Safety Tech Podcast [Episode 3] – Young and at risk: the challenge of making digital spaces safer for children

Tue, 4/19 11:25AM • 29:32

SUMMARY KEYWORDS

children, safety, people, safer, online, bullying, harms, terms, kids, create, organisation, moderation, impact, reports, platforms, experiences, young, called, tech, developed

SPEAKERS

Marija Manojlovic, Henry Platten, Ben Whitelaw, Alex Holmes, CLIP

Ben Whitelaw 00:00

Welcome to the safety tech podcast brought to you by the safety tech innovation network. My name is Ben Whitelaw, and I'm the founder and editor of everything in moderation, a weekly newsletter dedicated to online safety and content moderation. One in three internet users is a child, and yet very few online spaces are designed specifically for kids. In today's episode, we look at the myriad harms that children are at risk of online and talk to some of the people working to make the web safer for its youngest users. Thanks for joining us. If you're a child in the UK, it's likely that you have been bullied at school at some point. According to the UK Office for National Statistics, half of young people have experienced it higher than many countries around the world. And what's more, one in five children aged 10 to 15 say they had been bullied online.

Alex Holmes 01:08

My name is Alex Holmes and I am from an organisation called the diner vote.

Ben Whitelaw 01:14

The diner Award is an organisation established in memory of Diana, Princess of Wales, to helps young people develop and fulfil their potential.

Alex Holmes 01:22

One of the biggest ways is through our schools programme where we train young people to be ambassadors who learn to stand up for themselves and for others to harmful impacts like bullying online and offline.

CLIP 01:37

I think it was decided by the world before I discovered myself that I was gay. And I was picked on for people used to say really hurtful comments to me about my hair and my parents really any part of me that wasn't stereotypically why I was relentlessly sent malicious text messages by a group of boys on different social media platforms about my looks, they took pictures of me without me knowing they didn't think I belonged in this country.

Ben Whitelaw 02:09
This is a clip from a campaign run by the Diana reward, encouraging young people to speak out about the bullying they faced. Alex himself experienced bullying as a child. However, he's glad it took place during the pre Internet era.

Alex Holmes 02:23
To me, a lot of the bullying was around racism and homophobia. And it was verbal, I think if I did have that online element, it you know, would have been a lot more difficult to handle. And at least I could perhaps walk out the school gates and escape some of that bullying. But for a lot of young people it can feel like it's really hard to escape. And that follows you into your bedroom and into your life.

Ben Whitelaw 02:50
What are the most common forms of cyber bullying that you come across in your work?

Alex Holmes 02:55
Probably things around online fights and harassment and insults. I think secondly followed by denigration and accusations. So actually, for a lot of young people, they might be on the receiving end of rumours or gossip or, or quite serious accusations. And they've been targeted, and the very viral nature of the internet means that quite quickly, more people can find out. So I think that can be quite scary. And I think the other thing for young people is if they if they have protected characteristics, perhaps their race, religion, or even around gender, they can be adversely affected often by online bullying.

Ben Whitelaw 03:33
What role do parents have to play in the kind of mediation between their children and the tools?

Alex Holmes 03:39
Yeah, I think my parents have a really, really difficult job. It is difficult to keep up with the pace of, of technology and the change of apps and culture that goes with it. Some of the things that we face, offline, face to face, some of the challenges can be solved in a similar way through dialogue through conversation. So actually, if you're able to speak to your child and help them understand who their support network is, developing some resilience and some inner strength, so that actually, if some things go wrong, they know what to do they know who to turn to, and having that conversation with them saying, if this happened to you on the internet, what would you do? And then outside of that parents can use resources you some great websites, think about sort of parental locks and controls and so on. We do need to technology companies to do more.

Henry Platten 04:38

One of the most enduring experiences that I've had in working in safety tech for so many years, is that the innovators, the companies, the entrepreneurs, who are all part of safety tech, care deeply about creating a safer environment.

Ben Whitelaw 04:54
Henry Platten is the founder of the UK based safety tech company, go bubble. One of a number of companies that is tackling the growing problem of online bullying,

Henry Platten 05:02
what drives most safety tech companies, certainly drives us, is the wants and the need to make the internet safer. Now, for me personally, obviously, I've got a strong tie back to my time in the police with regard to that. But also, more importantly, I'm a dad. So my two kids here Rocco is five, nearly six, Sophia is nine, nearly 10. I want the internet to be safer for them. And when you look at the global population of children, you look at the open access to devices, how that's getting younger, certainly here in the UK. But the role of digital safety digital citizenship, safeguarding online safety really is one of the critical and pivotal touch points for anyone concerned about a child's welfare.

Ben Whitelaw 05:56
Having seen firsthand the damage that toxic unfiltered content can do. The company developed technology to allow children to communicate safely with one another online.

Henry Platten 06:06
Google really came about because of a problem. And the problem in particular was around practical content moderation and detection. Personally, I was a sergeant in the police. And having seen the role of social media and content, even going back to very early stages of Bebo, MySpace, was aware of how content played a key role, and also the distress and anguish that that could cause fast forward to where we are now. And I'm so blessed to work with the stellar team that we have. Our chair is Patricia kartha Andrez. Ex Global Trust and Safety from Twitter also worked at Google and Facebook. Danielle CEO was one of Twitter's Safety Advisory Panel members, Tim our head of legal was Facebook's first Director of Public Policy. So we have this wonderful combination of industry and law enforcement, you've come together really to look at it through a practical lens in terms of what does tech actually need? And also, more importantly, what does society deserve in terms of content, moderation and detection?

Ben Whitelaw 07:14
So can you tell us a little bit more about how go bubble works, then

Henry Platten 07:18
go bubble works in terms of full multimedia. So we're looking at text emoji, photo, video, audio 21 different languages, very much in terms of context. Content moderation, historically, has always operated on keywords, that unfortunately, just doesn't deliver the impact that society requires and that clients are looking for. The sophistication that can come through in terms of contextual analysis, where you're looking more at the behaviours and the inherent behaviours that can come through. That's where you can pick up on the subtle differences. And we're proud with the clients that we work with, across

sports media, community platforms, dating apps, that what they value. First and foremost is the user experience, and a safer user experience. When you're looking at that kind of deeper contextual machine learning that we can bring to the table. That's where you're actually helping a community to grow and thrive, and to be supportive of each other. It's there as the community enhancement tool to amplify the good and to reduce the bad.

Ben Whitelaw 08:30
I'm really interested in one of the claims on Goba was website about how it can impact kindness and reduce bullying. Can you tell us how you've been able to quantify that

Henry Platten 08:40
one of the important areas for us has been the impact of the technology on individuals and on their personal positive mental health and their outlook. One of the areas where we actually test it out on machine learning, first of all, is that we built a layer on top of it called Go bubble school.

Ben Whitelaw 08:58
Go bubble school is a product that effectively creates a small scale social media network, used exclusively by school staff and students.

Henry Platten 09:07
We were really honoured in terms of seeing the uptake that it had within schools, both pre COVID. And during COVID, with schools in more than 70 Different countries registered. The really interesting thing that came back from parents, from teachers and from head teachers, is that by using go bubble school, and therefore go bubble wrap our content moderation that was powering everything with machine learning, actually, it was delivering through a more positive experience for the young people. We were also able to pass back messaging and feedback where things have been detected by our system and prevented from appearing. We can actually provide advice and feedback to the content creators and the authors as to why that's really important when we're looking at a learning loop that can come through for a community If you're simply blocking and not giving feedback to the author as to why all you're simply doing is perpetuating an issue. And that's one of the key things that we look to try and solve. Now from the teachers, the parents or the head teachers, they told us anecdotally that they'd actually seen an improvement in offline behaviour, and traditional face to face bullying. Because of the interventions that go bubble the machine learning was delivering through global school. They said that the impact that came through providing children with the opportunity to have authentic, safer, healthier and kinder engagement, ahead of starting at secondary school actually impacted them come to September, when they were starting at this new school. They built up friendships that were meaningful, they built up experiences, which were authentic. And that really helped in terms of the transition.

Ben Whitelaw 10:56
How can other platforms used by children prevent cyberbullying,

Henry Platten 11:02
there's a great challenge for platform creators and operators. There's a balancing act between creating environments as a platform which fit your ideology that you want to be able to create and solve and run.

And now very much in terms of the pivotal role of safety tech, have the Children's Commissioner have the correct code in terms of safety by design, for the platforms that are operating in a child space? They have already, I would suggest they're already aware and adopted safety by design principles that exist within the platform. And then it's a case of okay, how can we go further, you know, we can put certain restrictions in place, we can put certain protocols in place that will make it safer to enter. And we'll pick up on some of those bad actors who may pretend to be children to come into that environment. And then it's a case of okay, how do we want our community of young people to interact with each other. There's fantastic work being done by the likes of the British Esports Association, looking at how to address this area, in terms of creating a safer community environment. For example, within eSports

Ben Whitelaw 12:23
The British Esports Association is a not for profit national body, which seeks to promote competitive video gaming in the UK. It reported in 2019, that a whopping 81% of children regularly play video games, and often did so with their friends.

Henry Platten 12:38
So giving the opportunity for young people to come together to connect to collaborate, but how would you do that in a safer way? So there are aspects there of using forms of digital ID and age verification using providers like Yoti. But then they're going one step further and saying, Okay, that's great that we know that a young person is a young person coming into our environment. But then how do we look at the day to day interaction, which again, is very much where content moderation and detection comes through? And using forms of nudge theory? Or how do we want our community to interact? How can we help to further the outlook of eSports in terms of being a world connected, which is what very much one of the drivers of the global Esports Federation, of which I'm proud to be a safeguarding advisor. So for platforms who operate in the child space, we have to look at safety by design, which most, if not all, are doing now. Part of that is around digital identity and age verification, but then it's also having a meaningful form of content moderation, and detection. AI can solve this problem, and we've evidence that it can solve it and solve it at scale. What's always important, as I mentioned earlier, in terms of looking at the experience that we've gone through, and testing out the whole machine learning and why we looked at testing it out within a school environment in particular, is that children's language and language patterns are so dynamic. they fluctuate all over the place. But also they'll they'll try and find gaps in protocols. For example, there are without naming names there are very big, open source gaming platforms within a child sphere who have some elements of content moderation, but there are still ways around it. One of the patterns of behaviour that young people have identified is that if you want to send a bad message, intersperse it with a good message. So in other words, if you want to send an expletive or profanity, send it letter by letter but intersperse a positive message in between each one - still gets through because the moderation isn't using forms of sophisticated fuzzy logic, which are systems that we use where you can actually find those issues, young people are very creative. So I always say try and involve them as much as you possibly can when building out the tech and also implementing the tech.

Ben Whitelaw 15:14

Unfortunately, bullying is just one of many online harms affecting young people today. Marija Manojlovic is the safe online director at the end violence Global Partnership, an organisation that tackles the broad scope of harms that children face in their daily lives.

Marija Manojlovic 15:32
The End Violence partnership is is actually by far I think, the largest Partnership for Children that was ever created. And people don't know that, I'm really proud of that fact. It represents a really a new model of generating change at scale, because it brings together more than 700 organisations under one laser focused mission, which is to prevent all forms of violence against children everywhere. And what we do with our partners is we convene to raise awareness. catalyse leadership commitments with governments and other sectors, but also mobilise new resources promote evidence based solutions to end all forms of violence, abuse and neglect.

Ben Whitelaw 16:08
The safety of children online has become one of the organization's key priorities.

Marija Manojlovic 16:13
what my team does within the organisation is we look into how connectivity and digital technologies impact existing and new forms of child abuse. But we're specifically focusing on the worst form of that abuse, which is the sexual exploitation and abuse. We have a flagship research project called disrupting harm, which is a huge data collection effort across 13 countries in southern and eastern Africa and Southeast Asia, which are basically like mini national threat assessments of how children's experiences look online, what harms they're facing, what law enforcement is doing, what social services are doing and like creates an overall kind of like data landscape of child experiences in specific national contexts. What we can also learn specifically from children's experiences, that is that prevalence is much higher than we expected, especially because this is only reported prevalence. So you can imagine this is just the tip of the iceberg. But for example, in Thailand, 10% of the kids who use internet have reported being actually victimised chemic, suffered child exploitation and abuse online. And these experiences relate to for example, being extorted being blackmailed for sexual acts, or being paid or offer gifts for sexual acts online. This is a high number, it's huge, similar figures we have seen in Kenya, Cambodia, in Uganda and other places that the research was conducted. One thing that I think people don't talk often enough about is the self generated child abuse material. And I think there has been a huge explosion of this imagery. People think about, like, you know, it's explicit imagery or child that appears to have been taken by a child itself. But it can result from both conceptual but also courses of experiences. And we have seen huge increases, for example, IWF noted 168% increase in self generated images, compared to 2020. And there was a huge increase compared to 2019. 21% of kids aged nine to 12, nine to 12, like super young kids, agree that it's normal for people my age to share nudes with each other. Totally normalised. 17% of kids have shared their own nudes with others, out of these kids who have shared 50% have share these images with people they have never met in real life. And it just tells you how normalised and prevalent this behaviour is. What really is then in danger kids is that these initially kind of naive shares of images can result in really, really serious harm. But then the sense of like, shame really prevents them from reporting and are seeking protection and help. And that's a really big issue. And this is something that we've really tried to kind of grapple with right now. And put in front of people to kind of like think about

Ben Whitelaw 18:55
Which type of companies or organisations have you funded and where have you seen impact as a result of that funding.

Marija Manojlovic 19:02
So we have funded a lot, a lot of tools that are used for detection of child abuse material, for example. So we have supported Thorn, because supported development of their classifier for automated detection of child sexual abuse material, or CCM as we call it, especially looking for example to how to improve the accuracy of those classifiers on various skin tones. That's a really critical piece. Or we have supported in probe network of hotlines to develop their automated tools for classification of sea sand reporting. And why is this important is because they receive such high volume of of these reports, and they need to review them sort them - as they're doing that some of the reports may get lost and some of the children can be not saved because we are delaying these processes. So what this classifier is actually doing is helping them give red flags to the most urgent reports and for example, the reports that are like most severe interviews that really help save lives. Another set of tools that you get funded relates to anti grooming tools, so we can support a few interesting tools that are combining linguistics and AI to spot online grooming in real time. So basically, these tools are identifying specific patterns of language or behaviour in predators behaviour that they're using throughout the grooming process. Because sometimes the language is not straightforward. It can sound is like a really loving and nice, you know, friend who's talking to you. And the next thing you know, that friend is asking you for nude or sex related favour. We have also funded a couple of automated chatbots. And I think really interesting one is developed by Internet Watch foundation, it detects potential offenders and refers them to self help programmes. Why this is really important is because you're trying to interrupt the cycle of perpetration where it's starting, basically. And then you want to make sure that you're reducing the demand for child abuse material. Another chatbot that is really interesting is developed in India. And it's using English language, it's called snack AI. This chatbot provides young people with a safe and trusted space for educational information. And what we have learned from this project specifically is, for example, that children don't really care whether they're speaking to a person or a robot, as long as they get the right information. Sometimes they actually prefer speaking to a robot because it's not going to judge them. So this is a really interesting kind of combination of kind of like what we have learned from kids and what their needs are. And then lastly, I'll just mention two educational digital games that we have developed. One I'm super proud of, because it's something that nobody has been working on so far. And it's working with deaf kids International. It's an interactive and accessible digital platform for Deaf Kids to Learn to protect themselves online. And this game is teaching them how to identify warning signals and or inappropriate chat conversations and stuff like that. And the last one is also a really, really interesting is developed by Huddersfield University. And it's an immersive pro social game that is working with young boys. It's called Emilio, and it's in Portuguese now being tested in Brazil. But what this game is doing is really attacking a specific target group that is at risk of perpetration. What people don't usually think about is that perpetration actually starts really early. And the latest research that we also funded shows that 70% of consumers of transsexual abuse material on the dark web started consuming child abuse material when they were younger than 18. 40% said, they were younger than 13 years of age when they started consuming this material. That gives a sense of like, a huge gap between here in prevention efforts that we are missing an entire cohort of kids we can work with to

prevent this abuse in time. So this is why this game is really critical for us, because we believe it's really targeting the right group of kids.

Ben Whitelaw 22:38
I mean, one thing that, that I've been thinking a lot about, and which we've discussed on the podcast is, is whether it's possible to reduce the risk of online harm to children to zero. What would you say to that?

Marija Manojlovic 22:51
But of course, that I would say no, but I think I mean, the Keith or allocating harm to children online and offline is like a huge transformative shifts in the overall ecosystem and the society. But to get there, we have a lot of short term things we can do. And you will often hear analogies that people are making to like road safety. But what I like to talk about is child proofing and apartment because I have a one year old. So like everything to me is now about child proofing. So what you do is basically you look at your apartment and start thinking about what are all the bad things that can happen, right? Sharp edges, glass vases, your favourite Lego Saturn five set or something. And then you're like, make the plan and you're like, Well, I will not wait for my daughter to take you know, a fork and put it in the electricity outlet. I'll put a plug in first, right? So that's all we ask it's doing assessment, see how your digital platform can impact the child and then make sure that you're putting safeguards in place on time. So for example, in my case, will my daughter fall every day and hurt herself? A little bit? Yeah. But will I have reduced a serious harm risk? Yes, of course. These safety by design questions need to be at the heart of the business process of companies because you can't iterate as you go, you can't allow harm to happen. And then oh, then I will deal with it.

Ben Whitelaw 24:04
protecting children online is often used by politicians and policy makers as a significant reason for regulation. How worried do we need to be about some of their intentions, and also the possible risks as far as privacy goes? And the kind of degradation of things like end to end encryption?

Marija Manojlovic 24:26
Yeah, I mean, I think there is like a, maybe rightfully throughout the history, you know, distrust in government intentions. But there are a lot of things that I think governments are trying to do, right. And I think regulation is critical, because with regulation, you're setting principles or standards that everybody should strive towards. And if you set right frameworks around what needs to happen, for example, demanding detection, reporting or child abuse material through a set of really strongly defined principles and rules, I think that can be both respective of privacy and protection. And one thing that I really am trying to kind of like, dismantle in terms of Mists, it's ensuring children's privacy protection is core to their safety. And it's not either or, like you can't think about privacy and safety in exclusive terms, especially when it comes to children. Because keeping children safe online does not mean surveilling them, or others. Because we know that fear based approaches, for example, don't work on kids. If you tell kids like don't do this, or else, they'll certainly do it. So like, that doesn't work. But if you think about how you can work with industry to create this regulatory environment in which they have certainty, they know what's expected from them, you have a sense of how the safe infrastructure can look like, then digital safe spaces can be safer for kids to navigate and explore. We also need to kind of

think about like what governments can do, and industry in terms of enabling parents and educators to give kids the tools to keep themselves safe. I think industry can do much better in terms of also not collecting data from kids not having their personal data being used and collected for commercial purposes is really critical. But I think also like being really clear that once the harm has happened, and for example, images, or videos of child abuse are already online, it is respective of children's privacy to remove those images as fast as possible. So this is also thinking about the privacy of the child and throughout their adulthood, because these victims keep on being reminded of this abuse, through, you know, the ages 30, 40, 50, because these images are not removed. And I think that's also critical to say the privacy is also removal of these photos and detection of these photos. Privacy Protection work best when you give kids agency by creating safer spaces within - you know, industry does that, but also give the necessary tools to protect themselves and create regulatory behavioural institutional protective factors around them. So you can complete the whole system.

Ben Whitelaw 26:53
How is the end violence partnership, trying to kind of shape that regulatory piece then? What kinds of representations are you making in the countries where regulation seems to be coming down the track?

Marija Manojlovic 27:06
Yeah, so we have, actually, I think soon. And I think maybe I'm breaking some rules now. But I will make the announcement, we are soon going to be launching a Global Title non production toolkit. And the toolkit is actually a really practical guide to policymakers on the specific areas of child wellbeing online that they need to tackle to create a really favourable regulatory framework. And this doesn't only relate to harms, but it really relates only to kind of enabling environment around it. So we want kids to thrive online, because it's critical for the development. You will go through this toolkit and you will see what pieces of legislation need to put in place? What are the good practices from around the world that you can adopt? What are the regional and global framework, so kind of like it's a really big piece for us, because we've been working on this for quite some time with a partner that again, I will not announce now. But soon Watch this space, it will be a big, big announcement. But I think we're really trying to give the policymakers no excuse not to do it like you're giving them like you can give us any excuse not to implement this because it's laid down in front of you. And there are things you can do right now to make children more safe online.

Ben Whitelaw 28:14
regulation and safety technology both have a potentially critical role to play in alleviating some of the harms we've heard about in today's episode. But today's experts have also made one thing clear, there are possible pitfalls ahead if young internet users' needs and rights are not taken seriously by platforms and other digital services. We'll be taking a closer look later in this series, at how companies can adopt a proactive approach to safety, and how some of the world's top gaming companies used by millions of children are leading the way. If you'd like to learn more about the ways safety tech is tackling harms faced by children head to the safety tech innovation network, an international network dedicated to the promotion, collaboration, and industrial application of online safety technologies. Become a member to receive the latest information about safety tech events, and reminders about future episodes of the safety tech podcast. Thanks for joining me, and I'll see you next time. This has been a 4kicks production.