“You need to do safety by design and have detection on the backend”: Julie Cordua of Thorn on designing technology that protects children from sexual abuse and exploitation.

Rollo Romig

December 13, 2024

Rollo Romig: Thanks again for spending some time with me today. Could you start by introducing yourself and giving a bit of an overview of your work?

Julie Cordua: I'm Julie Cordua. I'm the CEO of Thorn. We're a non-profit that works to transform the way the world responds to child sexual abuse and exploitation in the digital age. We've been around for about 13 years. If you think back to what the world looked like 13 years ago for children and social media, it was very, very different. Yet at that time, we had already started to see the role that technology was playing in the sexual abuse and exploitation of children.

We started researching it, but then quickly moved into taking on a role to drive technology innovation as part of the solution, and now today, that is really our sweet spot. We still focus on research with children and frontline responders in the field to better understand the trends and the reality of what child sexual abuse in the digital age looks like. We publish [the research] with the goal of making sure that we have an evidence base for this field. Then we actually build solutions to equip frontline responders–mainly press and safety teams at tech companies and

law enforcement–with the tools that they need to address both victim identification and harm identification and removal at scale in this world.

Rollo Romig: Could you go into more detail about some of these technological solutions that you've been working on? What are the main projects?

Julie Cordua: I'll start with what the problem is in each of the areas. On the open web, on tech platforms, you're dealing with an exponential growth in the volume of content or harm scenarios when it comes to child sexual abuse. What does that look like specifically? Image-based abuse, child sexual abuse material, video abuse, grooming, sextortion, live streaming abuse, generative synthetically created sexual abuse material of both [artificially generated "people"] and actual children where you've had generative abuse material made.

The mountain of content is growing. The harm vectors are growing. There is no way that humans can address the sheer scale of this issue. So we asked ourselves, how can you build specialized solutions to detect, report, and remove child sexual abuse material at scale? We have a product we built called Safer that is essentially a specialized content moderation system for the detection of child sexual abuse material on platforms. We can detect images and videos of known child sexual abuse material. That's probably the easiest thing because if it's been seen before, we can find it again and take it down. But over the past few years, we also have built child sexual abuse classifiers, which are machine learning tools that can predict if an image or a video is child sexual abuse.

This [tool] is very important because if you think about it, much of the known abuse material that's circulating on the internet is of children who've already been recovered. It's known because it was part of a case. It was already reported. It went to law enforcement years ago and the child is now recovered. If you are finding new content that has never been seen before, that very well may be a child who's being abused right now.

Adding this tool into our arsenal over the last few years has been incredibly important. We also have text harm detection that we've released that can be very useful for identifying sextortion and grooming on platforms. We also have those classifiers that detect child sexual abuse material. We also have integrated forensic tools for law enforcement. If you think about the entire ecosystem, where tech companies detect child sexual abuse material, take it down, and they're legally required to report it to the National Center for Missing & Exploited Children. Last year alone, more than 100 million files of child sexual abuse material were reported just from the companies that are looking for it.

But that is a tip of a very big iceberg. Of those 100 million files, it's the Center's job to figure out where in the world those files should be distributed to law enforcement to investigate. The downstream effect of good detection on tech platforms is that law enforcement is just flooded with CyberTip reports of images and videos of abuse. They need scaled technology solutions to sift through all of that and find the greatest harm, the most immediate need, so we've integrated the classifiers that we've built into the forensic tools for law enforcement.

When they're dealing with hundreds of thousands of image and video files from a hard drive or a CyberTip, they can use some type of machine learning and artificial intelligence to review the files and predict what's most important to surface, maybe images or videos of a certain age group. If they have to prioritize by prepubescent versus mid-teen, they can do that without manual review. Our tools help the prioritization in both areas, but it also protects. We have a big focus on the mental health of frontline responders. Imagine, if you're an investigator and your job is to look at 10,000 files of child sexual abuse material every day; first, you can't do that in a day, and two, it takes a mental toll on you.

The more that we can pair technology with human intervention, the more effective we can be in finding and removing kids from harm, but also the more we can protect the mental health of those on the frontlines who do this work.

Rollo Romig: That makes sense. Would you say that your primary audiences for these technologies are the tech platforms and law enforcement? Is there anyone I'm missing from that or is it mostly those two?

Julie Cordua: Yes. For technology, it's those two. For the groups that we talk to when we publish our research, we have a broader reach. Obviously, it's the general public, but we often consult with policymakers to make sure that they're informed with the latest research. We talk to adjacent groups in the space, such as those working on advocacy and we may communicate with educator networks or parent networks. But law enforcement and tech are those who use our software the most.

Rollo Romig: That makes sense. Since your research reaches a broader audience is there any sense in which the research and your messaging based on that research is aimed at changing social norms? If so, in what regard?

Julie Cordua: The main social norm that we are trying to affect at this stage is making a case for this issue and bringing it out of the darkness. The social norm around child sexual abuse is:

"We don't want to talk about that." Then pair that with a social norm of parents and many adults not understanding kids' digital lives. Pair child sexual abuse with not understanding kids' digital lives and you have just a perfect storm: if we don't understand it or we're not willing to talk about it, we can't help fix it. It's not necessarily a social norm. Maybe the norm around not talking about sexual abuse is, but the main thing we're trying to shift right now is get this to be a conversation we have as a society.

Rollo Romig: That makes sense. I think there is a certain social norm about thinking that it's okay not to be entirely plugged in as a parent. I'm the father of a 13-year-old, and it's really difficult to even know what I should be paying attention to. What's something that you think parents ought to know or parents don't seem to know well enough about what's going on out there and how best to protect their kids?

Julie Cordua: This is my opinion–not anchored in a robust study and evidence-based research–but given how many parents I interact with and what I hear the most, I encounter people often in a state of fear. There is a lot of fear-based parenting when it comes to technology because we don't, as parents, know exactly what's happening behind the screen of our kid's device. To be fair, this is biased because of what I work on.

I think one of the things that could be most helpful is shifting to a curious-based mindset. I'm not saying I disagree with advocating for shifts in policy or regulation or limiting screen time or delaying when screens are introduced. All of that's great. That's a family decision. The reality is our children will encounter the internet. Even if you keep it out of your home, they're going to encounter it at school, they're going to encounter it at a friend's house, they're going to encounter it somewhere, and most likely they're going to encounter it in your home. Either they have a device or you have a device. Starting from a very, very young age, talking about the role of technology in your life, in a way that empowers the human and makes kids start to understand the role a device or technology should have in their lives as a vehicle to achieve what the human wants to achieve and not the other way around.

Second is having a curiosity around as you “inch” your child into being online: what are they interested in and how are they using it so that the child feels okay talking to the parent about their experience. When you talk to kids, there is very much a blurred line between their offline and online lives. One example I've heard folks talk about is they often talk about the danger of strangers online and say, "Well, you don't know that person. That's not your friend."

Actually, to kids, many of these people are their friends even if they've never met them. Already that language doesn't resonate with kids. Many kids consider who they meet online as friends, and for many kids, those are good friendships and they're fruitful and they're helpful, especially for kids who may be ostracized in their physical environment. You know what? In-person relationships can be harmful too.

So how do we think about talking to our kids about relationships and what are potential red flags for relationship harms? They could happen offline. They could happen online. Help teach them what to look for instead of "If your friend is someone I don't know, I'm taking that phone away." They're not going to tell you if something goes wrong because [they don't want you to take their phone away]. How do you lead with curiosity, with empathy, with education, and try to create a dialogue with your child so that they have a safe space to get the best out of technology while reducing potential harm?

Rollo Romig: That's a great answer. I really appreciate that. The technology that enables abuse of children online is accelerating at such an insane rate and your technology has to match that. How do you keep up both with your analysis of what's going on and with the technology itself?

Julie Cordua: There are two things. One, we can try to continue to build solutions that keep up with the potential harms. How we do that is why I think Thorn is special as an organization. We have folks internally who have deep subject matter expertise. They've been victim identification specialists. They've worked in trust and safety teams. They've been law enforcement [officers] who have knocked on the door and recovered these children. We pair them with incredibly smart engineers, incredibly smart data scientists and AI specialists. When you bring those two worlds together, you can stay on top of the trends. You can innovate quickly. You can deploy at scale. That's how we drive an innovation engine.

The second part of it is, I would say, we're finally getting to the place–although it's taken us a while–where companies are at the table. When we started, we were cleaning up the mess of social media companies not paying attention to this issue for the first 10 years of their existence. We were playing a lot of catch-up, and we still are, to be fair. But I do actually see some hope in where we're headed with generative AI because even though there's a lot of harm and a lot of things happening with the new companies, at least those companies are at the table. They know that their tools could create potential harm.

The people who started social media companies didn't even think about the potential harms back then because if you remember the beginning of the internet, it was the first time (they were

being built). These were built in dorm rooms. There was nothing called Trust and Safety when Facebook was created. The first Trust and Safety conference was three years ago.

The discipline of having a trust and safety mindset in a tech company did not exist until probably 10 years ago. Another way of keeping up is that we're now shifting—or adding, I would say, this concept of safety by design. You can work with innovators and designers to think ahead of time about how your innovations and technologies will create harm and design the environments to minimize the likelihood that those harms will be created. That is the ideal way to stay ahead.

You need to pair both. You need to do safety by design and have detection on the backend, but we're finally getting to a place where that safety by design is coming in as well.

Rollo Romig: That makes sense. Tell me, in your interactions with tech companies, what do you find to be most effective in getting them to sign on to your program and to be concerned about the things that they should be concerned about? Many of them have a reputation of not making any changes that don't increase their bottom line unless they're forced to. What do you find effective in getting them on board?

Julie Cordua: It's a mixed bag. You have some companies that want their platform to be known as a good citizen, and they do it proactively, and they seek out experts and partners, and they yell from the rooftops. You have others that want to be good citizens, but maybe it's such a big bureaucratic behemoth that the idea of fast-tracking safety has to go through 20 layers of approvals. To your point, it's not the one thing that will drive the bottom line, so it’s probably not going to be prioritized. Then, you have others who just maybe don't truly understand. They're new, younger. Then, you have some that are just blatant bad actors.

Different things work at different stages. On the first end of the spectrum, just simply being collaborative, sitting down, offering real solutions in their language is helpful, and you can make a lot of change. This is where the ecosystem approach comes into play. This is where regulation will start to play a role. This is where litigation will start to play a role. This is where bad press that they don't want will start to play a role. If you think about systems change, that's where different levers in a system come into play, to be effective.

Rollo Romig: It makes sense that it's a really big range of approaches depending on the attitude of the company. It's not like a one-size-fits-all thing. Tell me, is there an example that you like to share that illustrates the impact of your work?

Julie Cordua: I'll give a few. On the research side of things, we've had two of the largest social media companies take not only our research, but (take part in) one-on-one readouts with companies to share with them what we're seeing on their platform in particular. Twice, we've had companies make specific feature changes because of our research, which, if you think about scaled change: we can go company to company or person to person.

If you're a platform that has several million children on it, and you change how a child can report that they have been groomed or asked for-- and that change results in a 50 percent increase in reports, that is scale, because you're already at millions of kids on that platform. That's quite big.

We've had that happen twice, where you affect millions of kids with just the shift of a new feature implemented into an embedded platform. I actually just had the privilege of hearing from an investigator just this week who uses our CSAM Classifier in his forensic tool. He was talking about just the sheer volume of CyberTips that he has to go through. He shared one case where from the time they got the CyberTip, to the time that they issued the search warrant for the individual, to the time that he was able to find the video on the person's device–it was a video of a very young child–of doing that within two minutes.

The device had hundreds of thousands of files on it and allowed them to charge the case with the highest charging level, which would mean this person would go away for quite a while. You had to find a certain kind of file that was egregious enough to account for a certain sentencing. Before our technology, that [search] may have taken that agent weeks to sift through that device and look through everything. Now he can just plug it in, run our classifier, and it will pull it up. And it saved his mental capacity. [The classifier helps law enforcement] be more efficient, move on to another case where they may be able to identify another victim.

With the tech companies that we work with, we help identify millions of pieces of child sexual abuse material every year which they then take down. One of the most impactful stories I've heard was a few years ago. It was right when we were launching our child sexual abuse classifier. This company had committed to work with as one of our first launch partners. This meant that they would give us feedback on how it was performing.

They launched the classifier. It's a predictive tool so the trust and safety team at the company has to review every image that it predicts on and tell us whether we were accurate or not, so we can get false positives, and so that we can consistently train the tool. It was the first hit that this company had gotten off our classifier. Talking to the trust and safety agent who reviewed it, he

said before he opened the file, he was thinking, "Oh, well, I got my first false positive. There's no way this is going to be a hit because it was so quick."

He opened the file, and it was the active abuse of probably an eight to nine-year-old girl, so it was a real hit. He went into action. He looked at the timestamp on the image. The image had been taken within the last 24 hours. They then opened up the account, because it was a photo-sharing site. They went to the account of the person who had posted it. There were hundreds more images of abuse of this child. They were able to immediately call law enforcement and within 12 hours, law enforcement was able to go knock on the door of this family. It was the father who had been abusing his daughter, and the father was arrested and the daughter was recovered.

We help companies take down millions of pieces of content, but not all of them lead to the recovery of the child, but they do have a possibility of leading to recoveries like that. It is pretty impactful work.

Rollo Romig: Those are great examples. Tell me, what do you find to be the biggest challenges in this work, and how do you go about trying to address those challenges?

Julie Cordua: I think one is that it's a really big problem. It can feel never-ending. You just have to break it down into bite-size wins so that you can feel progress because if you don't feel progress, you're going to feel completely overwhelmed and burned out. The second is getting engagement from other parts of society.

Like I said, one of the biggest hurdles is just getting people to want to talk about it, to know that it's a problem, to know how big the problem is, and to know that there are solutions. (We also need) society to talk about it and tech companies to take it seriously. We need to find the right way to engage parents in an effective way, so they feel that they can be advocates for their children. We need donors to believe that you can put money into this and we can create real change. I think those are the biggest challenges.

Rollo Romig: To what extent are partnerships or other collaborations and coalitions a part of your work? Who are your main partners in getting things done? How do you cultivate those relationships?

Julie Cordua: Partnerships are critical to what we do. We definitely think of our work as part of an ecosystem. We have a huge amount of partnerships on the tech side because again, we serve. We say our goal is to equip the front lines. We are not a one-to-one effective organization.

We are a one-to-many organization. Our job is to mobilize tech companies with actual tools that they can use, so we have dozens of relationships in the tech industry. Then, we're a one-to-many with law enforcement as well. We work with hundreds of law enforcement agencies all around the world.

We have more than 3,000 agents who use our software. We also are part of policy and advocacy groups, because we use our research to inform that movement. Often, we have collaborative donor groups that fund us and we are a part of those initiatives.

We also participate in other groups, whether it's the World Economic Forum on trust and safety or other areas. I think those (engagements) are key because I look at the nine levers of creating long-term systemic change, and no one organization can do all of them. You have to really think about what your role is in those different levers. Where are you uniquely positioned to add value? How do you create these technical connections to the rest of the ecosystem where they can go do their own thing, but you can identify where you work well together and collaborate? I think the most important thing in a fruitful partnership is that you're both adding value. You're adding value. You're honest and trustworthy. Even if it's not pretty, you're honest and trustworthy. Those to me are the critical parts of creating a strong partnership.

Rollo Romig: Let's say hypothetically, you are talking to someone who was just getting started in this work of trying to end sexual violence against children. What sorts of insights or a teachable lesson you've learned from your own work?

Julie Cordua: One of the things that I was most grateful for in my journey getting into this work is that I don't know that everyone has the privilege to do it. I want to recognize it was a privilege to have this, and everyone might not, but I'll say it so if people can create the space (for others)., I entered this field 13 years ago not knowing anything about the field. My background was in technology. I'd never worked at a nonprofit. I was starting from ground zero. I knew how to start social enterprises. I knew how to bring together the strengths of business with a social mission to do good in a measurable way. I had so much to learn when it came to child sexual abuse, online child sexual abuse, any sexual violence, nonprofit work, all of that. I probably spent an entire year learning.

That is a luxury because I was funded to do that. I went out and I probably interviewed 100 people. I talked to tech companies. I talked to law enforcement. I talked to survivors. I talked to non-governmental organizations. I asked, "What are you experiencing? What are your risks?

Where are the gaps? What are the trends? What do you wish you had?" From there, I developed a thesis for how we could be useful in this field.

I would say for someone starting out, learn as much as you can before you get started. Then, the second thing is to be super-disciplined about only doing something that isn't already being done. A lot of people come to me and say, "I want to start a nonprofit." I say, "Don't start there. Start with what problem do you want to solve? Why should you be the one to solve it?" That really forces you to think about what you are going to do differently?"

If you're trying to solve a problem many people are trying to solve, but you have a new approach, do it, because it's a problem. If you're trying to solve a problem someone's trying to solve and someone's doing exactly what you said you want to do, join forces and do it with them, or go do something else. Take a lot of time to learn and audit.

Third, I think the best new ideas come from when you have aggregated disparate information. If you think, "I'm only talking to nonprofits in my field. That's it," then I think you're going to have limited new ideas. Let's say you're working in sexual violence, but why don't you go study the climate change movement? You've got to bring in information from different fields to come up with new ideas and drive things forward. Those would be my three areas of advice.

Rollo Romig: You mentioned earlier that part of the purpose of Thorn's technology is to help protect frontline workers by limiting exposure. But this is taxing work for anyone who's working on it. What have you learned about how to look out for yourself while working in this area?

Julie Cordua: It needs to be a constant question. I've learned that it will go up and down and you have to recognize that you're not always going to know when you're going to be okay and when you're not. Having consistent wellness practices in place is good. Giving yourself grace is good. Setting goals that are achievable and celebrating when you achieve them. Otherwise you just feel you've put your mind into hell and you're not making any progress which is not a good place to be. We say in this field of work, "Once you know, you can't unknow." You must be able to mark your progress along the way.

Rollo Romig: Absolutely. Thinking about the field in general and the collective efforts to tackle this problem, what do you think is most needed right now? What do you think the gaps are in tackling the problem?

Julie Cordua: We need to be better at telling the good stories of the progress that's being made. The problem is that we're building solutions for an issue that is rapidly increasing. If we look at

global health or disease eradication or climate, those fields are way older than us. They can tell a story of how many diseases have spiked and are now on the downstream. It feels like, "Oh, we're almost there. We're winning." If you think about the harms that technology will bring to our society, I think we're still in the beginning whether we like it or not.

The rate of innovation is exponentially increasing. It took 60 years for there to be 100 million cars on the road from the time the automobile was introduced. It took three or four years for there to be 100 million smartphones in the world. Now that there's smartphones in everyone's hands. It took two months for 100 million users to get on ChatGPT. Now that we have devices in everyone's hands, the adoption of new technologies is going to be exponential. We're at the infancy of this field and we are making progress in getting better. Like I said, 13 years ago we were cleaning up the mess of a social media field that did not understand potential harms the 10 years prior. I've been in it long enough to know that it's different right now with generative AI companies. There's still harm, but I can tell you it's different than it was when social media was created, and that it's better than it was.

We will get better, but we have a hard time telling those stories of progress because we're in a fast-growing harm field, and I think we need to get better at (telling our story of progress.) We are starting to see more progress and innovation in the regulatory and litigation space. I think that's been missing, and I think it will be interesting to see what comes out of that. That's going to add new pressure points in the field. And there's a third. I think one of the gaps in this field is what we talked about from a parent perspective. That may actually be one of the biggest gaps. We're making a lot of progress on the tools for companies to take things down and to create safer environments. I think we're making progress on the regulatory front. There's still a long way to go. I'm not saying it's solved. Just progress. On the parent front, there's a ton of information out there for parents, but I just don't feel like there has become a mechanism for how parents really know how to parent kids in the digital age.

Rollo Romig: Looking forward to the next five years or so, what actions or policies do you think have the potential to have the biggest impact on making change on this problem?

Julie Cordua: I'm not a policy expert, but one thing I think that could make progress is transparency legislation. I like it because I think it's doable. It's achievable and I'm a big progress over perfection person. We're starting to see some of this in Australia and other places requiring companies to answer this question: What are you doing to design your systems to be safe for kids? [They aren't asking companies] to do certain things, but saying you got to at least tell the world [what you will do].

The more that we can get the information out in the world, then we can figure out what to do with it. At least then parents could make informed decisions about where their kids are online and that regulators could have a better sense of what's happening. I think that that's a fair ask of tech companies. We're not going to tell them what to do, but we are going to say that they need to be explicit about what they [plan to do] and they need to share [their plans] publicly.

Rollo Romig: I really appreciate you taking the time with me today, Julie.

Rollo Romig: (he/him) is a freelance journalist who writes most often for The New York Times and The New Yorker. He is the author of the book I Am on the Hit List: A Journalist's Murder and the Rise of Autocracy in India. He teaches writing at The New School in New York City. He was born and raised in Detroit.

***This conversation has been edited and condensed.*