youth online risk are now one of the critical concerns to many [6–8].

Youth online safety solutions have been investigated through a sociotechnical lens as they are instances of work systems that Sociotechnical Theories (STS) view as being comprised of social and technical subsystems [9]. Literature on the sociotechnical solutions to teen online safety often puts a strong emphasis on parental control strategies (i.e., parental control app) [10–12] and this has led to sociotechnical design practices for creating privacy-invasive parental control solutions that overly restricts and monitors teens' online activities [13, 14]. Such surveillance-based and privacy-invasive strategies may shield teens from online risks but at the cost of exacerbating the privacy tensions between parents and teens [15–18].

Researchers have conceptualized more collaborative technologies that move away from surveillance-based approaches to ones that engage both teens and parents for digital well-being [13, 14, 19–22]. Existing research on adolescent online safety tends to focus on exploring the perspectives of teens and parents, the primary stakeholders, in terms of safety strategies implemented by parents, tensions in parent-teen relationships, privacy implications, and more. However, the perspectives of the secondary stakeholders (e.g., IT professionals) who play a pivotal role by creating the online safety interventions used by the primary stakeholders are rarely explored. In this work, we bring into focus the viewpoints of members of this population by expounding on their perspectives as inventors, implementors, and investigators of youth online safety solutions. In doing this, we address the following research questions:

- **RQ1**: *How do stakeholders in the domain of youth online safety approach the problem of youth online risk to keep youth safe online? And what are the challenges associated with sociotechnical approaches towards adolescent online safety?*
- **RQ2**: *How might we overcome these sociotechnical challenges and make strides in protecting youth online?*

To answer these questions, we conducted 38 semi-structured interviews with a variety of youth online safety stakeholders. In analyzing our results, we found that stakeholders take three main sociotechnical approaches to youth online safety: 1) online risk detection using Artificial Intelligence (AI) and Machine Learning (ML), 2) generalized approaches such as providing implementation support, and 3) creating parental control software. However, they identify significant challenges in these approaches such as identifying accountability, issues collecting and using youth data, issues providing privacy guarantees, and providing accurate results. We found that the solutions

to these issues are thought to be rooted in creating sustainable solutions that have temporal resilience, and ensuring technology solutions are designed using multidisciplinary input. By engaging closely with the secondary stakeholders' views, our paper is the first research paper to systematically unpack the ecosystem of online safety, the tension between privacy and protection, the role of legislation, and the expectation of parents fulfilling a role. We highlight the novel insights from those stakeholders such as increasing antagonism towards big tech and how this tension can put companies off trying to develop tools in this space, and how there can be a lack of access to data for small players in this space. Based on our findings, we recommend specific solutions that can be actually implemented for concrete changes and impacts they may have on adolescent online safety. The rest of this paper is structured as follows: Section 2 provides study participant criteria and recruitment strategy, Section 4 presents the qualitative analysis of the participant interviews, Section 5 presents an evaluation of the results including implications for designing youth online safety solutions, and Section 6 concludes the paper.

2 Related Work

Youth Online Safety: The growth of online and social networks has spawned more privacy and security risks to youth than ever before. In response, youth online safety literature has emerged over the past two decades. Scholars from various fields have examined a wide range of topics such as the definition, characteristics, prevalence, and perceptions of different online risk experiences [23–28], adverse impacts of online risks on youth [29, 30], and strategies to counter online risks [19, 31–35]. Existing literature documents that cyberbullying and harassment [35, 36], sexual solicitation [33, 37], exposure to explicit content [34, 38], and privacy breaches [19] are prevalent risks youth encounter. Cyberbullying, one of the most prevalent online risks, is associated with emotional stress and negative emotions such as anger, fear, and depression [39].

To mitigate youth online risks and promote their digital well-being, multiple approaches have been explored. Parental mediation or digital parenting [31, 40], and family communication [41] are non-technical and family-centered approaches to promote teen online safety. Digital parenting includes, for example, the monitoring of teens' technology use, rule setting and enforcement, education of digital skills, and exploration of online opportunities [31]. However, parent- and family-focused approaches to adolescent online safety assume a significant level of privilege in many ways. The most vulnerable teens to online risks (e.g., foster youth) often do not have parents who can actively engage in ensuring their online safety [42, 43]. Even in traditional families, communication regarding the risks teens experience online is particularly poor [18] because in many cases, parents are overly judgmental and overreact when teens disclose their online risk experiences, exacerbating the problem instead of teaching teens how to effectively manage online risks[17]. Researchers have focused on the effects of safety strategies implemented by parents [44], the views of parents [45] and youth [15] on online privacy, and exposing public opinion on the regulations enacted by governments [46]. The views of the creators of sociotechnical solutions on these same points have not garnered as much dressage in the literature.

Sociotechnical Solutions: Youth online safety solutions have been investigated through a sociotechnical lens [9, 47] where environment within which youth exist is a social component, while the digital technology they use (e.g., internet platforms and online safety tools) is a technical component. Literature on the sociotechnical solutions to teen online safety often puts a strong emphasis on parental control strategies (i.e., parental control app) to restrict and monitor teens' online activities [12–14]. The use of online surveillance by parents may shield teens from online risks, but at the cost of teens' autonomy [48], trust between parents and teens [49], and positive family value as whole [14]. Al-based algorithmic risk detection has also been proposed to automatically identify cyberbullying [32, 50–52], sexual risks [33, 53, 54], and other risks faced by teens in online contexts [34]. Yet, creating the datasets to train machine learning algorithms raises additional privacy and surveillance issues as it involves the use of teens' high-risk interactions (e.g., predatory sexual grooming) that are more likely to occur through the private sphere [33].

Some researchers have conceptualized more collaborative technologies that move away from surveillance-based approaches to ones that engage both teens and parents for digital rule-setting and managing online activities [19, 21, 22]. Others have called for new sociotechnical solutions that promote teen self-regulation of online risks [13, 14, 20]. Yet, very few researchers have studied evidence-based sociotechnical interventions for adolescent online safety [26]. In recent work, perspectives of secondary stakeholders in adolescent online safety (i.e., social service providers (SSPs)) were explored [55]. They uncovered that SSPs faced challenges such as heavy reliance on youth self-reports to discover their online risk experiences and that SSPs suggested AI as an early detection system to monitor youth online safety. However, SSPs were concerned that such a solution would not be feasible due to a lack of access to the necessary digital trace data (e.g., social media) and concerns about violating the trust relationships they built with youth. In this work, rather than focusing on a specific stakeholder group (e.g., SSPs), we interviewed multiple secondary stakeholder groups (e.g., IT professionals, entrepreneurs, clinicians, educators, and researchers) to understand the entire ecosystems of the online safety domain, challenges that various stakeholders are experiencing, and potential solutions to mitigate those challenges and move forward. By taking a broader view of the youth online safety ecosystem, we seek to start addressing the gap in the literature between primary and secondary stakeholders while uncovering possible paths forward where the voices and needs of all stakeholders in this context are openly considered.

3 Methodology

Over a six-week period from June 2021 into July 2021, 38 semi-structured interviews were conducted within the United States with Entrepreneurs, IT professionals, Educators, Clinicians, and Researchers. We started this work as part of a greater project under the NSF Innovation Corps (I-Corps^{TM}) program [56] where researchers acquire the knowledge and tools to become entrepreneurs through training on customer discovery and monetization of research outputs. The program required interviewing prospective customers as well as subject matter experts within the interest area of the

possible product or technology, that being adolescent online safety and online risks in this instance, to understand the present needs in the field and how those needs could best be met. In order to obtain valuable responses from participants, we used a semi-structured interview methodology. We focused on asking questions that would help us understand how they approach the issue of youth online risk, what obstacles they encountered, and mitigation strategies for those challenges to keep teens safe online. While the interview questions mentioned the threats of cyberbullying, sexual solicitations, or sharing of inappropriate content, the interviews were semi-structured in nature to allow participants to speak about the youth online risk detection space, in general, to obtain a fuller understanding. The semi-structured interview questions are included in Appendix A.

3.1 Participant Recruitment and Procedure

We recruited participants who were 18 years of age or older, residing in the United States, currently working with adolescents (minors between the ages of 13 and 17), conducting online safety research on/with youth in the target age range, or working on technology used by such youth. If participants performed such duties in the previous 5 years, we considered their input as well. We developed our inclusion/exclusion criteria to figure out the key secondary stakeholders in the adolescent online safety domain and use them to create an initial list of stakeholders to contact. Also, by the end of each interview, we asked our participants to suggest other secondary stakeholders who may be able to provide insights into the same topic. We used the two strategies to complement one another. Using these methods, we met with a diverse group of participants including Entrepreneurs ($n = 13$), Information Technology Professionals ($n = 9$), Researchers ($n = 8$), and Clinicians/Educators ($n = 8$). The participants' characteristics are listed in Appendix B.

Participants were briefed with an IRB-approved clarification of the research and given the option to not participate if they so desired. All interviews were held online via Zoom video conferencing to accommodate for COVID-19 limitations at that time. Each interview ran for 30 minutes with some lasting longer. Most of the interviews had two interviewers present with one leading the interview and the other writing notes. All were asked for consent to record audio and video at the start of each interview. No financial incentive was awarded in this study due to the constraints of the program.

3.2 Data Analysis Approach

All interviews were conducted via Zoom and were recorded (both audio and video). Each interview was automatically transcribed using the live transcription feature of Zoom. In order to correct the transcriptions for errors if any, two research assistants observed each video and/or listened to the audio recordings. For qualitative analyses, we followed the thematic analysis approach [57]. The first and last authors identified the key components for which to code the transcribed data and established the first codebook of emergent codes. To address our research questions, the interview transcripts were first coded for the types of key stakeholders in youth online safety and the approaches they took to deal with the problem of youth online risk to keep youth

safe online. Next, they were coded for the challenges in sociotechnical approaches to keep youth safe online **(RQ1)**. Lastly, the interviews were coded for ways to overcome the challenges related to sociotechnical approaches to keep youth safe online **(RQ2)**. The first two authors equally split and coded the interview transcripts based on the first codebook with allowance for additional codes to arise throughout this process. All revised codes were finalized after discussion amongst all authors using consensus building [58]. The final themes that correspond to our main research questions are shown in appendix tables C2, C3, and C4.

4 Results

In this section, we present the major themes found in our study. We first present the sociotechnical approaches to online safety the participant groups undertake. Then, we present the challenges they identify to working within or creating solutions in the area of online safety (RQ1). We then present the solutions to those challenges (RQ2), and finally, any needs identified by the participant in the area of online safety.

4.1 Sociotechnical Approaches and the Challenges Related to Sociotechnical Approaches to Youth Online Safety (RQ1)

There were three major sociotechnical approaches that our participants took in addressing youth online safety within the context of their job: 1) online risk detection using Artificial Intelligence (AI) and Machine Learning (ML) techniques, 2) taking a general approach to online safety, and 3) implementing parental control software. Stakeholders shared that they use AI and ML techniques to identify instances of cyberbullying, privacy violations, and scanning devices belonging to youth for pornography. They also shared some general approaches to online safety such as using privacy-enhanced web browsers and/or internet router functionality to block access to inappropriate sites. They also mentioned social media sites restricting access to their platforms to those below certain ages as measures that can are currently taken to protect youth online. Finally, parental control software mentioned by our participants included tools and platforms that allow parents to monitor what youth do online, as well as implement restrictions such as screen time management and content filtering. Overall, our participants shared the challenges associated with sociotechnical approaches to keep teens safe online such as the "Trade-offs between Privacy and Protection" and "Tensions around Diffused Accountability" in the area of youth online safety. "Risk is Subjective" was the next most prevalent theme, followed by "Technical solutions are costly" and "Lack of Necessary but Proprietary Data". In the sections below, we discuss the major challenges faced by each stakeholder group.

4.1.1 Trade-offs between privacy and protection

In our dataset, trade-offs between teen privacy and protection emerged as the major challenge associated with sociotechnical solutions to keep teens safe online. Under this theme, participants discussed topics such as teen privacy issues and legal liabilities around the use of teens' behavioral data. For entrepreneurs, teen privacy is

an unavoidable and complicated issue associated with online risk detection systems (as a product and/or service) that they provide to their customers for profit. They perceived teen privacy issues as a conflict between their business model and the responsibility to protect teens from online risk. Some entrepreneurs expressed these conflicting feelings where they respect teen privacy to be protected, while they need to work with fine-grained teen data (which can inherently invade teen privacy) to provide efficient risk detection services to their customers.

"For sexual abuse manipulation, a lot of that happens in direct message right so like if the platform is really going to touch on any of that... You know some of these communication platforms are very social and their appeal and their purpose is the privacy and lack of moderation." - P23, Entrepreneur

Others mentioned that teen privacy issues are hard to address because of the subjective and varying nature of privacy perception. That is, the degree to which teen privacy should be protected varies depending on parenting style, which makes it harder for entrepreneurs to come up with technical solutions that satisfy all customers. The varying degrees of teen privacy concerns among customers were in part, closely related to the subjective nature of online risk itself.

"...some parents are not going to be willing to invade their children's privacy and that becomes a parenting question on how the parents' philosophy is versus different parents." - P32, Entrepreneur

IT professionals contributed the most to the trade-offs between teen privacy and protection theme. Similar to entrepreneurs, IT professionals spoke mostly about teen privacy issues associated with risk detection systems but focused more on legal and technical perspectives. For instance, some IT professionals were having difficulties with training online risk detection systems due to legal responsibilities such as mandated reporting of child pornography. They mentioned that although they need to work with teens' data to train risk detection algorithms, they do not even attempt to do so because they were terrified by the legal liabilities and responsibilities that they need to deal with.

"we are only looking at 19 and up because of those (legal) concerns ... managing adolescent data is terrifying." - P22, IT Professional

Similar to entrepreneurs, IT professionals perceived teen privacy issues as an inevitable pitfall of data-driven online risk detection systems. This was because fine-grained and high-quality teen data is needed to create effective automated systems to detect online risk. At the same time, privacy risks arise at every stage of the data pipeline. That is, for IT professionals, teen privacy issues were ethical dilemmas between preserving teen privacy and protecting teens from online risks.

"It's a big challenge, I would say, like the balance of protection and privacy, because, in order to protect your kids you do have to have some of that privacy." - P24, IT Professional

Researchers shared that coming up with sociotechnical solutions to keeping teens safe is challenging because of a strict legal framework that applies to teen data. Even with complete compliance with the legal framework, researchers also felt that challenges still remain for many since data-driven risk detection technology depends heavily on the use of teens' intimate data.

"people going to be upset that Facebook or Google or Apple is doing something with their intimate images, in order to detect child pornography." - P14, Researcher

Meanwhile, unlike other stakeholders, clinicians/educators spoke about teen privacy issues focusing on parent-teen relationships. For instance, they worried that implementing parental control software without teen assent is problematic because oftentimes, it is not the teen but their parents who consent to the use of such solutions to monitor teen online interaction. They saw this mismatch in privacy expectations between parents as the root cause of conflict between them which will eventually hinder them to build a trusted relationship to discuss online safety issues.

"The other problem is this kind of expectation of privacy... teens don't want us to see what they're actually doing online that's causing a lot of conflicts" - P12, Educator

4.1.2 Tensions around diffused accountability

Tensions around diffused accountability among different stakeholders arose as the second major challenge theme in our dataset. Under this theme, the participants touched upon various topics such as challenges in deciding the right stakeholders to involve and catering to different opinions/standards, and also complaints about big-tech companies for failing to play their roles to protect youth from online risks. The entrepreneurs contributed to the majority of this theme by criticizing big-tech companies' current practices in addressing youth online safety issues, in many cases, due to *closed and proprietary nature* of big-tech business. In their view, big-tech companies are shirking their responsibility to address teen online safety issues when they are the ones who have the ability and resources to do so.

"I think the social networks are like the place that should be addressing this and I don't think they're going to until something catastrophic or some real threats to their revenue." - P26, Entrepreneur

Meanwhile, IT professionals spoke mostly about tension around diffused responsibilities among different stakeholders. None of the IT professionals complained about big-tech companies which other stakeholders often did in their interviews. Rather, one IT professional from a big-tech company (anonymized upon request by the participant) addressed the criticisms toward their company and acknowledged the social

responsibilities that other stakeholders expect from their company.

"We're in an interesting position where we're constantly under the microscope and I think it's deserved and I think that's a lot of the criticism that comes our way is also deserved and I think that we need it." - P7, IT professional

Researchers also addressed the diffusion of accountability among different stakeholders as a strong challenge associated with sociotechnical solutions. In their views, the current teen online safety community is a highly *decentralized* community with various perspectives and interests. One of the major concerns with the decentralized community for researchers was the *bystander effect* where everyone waits and sees but no one takes the lead to intervene in teen online risk situations. Some researchers specifically blamed tech people for being too competitive and dishonest. At the same time, others recognized that it is not about specific members, but rather building a centralized community in the multi-disciplinary field is not easy because it is not static but dynamic in nature.

"Most of the issues and things come from is around that sense of responsibility and accountability. The biggest reason for the bystander effect is that diffuse means of responsibility" - P2, Researcher

Educators/clinicians also expressed critical views toward the big tech companies. They almost demonized big tech companies for polluting the water and hence, claimed that the companies should hold social responsibility for it. Some educators manifested strong skepticism toward big tech that they do not even expect any protective measures placed by big-tech companies; even if they did, the protective measure will be inherently monopoly. Finally, they resonated with us an important point that when corporations shirk their responsibility to protect teens, and parents cannot afford to do so by themselves, it is *teens* who will be left on their own.

"Corporations themselves aren't going to fix the problem, and parents don't really have the tools or the understanding to fix the problem, then it does really fall on the kids to figure it out." - P27, Educator

4.1.3 Risk is subjective

The next major challenge theme that emerged from our interviews was the *subjective and volatile* nature of risk. Under this theme, participants addressed mostly the difficulties with developing accurate and stable risk detection systems due to the volatile, nuanced, and contextualized nature of online risks. Entrepreneurs contributed the most to the challenges related to the subjective nature of risk. As a product and service, entrepreneurs perceived that defining and detecting subjective risks in ways that satisfy all customer groups is nearly impossible. Even with a set of risk definitions that satisfy all, it is still challenging for entrepreneurs to come up with stable and valid technical solutions to keep teens safe due to the volatile nature of risks (how

fast the ways predators attack the users are changing).

"You're defining the standard of that what that grooming threat is and what is bullying is and parents are not going to like that"- P13, Entrepreneur

Establishing ground truth (e.g., certain interaction is risky vs. non-risky) to build valid and accurate online risk detection systems was one of the huge challenges for IT professionals due to the nuanced nature of online risk. Similar to what entrepreneurs mentioned, IT professionals emphasized difficulties in first, defining what harmful content is and second, continuously coming up with new solutions given how rapidly online risks evolve.

"...as soon as you come up with a defense, the people you're defending against are going to come by attacking it to come up with a way to breach it."- P18, IT Professional

The subjective nature of the risk was also mentioned by the researchers and educators/clinician groups. They both spoke about low accuracy and potential bias inherent in online risk detection/monitoring systems to detect nuanced risks. On the one hand, they expressed concerns about highly contextualized risks (e.g., sarcasm) and emphasized the importance of *human involvement* in building youth risk detection systems. At the same time, they acknowledge that such approaches (e.g., human coding) require extensive and time-consuming efforts.

"...they went through and did it by hand which is enormously time-consuming because the algorithm was having trouble with some of the aspects of the interactions that are just hard to explain."- P31, Clinician

4.1.4 Lack of necessary but proprietary data

For the data access theme, participants mainly debriefed the difficulties with a lack of necessary data to build technical solutions. The challenges under this theme were dominantly addressed by entrepreneurs. Particularly, they shared the difficulties of having access to the necessary data to train AI-based online risk detection systems. Once again, for many entrepreneurs, the *closed* and *proprietary* big-tech industry that owns the majority of necessary data was to blame. They lamented that due to the financial burden to access proprietary data, oftentimes, they have no choice but to adopt cheap solutions which will not benefit anyone.

"Oftentimes what happens is that they go the cheap route. And the cheap route never gives them the minimum of what they need."- P29, Entrepreneur

IT professionals were also experiencing challenges due to a lack of access to necessary data. However, unlike the entrepreneurs, IT professionals spoke less about the proprietary nature of necessary data that the big-tech industry monopolized and more about the lack of *quality* data that they need to build effective online risk detection

systems. They perceived that the teen data that is currently available for them is not good enough to build effective risk detection systems and that with more fine-grained data in scale, a huge leap in the advancement of the risk detection system is possible.

"...we don't have access to the data about what Apps their kids are using and what is actually going on in that page and that level of detail is not something that we have today."- P24, IT Professional

Challenges associated with the lack of necessary data were rarely mentioned by both the researchers and clinician/educator groups. They worry that technical solutions are stalled because in most cases, they require personal data when the perception of privacy is becoming a more sensitive issue.

"I found technology solutions are very lockdown just because it's much harder to do on the phone because they try to keep things more private and, like the perception of privacy is very important for these devices."- P4, Educator

4.1.5 Technical solutions are costly

Finally, participants shared the difficulties with having technical solutions to promote teens' online safety due to their costs. For instance, as a parent and a lead of a business, one entrepreneur expressed the dilemma that they are having between the cost and the scalability of risk detection technology.

"I think part of the challenge, unfortunately, from an individual parent perspective, it really depends on what the end result cost"- P32, Entrepreneur

The difficulties related to the costs of technical solutions were often addressed by IT professionals. Particularly, they pointed to the enormous volume of data and the following financial costs that are needed to build effective systems. Some IT professionals mentioned that the market for online risk detection systems is still limited and hence, it is difficult for them to come up with solid budgeting plans to convince other stakeholders.

"We would also have to justify that it's worth the investment and that the cost to use this service would be worth it."- P36, IT Professional

Similar to entrepreneurs and IT professionals, researchers acknowledged that having sustainable solutions to develop technical solutions to promote teens' online safety is challenging due to the vast amount of data that is needed, and the costs associated with getting access to and processing data. Unlike the other three stakeholders groups, none of the clinicians/educators mentioned the challenges related to the financial aspects of sociotechnical approaches to teens' online safety.

4.2 The Solutions to Overcome the Challenges Related to Sociotechnical Approaches to Youth Safe Safety (RQ2)

At a high level, the solutions posited by the participants were grouped under 4 themes, "Centralized and Multi-disciplinary Collaboration (CMC)," "Sustainable Plans to Keep Teens Safe Online," "Prioritizing human-centered approaches," and "Leveraging the state-of-art AI." In the following sections, we present the views on these themes shared by the participant groups.

4.2.1 Centralized and multi-disciplinary collaboration (CMC)

Participants discussed topics such as providing open-source solutions, application programming interfaces (API), scraping internet data, and integrating with multiple systems when discussing solutions that were categorized under CMC. Having open-source or low-cost system integration with youth online safety solutions and forming partnerships were the two areas entrepreneurs shared as solutions. In speaking of integration, they spoke of companies who recognize their product offering needs to protect youth but lack the resources, both human and capital, to do so. Providing a centralized, standardized API that could be accessed in an open-source model was spoken of in a favorable manner. Forming partnerships was mentioned slightly more than system integration in this group. In speaking about partnerships, entrepreneurs discussed the need to reach out to stakeholders in youth online safety on common grounds so strategic partnerships can be made to incorporate different skill sets.

"*Tailoring your message to each of the specific focus groups... we have to look at all the different ways to effectively go into different groups with their mentality and show the value that we bring.*" - P10, Entrepreneur

IT professionals mostly spoke about creating partnerships through technology solutions integrations when discussing CMC. In contrast to the entrepreneurs whose viewpoints leaned towards convincing other parties to work together, IT professionals viewed collaboration between competitors to protect youth online as something most of their peers readily agree upon. In this light, they viewed partnerships and collaboration as the major method of youth online safety. Efforts to bring together youth online safety stakeholders in conferences, consortiums, and systems integrations were put forward as solutions to the challenges. This was seen as a solution as subject matter experts from different realms are included to enhance solutions and thereby improve the protections designed. To a lesser extent, IT professionals also spoke of sharing data sets, making data sets open, and government efforts to have data sets created specifically for companies to design online safety solutions.

"*Companies that typically would be considered like our competitors...When it comes to health and safety, they are coming together on a safe community of practice and they're sharing insights and learnings*" - P36, IT Professional

CMC solutions posited by researchers were on the topic of forming partnerships. Discussing this, researchers spoke of the need to form collaborations with non-profit organizations, non-governmental organizations (NGOs), and groups who work towards social/public good, for the purpose of collecting information for studies, and also sharing findings from studies that might help the particular group in their own efforts. Educators/clinicians also only discussed partnerships in this manner.

"*I think it's know your community, find somebody in the community or communities you're working with that is behind you and has your back and is going to support you and you know help, I would say, is the best approach.*" - P16, Researcher

4.2.2 Sustainable plans to keep youth safe online

The second major theme for RQ2 was sustainability. In discussing sustainability as a solution to youth online safety challenges, participants talked about pricing strategies, finding funding for projects, and targeting the right customers, otherwise known as market segmentation. The market segmentation was especially contributed by Entrepreneurs. They talked about making sure downloadable products are intentionally placed where the particular market segment has a high probability of seeing the product, as well as strategies such as search engine optimization (SEO). They also stressed that there needs to be confidence in the created product such that it can be demonstrated to investors without issue.

"*You have to be able to target a message to people of different socio-economic backgrounds...you have to show the value of what's happening in the perspective space and to that targeted market*" - P13, Entrepreneur

IT Professionals discussed sustainability as a solution through the lens of ensuring the business offers support in the form of maintenance and product development to clients. They saw these add-on services as a way of obtaining recurring income and up-selling core products to keep the business running. In doing so, they talked about finding *gaps* in the client's business which they could fill to offer more complete or holistic solutions.

"*...you can internally grow what you're doing for that entity, if you show the capability to provide the support that the entity requires.*" - P18, IT Professional

Only a single instance in the researcher category contributed to the sustainability theme. When discussing this theme, the researcher spoke about the need to be innovative to secure funding for building products and growing businesses. Educators/clinicians contributing to the sustainability theme also discussed market segmentation as the solution but were very direct at stating what that segment is. They suggested targeting medical practitioners, insurance companies, and therapists as a means to get online safety solution products used as these groups were said to be in contact with individuals who would be in need of the product.

"*If you could keep 10, 20, 30 kids out of the hospital, you can recoup your investment easily* " - P31, Educator

4.2.3 Prioritize human-centered approaches

The third major theme in solutions to youth online safety challenges that participants shared was categorized under prioritizing human-centered approaches to youth online safety. Participants discussed topics such as empowering users to make healthy decisions while taking part in online activities and also making sure the technology or products are easy to use. IT Professionals did not contribute to this theme. In discussing this theme, Entrepreneurs focused on making sure the tone of messages used in marketing is more strength-based rather than deficit-based. This is done by focusing on the benefits as opposed to highlighting restrictive capabilities. Some entrepreneurs also spoke about making sure solutions are easy to use, both for end-users who may use the solution directly and for third parties who may integrate solutions into their own products.

"*I would want something that helps that my kid would want to use for themselves to monitor their own mental health ...I would want them to have something that was easy, simple, reliable, engaging all of those things*" - P26, Entrepreneur

Researchers contributed to this theme by speaking about empowerment methods. While they also discussed the messages used by solutions, they were more concerned with the messages produced by solutions which would then be seen by end-users. They spoke about this from the perspective of automated tools which process data and provide summaries or suggestions about the perceived state of the data provider. They noted that care must be taken with this messaging as it could be harmful to the end-user, and also push them away from using solutions.

"*Like it could also be potentially really harmful... if they don't have the type of structure support structure around them that once you're bringing this up, they have someone to like unpack and deal with those things*" - P5, Researcher

Researchers also mentioned that *holistic approaches* to youth online safety are necessary which take into account the offline factors faced by youth and also provide novel education strategies (e.g., social media simulation) for them to learn how to handle risky situations in a safe setting.

"*kind of gives them this place to kind of get hands-on experience to try out what it's like to be on a social media feed to you know get some experience on that...And so it's the same idea of those like virtual driving simulators you know, like it's not real you can go into the street you're not going to get hurt.*" - P2, Researcher

In discussing human-centered solutions to youth online safety, Educators/clinicians shared that getting parents and guardians to take more strength-based approaches is the solution. In this way, *empowerment strategies* targeting parents came out as the

major solution method. They shared that technology if it is all used, should provide talking points to parents on areas of interest of youth or areas where youth have issues so discussions can take place.

"*The challenge is to have parents really fully grasp the long-term implications of a child sending an inappropriate picture or starting to engage in a conversation with a 45-year-old man who has nefarious intentions for their child.*" - P28, Educator

4.2.4 Leveraging the state-of-art AI solutions

The fourth major solution theme put forward by participants was to use state-of-the-art technology to mitigate issues and keep updating the technology to remain current. In discussing this, they talked about datasets being used in AI solutions being kept current, as well as experts being involved in the modeling and prediction stages of those technologies. Entrepreneurs who contributed to this theme stated that concerns being raised in society about data access, transparency, and bias in AI solutions needed to be addressed by using *participatory design* to create products with individuals from the prospective target market. This was shared as a solution to especially address automated solutions which process text shared by youth online as the banter between youth can often be misinterpreted by adults.

"*There has to be validation of the data right... if the information is validated by a group of subject matter experts that are not necessarily technical people.*" - P32, Entrepreneur

IT Professionals mostly discussed being very transparent about what data is collected, how that data will be used, and who will have access to the data as a solution to the challenges. This was shared from the perspective of quelling any fears customers would have about sharing their data, as well as showing the personalized benefits they could gain by sharing data. One participant went as far as to state that data usage assertions should be embedded in the solutions themselves.

"*...every place where you can introduce some form of check or verification that you are collecting the data for the purpose that you said you were, that it's the right shape and the right reasoning, you should encourage that from a technology platform perspective.*" - P22, IT Professional

Researchers also shared transparency as a solution under the theme of leveraging state-of-the-art AI. While recognizing that how data is going to be used needs to be clearly stated, researchers also shared that the potential *risks* involved with sharing the data also needs to be clearly outlined to individuals so that they get a fuller picture of solutions. In this regard, they said individuals should be given a chance to opt-in or opt-out.

"*There needs to be much clear messaging around what it is, why it's being used, and not just the benefits, but the potential risks and allowing parents to opt-in for*

that. - P5, Researcher

Finally, Educators/clinicians shared transparency as a solution from the perspective of technology providers making provisions to end-users about how much data is being used to generate any predictions. This was seen as a way of helping end-users to understand how youth online safety solutions work and how their data is being used.

"*I think when we have a good understanding of the inputs that go into it and with all the biases that are included in it, you have a much better sense of like what you can do, on the other end*" - P4, Educator

5 Discussion

5.1 Sociotechnical Approaches Towards Promoting Youth Online Safety

In this study, we found that many platforms provide parental control software as tools for *parents* to monitor their own teens; however, our participants shared various challenges related to parental control software, focusing on teen privacy and family relations. Previous literature points to various reasons for this difficulty such as parents' struggle to keep up with new emerging technologies to understand and monitor their teens' digital interactions [15], disagreement between parents and teens on what types of content should be monitored [59], and the fluid nature of boundaries to teens' privacy rights [15]. Our findings add empirical evidence to the literature by highlighting the tension between teens and parents in that parental control software can indeed exacerbate parent-teen relationships due to a mismatch in privacy perceptions between the two. At the same time, in contrast to existing literature, we have found that while some teens and parents find online safety tools invasive and restrictive, others do not. The difference was rooted in the differing level of privacy risk exposure and various parenting styles to approach teens' online safety. This was well addressed as the challenges for entrepreneurs when trying to provide online safety solutions that can cater to a wide range of customers. The two conflicting views regarding restrictive parental monitoring bring us to the open questions of whether or not privacy should be maintained at all costs, what privacy actually means in the context of keeping teens safe online, and whether or not the data can be handled in a meaningful way if privacy is maintained, all of which can be further explored in future work. We also acknowledge the positive role of parental oversight if managed well, especially for early teens [60]. Yet, such approaches have been examined from the perspectives of teens and parents, leaving out the most vulnerable teens (e.g., foster youth) who do not have parents who can actively engage in ensuring their online safety. Hence, there is a need for a paradigm shift where teen online safety becomes a shared responsibility for all including those who design and develop online platforms serving teens.

In RQ1, we observed that our participants had strong opinions toward big tech companies for being irresponsible in their social responsibilities to address youth online safety issues. During the interviews, some educators almost demonized the big-tech

industry with the analogy of *"those who pollute the water should bear social responsibility"* (P3, Educator) or *"those who give a rat's ass about our kids"* (P28, Educator). More importantly, it resonated with us when one participant stated that if no one takes accountability, *"then it does really fall on the kids to figure it out"* (P27, Educator). This antagonism towards big tech is not new given multiple claims that they fail to provide guardrails to protect teens from such risks [61–63]. Our participants' view toward big tech adds voice to this social criticism, especially the perspectives of domain experts who are actively working on creating online safety interventions. Their antagonisms toward the closed and proprietary nature of the big tech industry call for more collaborative efforts toward the goal of promoting teen online safety. As such, our findings once again point to a need for sociotechnical approaches that engage different stakeholders including the big-tech industry, instead of just making the matter of youth online safety the parents' sole responsibility.

5.2 Mapping Sociotechnical Challenges to Potential Solutions

In RQ2, when participants were asked for possible solutions to the challenges they presented, they often provided solutions that were also identified as, or part of, the challenges they or other participants identified. For instance, AI solutions not having nuanced predictions or taking broad action due to the lack of contextualized training data was an identified challenge. A mitigating solution to this challenge was the collection and use of contextualized data. However, the collection and use of fine-grained youth data is also a challenge from the perspectives of ethics and privacy, as well as determining and managing levels of access to youth data. The challenges and solutions were also compounded by the issue of having to cater to different policies when dealing with data across international borders such as data of European citizens vs. residents of the U.S. State of California [64–66].

In turn, a solution offered by participants to combat many of the challenges was to have more transparency in technology products such that end-users are informed on specific data collection and processing procedures. Recently, there has been a push towards having more transparency in sociotechnical systems (e.g., transparent algorithms and explainable AI) [55]. Our findings add voice to the current movement toward transparent systems, particularly in the youth online safety domain by providing major stakeholders' views on this issue. While there are efforts to create shared databanks that allow users to retain control of their data with options to select who has access, how the data is used, and for how long, these solutions are still in their infancy or are not widely used outside of highly contextualized cases. This means that currently, the issues that have been identified, much of which are about teen data use and privacy, are unsolved. We discuss this further in the following section.

5.3 Balancing Protection and Privacy

One of the major challenges that emerged in our study was the trade-offs between the protection of youth with the use of youth's personal data and the protection of teen privacy. The questions about surveillance, privacy, and safety are not new. Scholars, policymakers, security professionals, and advocates have continuously discussed

the effects of surveillance/monitoring technologies on individuals and society altogether [67]. In contrast to the European Union, where privacy is a right and robust data protection laws have been enacted (e.g., EU Data Protection Law [65]), privacy regulation in the U.S. is sectorial, with separate laws for different types of information, users, and situations [68]. This has led to more complexities and controversies in developing legal frameworks to promote youth online safety within the U.S.

In 2022, a comprehensive bipartisan legislation called the Kids Online Safety Act (KOSA) has been proposed by U.S. Senators [69]. The legislation requires social media platforms to proactively mitigate harm to minors such as the promotion of self-harm, suicide, and sexual exploitation. It also requires independent audits and supports public scrutiny from experts and academic researchers to ensure that social media platforms are taking meaningful steps to address risks to kids [69]. Yet, teen privacy advocates fear that KOSA would incentivize social media sites to collect *even more information about children* to prevent a set of harms to minors. More importantly, the advocates argued that KOSA can effectively be an instruction for social media platforms to employ a broad range of content filtering to limit minors' access to certain online content such as sex education for LGBTQ youth (which schools had implemented in response to earlier legislation) [70].

Given our participants' claims about big tech's repeated failure to protect youth from serious risks on their platforms, we embrace the idea of developing legal frameworks to mitigate youth online risk. At the same time, privacy laws will be necessary as safeguards from the excessive use of teens' personal data especially when sociotechnical solutions (e.g., AI-based risk detection) rely heavily on teens' intimate data. Hence, to achieve the balance between privacy and protection, we urge teen privacy issues to be part of a larger policy agenda, and advocacy organizations, educators, parents, scholars, and youth need to work together as part of a broad, social discourse.

5.4 Implications for Design

Based on our findings, we suggest design approaches that can empower teens and parents to manage online risks in meaningful ways. First, we suggest providing co-monitoring systems that can facilitate parent-teen communication. For instance, we can design co-monitoring applications where teens and parents collaboratively monitor their online interactions. Or we design a co-learning prototype in which parents and teens together engage in learning social media use. Given that mutually monitoring one another's apps installed on the mobile device helped parents and teens to increase transparency and hence facilitate communication [19], we believe that providing co-monitoring systems can be a way to support teens, parents, and family as a whole.

Second, we suggest strength-based design practices that can support teens in the long term. Given that teens often do not disclose their online risk experiences because they are concerned about punishment, we should consider supporting teens in a way that helps them cope with and recover from online risk exposure. More importantly, parent- and family-focused solutions can leave more vulnerable youth (e.g., foster youth) to online risks [42, 43]. Therefore, the long-term goal of solutions for teen online safety should be to empower them to be more resilient to online risks they can encounter. For instance, as one of our participants mentioned, we can consider

providing teens with simulated environments (e.g., simulated social media platforms) before they are exposed to the real online world to allow them to learn potential risk scenarios and how to cope with them. We can also design monitoring/nudging systems that highlight and reward teens for positive activities (e.g., posting encouraging comments), rather than just alerting and removing risky content.

In addition, we highlight "human-in-the-loop" approaches to designing sociotechnical solutions to promote youth online risk. Given that risk is highly subjective, nuanced, and fluid, it is almost impossible for risk detection systems to capture all risk scenarios without reflecting human interpretation. To mitigate these challenges, we suggest participatory design where parents and teens are part of the design process of the systems. With this approach, designers and developers can reflect on the unique perspectives of users to identify subjective and contextualized online risks. Finally, instead of designing online safety systems that fit all, we recommend implementing an interactive design [71] in which systems are personalized based on parenting style and the context of teens' online usage. For instance, personas can be used to help designers and developers to understand how best to support the needs of parents and teens with various expectations. The idea of interactive design can be applied particularly to privacy settings in the system. The options to opt in/out of various privacy features would allow parents and teens more control over their privacy.

5.5 Implications for Practice

While companies are within their rights to keep computer code and processing methods proprietary in nature, our findings suggest that more can be done for public awareness on how data is going to be used within those proprietary methods. Even when youth characterize themselves as being competent in Math and programming, they still struggle to understand AI systems and how their data is used and need explanatory support [72]. Hence, we argue for increased transparency in AI risk detection systems. We suggest this as a social movement away from the "black box" and "proprietary" framing to more "open" algorithms where sources are shared and data usage is explicitly stated and made publicly available. This will help promote building trust from users and hence, eventually contribute to sustainable plans to keep teens safe.

To make promoting teens' online safety a sustainable agenda, proper legislation and policy should be in place. Our participants reassured us that the lack of regulation in the U.S. has led to more complexities and controversies in developing sustainable frameworks to promote youth online safety within the U.S. The only exception is the Children's Online Privacy Protection Act (COPPA) [73] which applies primarily to commercial websites. COPPA provided a legal ground for the companies to protect children under 13 by adding safety measures specifically designed to address a wide range of practices on social media, mobile, and other platforms. Such safeguards, however, do not apply to youth over 13, who have been largely left out of public policy debates and self-regulatory industry programs [68]. Hence, we urge proper legislation and policy (which include a wide range of youth populations) to be discussed as part of a larger legal agenda including developers, clinicians, educators, scholars, organizations, parents, and youth involved in the discourse.

In summary, our stance is that it takes a whole village to keep teens safe from online risks. The protection of teen privacy and the promotion of teen online safety via sociotechnical systems is only possible when a larger community of policymakers, advocacy organizations, educators, parents, scholars, and youth work together as part of a broad social discourse. To mitigate this paradox of teen online privacy, we call for a collaborative effort among practitioners and researchers to develop evidence-based sociotechnical tools to keep teens safe in a way that respects their privacy.

5.6 Limitations and Future Work

We note some of the limitations of our study. First, we do not assume that our findings are representatives of all stakeholders' views. Participants in our study are not representatives of all experts in the youth online safety domain. For instance, experts from big-tech companies were not well represented in the participant listing. However, our recruitment strategies (i.e., referring other stakeholders by the end of the interview) allowed us to reach hard-to-reach populations such as participants who may otherwise decline to participate in studies. In addition, although we engaged in an iterative process where we meaningful themes emerged, there could be potential themes that we missed. Hence, we cannot argue for generalizing our findings.

In this work, we did not explore the effectiveness or user satisfaction of any specific sociotechnical solution. Instead, we asked about their technical approaches to addressing youth online safety issues, challenges associated with their approaches, and potential solutions in their own context. Therefore, based on our design recommendations, future work can develop prototypes to be evaluated by teens, parents, and practitioners for concrete feedback. Finally, during the interview, participants also mentioned non-technical approaches to keep teens safe online such as raising awareness, education, and family communication, which warrant further investigation. Future work can examine how non-technical approaches can help improve technical solutions to empower teens and keep them safe from online risk.

6 Conclusion

Youths face challenges as they use online platforms. While past research and discussion focused significantly on examining different parenting strategies from the perspectives of youth and their parents, the focus must change to also include discussion on sociotechnical solutions from multi-perspectives. Stakeholders agree that sociotechnical solutions are not without challenges. Their views on solutions for the challenges point to a need for inclusive dialogue among would-be competitors, legislators, academics, and end-users. Only through such collaborative discussion can the balance between the needs of youth, parents, and safety experts be met.

Acknowledgments. This research is partially supported by the U.S. National Science Foundation under grants IIP-2329976 and IIS-2333207 and by the William T. Grant Foundation grant #187941. Any opinions, findings, conclusions, or recommendations expressed in this paper do not necessarily reflect the views of the sponsors. We would also like to thank Karla Badillo-Urquiola, Afsaneh Razi, Ashwaq Alsoubai, and Abdulmalik Alluhidan for their assistance in the interviews and data analysis.

Appendix A Semi-Structured Interview Questions

Below you will find the interview questions used for our semi-scripted interviews of participants who identified as entrepreneurs or worked in the Information Technology (IT) sector (A.1) and of non-technical participants who do not work in the IT sector (A.2). The interviews were conducted as part of a project funded by the National Science Foundation (NSF) I-Corp program. For transparency, the NSF I-Corp program was introduced to participants in this study.

A.1 Questions for Entrepreneurs and Technical Participants

Interview Warm-up

1. Introduce yourself
2. Introduce the purpose of the study
3. Ask permission to use for research purposes
4. Ask permission to record

Interview Questions

1. Please tell me a little about the organization you work for?
2. What is your title and role within the organization?
3. Does your platform engage with teen users? If so, in what capacity?

Online Risks

1. Are there any legal or safety concerns you have encountered when catering to youth as end users? If so, what?
2. Has your company had to deal with online safety issues, such as cyberbullying, sexual solicitations, or sharing of inappropriate content among youth users?

Discovery Questions

1. If so, could you please give me a concrete example of something that has happened in the past? How does your company currently handle these issues?
2. Do you think that your current solutions are adequate? Why or why not?

Technology Interview

1. If not, what else could be done to handle these issues? For instance, could you think of a new product or service to solve these problems?
2. Does your company have the in-house resources to build these solutions yourself?

Closing Remarks

1. Is there any question that I didn't already ask that you think would be useful for me to add to future interviews? If so, what would it be, how would you answer it, and why do you think this question would be valuable to ask?
2. Could you please refer to me at least three other people who I would benefit interviewing about this project? I really appreciate your time and insights and would love to keep in touch with you as we work on this project.

A.2 Questions for Non-technical Participants

Interview Warm-up

1. Introduce yourself
2. Introduce the purpose of I-Corps
3. Ask permission to use for research purposes
4. Ask permission to record

Interview Questions

1. Please tell me a little about your organization?
2. What is your title and role?
3. In what capacity do you work with teens?
4. Approximately how many teens does your organization work with in a given year. If they are in a local system, as how many teens are engaged in this system within the entire U.S.
5. In relation to how you engage with teens in your job, what are your key performance metrics for success?
6. Which of these metrics are you most concerned about?
7. What role do you think online risks experiences of these youth impact these goals?
8. How many teens that you work with have experienced significant online risks?
9. What online risks are the most concerning or considered as the biggest problem?
10. How do you currently find out if a teen is experiencing these types of online risks?
11. What are the limitations of this approach?

Technology Interview

1. What else could be done to handle these issues? For instance, could you think of a new product or service to solve these problems?

Closing Remarks

1. Is there any question that I didn't already ask that you think would be useful for me to add to future interviews? If so, what would it be, how would you answer it, and why do you think this question would be valuable to ask?
2. Could you please refer to me at least three other people who I would benefit interviewing about this project? I really appreciate your time and insights and would love to keep in touch with you as we work on this project.

Appendix B Participant Characteristics

Table B1 Participant characteristics

Participant ID	Gender	Role	Job Responsibilities
1	M	IT professional	Technical Leader
2	M	Researcher	Assistant professor
3	M	Educator	Psychologist
4	M	Educator	YouTuber/Consultant
5	F	Researcher	Senior research scientist
6	M	Entrepreneur	Business owner
7	M	IT professional	Technical Leader
8	F	Researcher	Senior fellow
9	F	Researcher	Senior faculty
10	M	Entrepreneur	Business owner
11	F	Researcher	Assistant professor
12	F	Educator	Consultant
13	M	Entrepreneur	Business owner
14	M	Researcher	Director of youth research firm
15	M	Entrepreneur	Business owner
16	F	Researcher	Senior faculty
17	M	IT professional	Technical Leader
18	M	IT professional	Security professional
19	M	IT professional	Technical Leader
20	M	Entrepreneur	Business owner
21	M	Entrepreneur	Business owner
22	F	IT professional	Technical Leader
23	M	Entrepreneur	Business owner
24	F	IT professional	Technical Leader
25	M	Entrepreneur	Business owner
26	M	Entrepreneur	Business owner
27	F	Educator	Technical Leader
28	F	Educator	Business owner
29	M	Entrepreneur	Business owner
30	M	IT professional	Technical Leader
31	M	Educator	Psychologist
32	M	Entrepreneur	Business owner
33	F	Educator	Business owner
34	F	Entrepreneur	Business owner
35	F	Entrepreneur	Business owner
36	M	IT professional	Technical Leader
37	F	Researcher	Assistant Professor
38	F	Educator	Pediatrician

Appendix C Codebooks

Table C2 Participant identified sociotechnical approaches (RQ1)

Sociotechnical Approach	**Quote**
Online risk detection with AI/ML	*"What we are trying to do is use a combination of technology, so that would be machine learning and other advanced technologies like that, advanced data analysis and human intervention to help parents understand when and if their kids may be struggling with mental health issues."*
Other security/safety tools	*"I help them be very intentional about how they design their products so they design their products with the concept of safety by design to make sure that they have the product practices this the processes and the tools to ensure a healthy and safe online community."*
Parental control software	*"What we do is we give tools for parents in order to monitor their own kids in their own families activities so parents can set alerts based on type of content being accessed."*

Table C3 Emerging themes for RQ1 - Online Safety Challenges

Theme	Code	Quote
Trade-offs between Privacy and Protection	Teen privacy	*"But yeah it's a big challenge, I would say, like the balance of protection and privacy, because, in order to protect your kids you do have to have some of that privacy."*
	Legal liability	*"The company takes no risks with it right so like if it's found it's gone, it's reported... we're legally responsible for doing something about them."*
Tensions around Diffused Accountability	Demonizing big tech	*"companies could give a rat's ass about our kids. All they care about is making money. "*
	Membership	*"And where we see most of the issues and things come from is around that sense of responsibility and accountability. The biggest reason for the bystander effect is that diffuse means of responsibility"*
1-3 Risk is subjective	Accuracy	*"I'm not sure how much nuance you'd be able to get out of monitoring AI because a lot of like child predators they're not direct about it"*
	Risk variability	*"as soon as you come up with a defense, the people you're defending against are going to come by attacking it to come up with a way to breach it."*
Lack of proprietary but necessary data	Data access	*"This is a challenge, and the big challenge in this is that gaming systems are closed and their proprietary."*
Technical solutions are costly	Cost	*"all you have to do is stand up the system to accept it, it's going to be a ton of data so keep that in mind your costs are going to be astronomic ."*

Table C4 Emerging themes for RQ2 - Identified Solutions

Theme	Code	Quote
Centralized and Multi-disciplinary Collaboration	Partnership	*"Keeping everybody together…communities that is behind you and has your back and is going to support you and you know help, I would say, is the best approach."*
	Source sharing	*"being an open collaboration… it's totally open and so those opinions can come from anywhere, as opposed to select leaders from different product areas"*
Leveraging the state-of-art AI	Ethical data stewardship	*"we also do not store anything about the child…there's no information about the child, we don't even know their name."*
	Participatory Design	*"really important to have young people involved in whatever you're developing…I think the actual end users, even if they're on the team like they should have input."*
	Transparency	*"There needs to be much clear messaging around what it (data) is and why it's being used that."*
	Validation	*"explainable AI, understanding that the data that's coming out of that has been validated and a lot of times that's generally from human beings"*
Prioritize Human-centered Approaches	Ease of use	*"I would want them to have something that was easy simple reliable engaging all of those things that they want to use themselves and would help them stay on track."*
	Empowerment	*"our marketing is more about empowerment, we say parents can buy this for the kids so they are in better digital skills and learn how to manage their time online."*
Sustainable plans to keep teens safe online	Self-sufficiency	*"What you're offering should be a technology play to fill a gap somewhere or should you offer more holistic solution."*

References

[1] Vogels, E.A., Gelles-Watnick, R., Massarat, N.: Teens, social media and technology 2022. Pew Research Center (2022). https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/

[2] Anderson, M., Jiang, J., *et al.*: Teens, social media & technology 2018. Pew Research Center **31**(2018), 1673–1689 (2018)

[3] Wisniewski, P.J., Vitak, J., Hartikainen, H.: Privacy in adolescence. In: Modern Socio-Technical Perspectives on Privacy, pp. 315–336. Springer, ??? (2022)

[4] Vogels, E.A.: Teens and cyberbullying 2022. Pew Research Center (2022). https://www.pewresearch.org/internet/2022/12/15/teens-and-cyberbullying-2022/

[5] Bányai, F., Zsila, Á., Király, O., Maraz, A., Elekes, Z., Griffiths, M.D., Andreassen, C.S., Demetrovics, Z.: Problematic social media use: Results from a large-scale nationally representative adolescent sample. PloS one **12**(1), 0169839 (2017)

[6] Finkelhor, D., Walsh, K., Jones, L., Mitchell, K., Collier, A.: Youth internet safety education: Aligning programs with the evidence base. Trauma, violence, & abuse **22**(5), 1233–1247 (2021)

[7] Kowalski, R.M., Limber, S.P., McCord, A.: A developmental approach to cyberbullying: Prevalence and protective factors. Aggression and Violent Behavior **45**, 20–32 (2019)

[8] Madigan, S., Villani, V., Azzopardi, C., Laut, D., Smith, T., Temple, J.R., Browne, D., Dimitropoulos, G.: The prevalence of unwanted online sexual exposure and solicitation among youth: A meta-analysis. Journal of Adolescent Health **63**(2), 133–141 (2018)

[9] Bostrom, R.P., Gupta, S., Thomas, D.: A meta-theory for understanding information systems within sociotechnical systems. Journal of Management Information Systems **26**(1), 17–48 (2009) https://doi.org/10.2753/MIS0742-1222260102 https://doi.org/10.2753/MIS0742-1222260102

[10] Ghosh, A.K., Badillo-Urquiola, K., Guha, S., LaViola Jr, J.J., Wisniewski, P.J.: Safety vs. surveillance: what children have to say about mobile apps for parental control. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2018)

[11] Ghosh, A.K., Badillo-Urquiola, K., Rosson, M.B., Xu, H., Carroll, J.M., Wisniewski, P.J.: A matter of control or safety? examining parental use of technical monitoring apps on teens' mobile devices. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2018)

[12] Khurana, A., Bleakley, A., Jordan, A.B., Romer, D.: The protective effects of parental monitoring and internet restriction on adolescents' risk of online harassment. Journal of youth and Adolescence **44**, 1039–1047 (2015)

[13] Schiano, D.J., Burg, C.: Parental controls: Oxymoron and design opportunity. In: HCI International 2017–Posters' Extended Abstracts: 19th International Conference, HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings, Part II 19, pp. 645–652 (2017). Springer

[14] Wisniewski, P., Ghosh, A.K., Xu, H., Rosson, M.B., Carroll, J.M.: Parental control vs. teen self-regulation: Is there a middle ground for mobile online safety? In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 51–69 (2017)

[15] Cranor, L.F., Durity, A.L., Marsh, A., Ur, B.: Parents' and teens' perspectives on privacy in a technology-filled world. In: Proceedings of the Tenth USENIX Conference on Usable Privacy and Security. SOUPS '14, pp. 19–35. USENIX Association, USA (2014)

[16] Wisniewski, P.: The privacy paradox of adolescent online safety: A matter of risk prevention or risk resilience? IEEE Security & Privacy **16**(2), 86–90 (2018)

[17] Erickson, L.B., Wisniewski, P., Xu, H., Carroll, J.M., Rosson, M.B., Perkins, D.F.: The boundaries between: Parental involvement in a teen's online world. Journal of the Association for Information Science and Technology **67**(6), 1384–1403 (2016)

[18] Wisniewski, P., Xu, H., Rosson, M.B., Carroll, J.M.: Parents just don't understand: Why teens don't talk to parents about their online risk experiences. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 523–540 (2017)

[19] Akter, M., Godfrey, A.J., Kropczynski, J., Lipford, H.R., Wisniewski, P.J.: 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), 1–28 (2022)

[20] Ghosh, A.K., Hughes, C.E., Wisniewski, P.J.: Circle of trust: a new approach to mobile online safety for families. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2020)

[21] Nouwen, M., JafariNaimi, N., Zaman, B.: Parental controls: reimagining technologies for parent-child interaction. In: Proceedings of 15th European Conference on Computer-Supported Cooperative Work-Exploratory Papers, vol. 2017, pp. 18–34 (2017). European Society for Socially Embedded Technologies (EUSSET)

[22] Ko, M., Choi, S., Yang, S., Lee, J., Lee, U.: Familync: facilitating participatory parental mediation of adolescents' smartphone use. In: Proceedings of the 2015

ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 867–878 (2015)

[23] Maghsoudi, R., Shapka, J., Wisniewski, P.: Examining how online risk exposure and online social capital influence adolescent psychological stress. Computers in Human Behavior **113**, 106488 (2020) https://doi.org/10.1016/j.chb.2020.106488

[24] De Santisteban, P., Gámez-Guadix, M.: Prevalence and risk factors among minors for online sexual solicitations and interactions with adults. The Journal of Sex Research **55**(7), 939–950 (2018)

[25] Reed, L.A., Boyer, M.P., Meskunas, H., Tolman, R.M., Ward, L.M.: How do adolescents experience sexting in dating relationships? motivations to sext and responses to sexting requests from dating partners. Children and Youth Services Review **109**, 104696 (2020)

[26] Pinter, A.T., Wisniewski, P.J., Xu, H., Rosson, M.B., Caroll, J.M.: Adolescent online safety: Moving beyond formative evaluations to designing solutions for the future. In: Proceedings of the 2017 Conference on Interaction Design and Children. IDC '17, pp. 352–357. Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3078072.3079722 . https://doi.org/10.1145/3078072.3079722

[27] Livingstone, S., Smith, P.K.: Annual research review: Harms experienced by child users of online and mobile technologies: The nature, prevalence and management of sexual and aggressive risks in the digital age. Journal of child psychology and psychiatry **55**(6), 635–654 (2014)

[28] Wisniewski, P., Xu, H., Rosson, M.B., Perkins, D.F., Carroll, J.M.: Dear diary: Teens reflect on their weekly online risk experiences. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 3919–3930 (2016)

[29] Valkenburg, P.M., Meier, A., Beyens, I.: Social media use and its impact on adolescent mental health: An umbrella review of the evidence. Current opinion in psychology **44**, 58–68 (2022)

[30] Schønning, V., Hjetland, G.J., Aarø, L.E., Skogen, J.C.: Social media use and mental health and well-being among adolescents–a scoping review. Frontiers in psychology **11**, 1949 (2020)

[31] Modecki, K.L., Goldberg, R.E., Wisniewski, P., Orben, A.: What is digital parenting? a systematic review of past measurement and blueprint for the future. Perspectives on Psychological Science **17**(6), 1673–1691 (2022) https://doi.org/10.1177/17456916211072458 https://doi.org/10.1177/17456916211072458. PMID: 35816673

[32] Kim, S., Razi, A., Stringhini, G., Wisniewski, P.J., De Choudhury, M.: A

human-centered systematic literature review of cyberbullying detection algorithms. Proceedings of the ACM on Human-Computer Interaction **5**(CSCW2), 1–34 (2021)

[33] Razi, A., Kim, S., Alsoubai, A., Stringhini, G., Solorio, T., De Choudhury, M., Wisniewski, P.J.: A human-centered systematic literature review of the computational approaches for online sexual risk detection. Proceedings of the ACM on Human-Computer Interaction **5**(CSCW2), 1–38 (2021)

[34] Park, J., Gracie, J., Alsoubai, A., Stringhini, G., Singh, V., Wisniewski, P.: Towards automated detection of risky images shared by youth on social media. In: Companion Proceedings of the ACM Web Conference 2023, pp. 1348–1357 (2023)

[35] Atoum, J.O.: Detecting cyberbullying from tweets through machine learning techniques with sentiment analysis. In: Future of Information and Communication Conference, pp. 25–38 (2023). Springer

[36] Cohen-Almagor, R., Trottier, D.: Internet crime enabling: Stalking and cyberstalking. In: Advances in Information and Communication: Proceedings of the 2022 Future of Information and Communication Conference (FICC), Volume 2, pp. 843–859 (2022). Springer

[37] Alsoubai, A., Song, J., Razi, A., Naher, N., De Choudhury, M., Wisniewski, P.J.: From 'friends with benefits' to 'sextortion:' a nuanced investigation of adolescents' online sexual risk experiences. Proc. ACM Hum.-Comput. Interact. **6**(CSCW2) (2022) https://doi.org/10.1145/3555136

[38] Park, J., Lediaeva, I., Godfrey, A., Lopez, M., Madathil, K.C., Zinzow, H., Wisniewski, P.: How affordances and social norms shape the discussion of harmful social media challenges on reddit. Human Factors in Healthcare, 100042 (2023)

[39] Bottino, S.M.B., Bottino, C., Regina, C.G., Correia, A.V.L., Ribeiro, W.S.: Cyberbullying and adolescent mental health: systematic review. Cadernos de saude publica **31**, 463–475 (2015)

[40] Wisniewski, P., Jia, H., Xu, H., Rosson, M.B., Carroll, J.M.: "preventative" vs. "reactive": How parental mediation influences teens' social media privacy behaviors. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work amp; Social Computing. CSCW '15, pp. 302–316. Association for Computing Machinery, New York, NY, USA (2015). https://doi.org/10.1145/2675133.2675293 . https://doi.org/10.1145/2675133.2675293

[41] Rutkowski, T.L., Hartikainen, H., Richards, K.E., Wisniewski, P.J.: Family communication: Examining the differing perceptions of parents and teens regarding online safety communication. Proceedings of the ACM on Human-Computer Interaction **5**(CSCW2), 1–23 (2021)

[42] Badillo-Urquiola, K.A., Ghosh, A.K., Wisniewski, P.: Understanding the unique online challenges faced by teens in the foster care system. In: Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 139–142 (2017)

[43] Badillo-Urquiola, K., Harpin, S., Wisniewski, P.: Abandoned but not forgotten: Providing access while protecting foster youth from online risks. In: Proceedings of the 2017 Conference on Interaction Design and Children, pp. 17–26 (2017)

[44] Len-Ríos, M.E., Hughes, H.E., McKee, L.G., Young, H.N.: Early adolescents as publics: A national survey of teens with social media accounts, their media use preferences, parental mediation, and perceived internet literacy. Public Relations Review **42**(1), 101–108 (2016) https://doi.org/10.1016/j.pubrev.2015.10.003

[45] boyd, d., Hargittai, E.: Connected and concerned: Variation in parents' online safety concerns. Policy & Internet **5**(3), 245–269 (2013) https://doi.org/10.1002/1944-2866.POI332 https://onlinelibrary.wiley.com/doi/pdf/10.1002/1944-2866.POI332

[46] Apthorpe, N., Varghese, S., Feamster, N.: Evaluating the contextual integrity of privacy regulation: Parents' iot toy privacy norms versus coppa. In: Proceedings of the 28th USENIX Conference on Security Symposium. SEC'19, pp. 123–140. USENIX Association, USA (2019)

[47] Kyakulumbye, S., Pather, S.: Digital and big data initiatives for smart cities in developing countries: A socio-technical view for developing city contexts. In: Advances in Information and Communication: Proceedings of the 2022 Future of Information and Communication Conference (FICC), Volume 2, pp. 53–73 (2022). Springer

[48] Baumrind, D.: Patterns of parental authority and adolescent autonomy. New directions for child and adolescent development **2005**(108), 61–69 (2005)

[49] Williams, A.: Adolescents' relationships with parents. Journal of language and social psychology **22**(1), 58–65 (2003)

[50] Balakrishnan, V., Khan, S., Arabnia, H.R.: Improving cyberbullying detection using twitter users' psychological features and machine learning. Computers & Security **90**, 101710 (2020)

[51] Vishwamitra, N., Hu, H., Luo, F., Cheng, L.: Towards understanding and detecting cyberbullying in real-world images. In: 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) (2021)

[52] Gangwar, A., González-Castro, V., Alegre, E., Fidalgo, E.: Attm-cnn: Attention and metric learning based cnn for pornography, age and child sexual abuse (csa) detection in images. Neurocomputing **445**, 81–104 (2021)

[53] Vitorino, P., Avila, S., Perez, M., Rocha, A.: Leveraging deep neural networks to fight child pornography in the age of social media. Journal of Visual Communication and Image Representation **50**, 303–313 (2018)

[54] Biswas, R., González-Castro, V., Fidalgo, E., Chaves, D.: Boosting child abuse victim identification in forensic tools with hashing techniques. V Jornadas Nacionales de Investigación en Ciberseguridad **1**, 344–345 (2019)

[55] Caddle, X.V., Naher, N., Miller, Z.P., Badillo-Urquiola, K., Wisniewski, P.J.: Duty to respond: The challenges social service providers face when charged with keeping youth safe online. Proc. ACM Hum.-Comput. Interact. **7**(GROUP) (2022) https://doi.org/10.1145/3567556

[56] Malik, J.A.N.: Us administration makes efforts to increase inclusiveness in stem industries: www.whitehouse.gov/demo-day. MRS Bulletin **41**(5), 355–356 (2016) https://doi.org/10.1557/mrs.2016.103

[57] Braun, V., Clarke, V., Hayfield, N., Terry, G.: Thematic Analysis, pp. 843–860. Springer, Singapore (2019). https://doi.org/10.1007/978-981-10-5251-4_103 . https://doi.org/10.1007/978-981-10-5251-4_103

[58] Susskind, L.E., McKearnen, S., Thomas-Lamar, J.: The Consensus Building Handbook: A Comprehensive Guide to Reaching Agreement. Sage publications, ??? (1999)

[59] Moser, C., Chen, T., Schoenebeck, S.Y.: Parents? and children? s preferences about parents sharing about children on social media. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 5221–5225 (2017)

[60] Cox, M.J., Gabrielli, J., Janssen, T., Jackson, K.M.: Parental restriction of movie viewing prospectively predicts adolescent alcohol and marijuana initiation: Implications for media literacy programs. Prevention Science **19**, 914–926 (2018)

[61] Chaudhary, N., Vasan, N.: 3 Ways for Big Tech to Protect Teens From Harm. Wired (2020). https://www.wired.com/story/opinion-3-ways-for-big-tech-to-protect-teens-from-harm/

[62] Fung, B.: Senators blast Big Tech companies over kids' safety amid renewed push for legislation. CNN (2023). https://www.cnn.com/2023/02/14/tech/senate-online-kids-safety/index.html

[63] Journal, T.W.S.: the facebook files A Wall Street Journal investigation. The Wall Street Journal (2023). https://www.wsj.com/articles/the-facebook-files-11631713039

[64] Geyser, W.: GDPR and Social Media: What Data Protection and Privacy Mean for Social Media Marketers. (Accessed on 08/15/2022) (2021). https:

//influencermarketinghub.com/gdpr-social-media/ Accessed 2021-05-19

[65] Wolford, B.: What is GDPR, the EU's new data protection law? (Accessed on 08/15/2022) (2018). https://gdpr.eu/what-is-gdpr/ Accessed 2018-11-07

[66] Attorney General, O.: California Consumer Privacy Act (CCPA). (Accessed on 08/15/2022) (2022). https://oag.ca.gov/privacy/ccpa Accessed 2022-03-28

[67] Čas, J., Bellanova, R., Burgess, J.P., Friedewald, M., Peissl, W.: Introduction: Surveillance, privacy and security. In: Surveillance, Privacy and Security, pp. 1–12. Routledge, New York (2017)

[68] Montgomery, K.C.: Youth and surveillance in the facebook era: Policy interventions and social implications. Telecommunications Policy **39**(9), 771–786 (2015)

[69] https://www.blumenthal.senate.gov/: Blumenthal Blackburn Introduce Comprehensive Kids' Online Safety Legislation. https://www.blumenthal.senate.gov/newsroom/press/release/blumenthal-and-blackburn-introduce-comprehensive-kids-online-safety-legislation

[70] Feiner, L.: Kids Online Safety Act May Harm Minors, Civil Society Groups Warn Lawmakers. https://www.cnbc.com/2022/11/28/kids-online-safety-act-may-harm-minors-civil-society-groups-warn.html

[71] Moggridge, B., Atkinson, B.: Designing Interactions vol. 17. MIT press Cambridge, Cambridge, MA (2007)

[72] Greenwald, E., Leitner, M., Wang, N.: Learning artificial intelligence: Insights into how youth encounter and build understanding of ai concepts. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 15526–15533 (2021)

[73] Commission, F.T.: Children's Online Privacy Protection Rule (2013). https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa/ Accessed 2013-01-17