Episode 3: Deepfakes

[MUSIC PLAYING] KSENIA BAKINA: Welcome to The Cyber Armour, a podcast which champions voices and safety of women and girls in the digital world. This podcast is brought to you by the Center for Protecting Women Online and the Open University. I'm Dr. Ksenia Bakina, your host for the podcast. I'm also a research fellow and a stream lead in Law and Policy at the center.

On today's podcast, we are going to be discussing deepfakes and the harms they cause to women. To address this issue, some laws and regulatory measures recommend labeling of AI-created material on social media. So we will also explore what we know about deepfakes and how effective any labeling measures might be.

In order to take a deep dive into these issues, I have some amazing guests here with me today. I have Arosha Bandara, who is a professor of software engineering at the Open University, whose research and teaching focuses on software engineering for adaptive systems.

He also leads the Future of Responsible Techstream at the center. He is a member of the steering group for the Open University Center for Policing, Research and Learning, and is also Associate Dean and Director of STEM Research.

I also have Rose Capdevila, who's a professor of psychology and stream co-lead for human behavior at the Center for Protecting Women Online. Her research focuses on gender and digital spaces. She was an international collaborator for the European Funded SeGReVUni project on gender-related violence in universities.

And last but not least, I have Lisa Lazard, who is also professor of psychology and a fellow stream co-lead for human behavior for the center of Rose. Her research focuses on gender and how gendered identifications become located within relations of power.

Both Rose and Lisa are also currently involved in the gender equitable interactions online project, which is a four-nation European-funded project. So welcome to all of you. It's a huge pleasure to have you here with me today.

So I want to start this discussion by first mapping out an understanding of deepfakes, because I feel that it would be really useful for our listeners to first and foremost understand what deepfakes are, how they are spread, who they might affect.

And so I would like to start with the research conducted by Rose and Lisa, because you've done a wonderful study that does just that, maps out quite an in-depth understanding of deepfakes and in social sciences field. So Rose and Lisa, shall we start with understanding what are deepfakes, and what they actually do?

LISA LAZARD: So deepfakes are any kind of material that's generated using deep learning generative AI. And they're used to generate videos, audio, anything like that. And it's usually pornographic in nature.

KSENIA BAKINA: Do we know why we call it deepfake?

LISA LAZARD: It's been coined by a Reddit user in 2017, I think, which was a combination of deep learning, AI, and fake content on social media.

AROSHA BANDARA: Yeah, and so that notion of deepness comes from the architecture that's used to construct these AI systems. So they involve multiple layers of neural networks basically. So basically, synthetic ways of analyzing data in computers, where each cell or unit, sometimes it's called a perceptron. It represents a neuron in, that would be in our human brain, for example. So that's the analogous model that's used.

And by linking lots of these perceptrons or neural net cells together in multiple layers is how this technology is able to generate these images. So you start with an image that is completely random noise, and at each layer, you're adding more structure to it in a way that increases the likelihood that it matches the output that you're trying to generate.

So the training of these neural nets is based on looking at large corpora of images that are labeled with particular descriptions of the content in them, and the network learns from those images how to go from random noise to a relevant image, basically.

KSENIA BAKINA: And do we have any statistics, or do we know how prominent the use of deepfakes is online?

LISA LAZARD: It's an interesting question, because studying prevalence of deepfakes is, it's incredibly tricky. However, some recent reports have suggested that something like 98% of deepfake material is pornographic or sexual in nature.
ROSE CAPDEVILA: And the other thing that research indicates is that, of that 98% of deepfakes that are pornographic in nature, 99% of those are of women. And we have to just say in the context that not all deepfakes are negative, or it's really about the deepfakes that aren't consensual that we really have to focus on, and they mostly aren't.
But they are used also by people in sexual play. That's not what we're worried about. That's not our concern. Our concern really is the ones that aren't agreed with, that are done without the victim's permission.
KSENIA BAKINA: Yes, absolutely. And I also think that although women are-- there's a perception potentially that deepfakes often only affect those women who might speak up publicly, because obviously, we know that there's a lot of celebrities like Taylor Swift and political figures who've been targeted with deepfakes. But from my own knowledge, there was a case last year, in 2024, in Spain, where 15 schoolboys were sentenced for creating deepfakes of girls in their class.
And I think it's really important to also make our listeners aware of the fact that you don't need to be a celebrity, you don't need to be outspoken online in order to be targeted, because you could be targeted by your next door neighbor. And so what harms do you know that deepfakes generally cause to, the non-consensual deepfakes cause to their victims?
LISA LAZARD: There's a wide range of effects that's been identified in the literature. So things like reputational damage, potentially financial damage, and also a range of psychological impacts that can result from being a survivor of deepfake abuse.
But I think it's also really important to be aware that the communities produced deepfakes are not always about malicious intent. So actually, the scant literature that we've identified is actually suggesting or indicating that there's actually a wide range of motivations for creating deepfakes. And it's quite a complex area.
ROSE CAPDEVILA: And that's something we want to explore further. Don't the creators and consumers of deepfake and the motivations there, because some of them are definitely malicious. They're used in blackmail and they're used in all different kinds, but some of it is what's referred to as homosocial bonding.
The example that you give in Spain, young boys think it's funny. They think it's funny. They're not really thinking about other people or the impact it can have on the victim. They're just thinking about trying to look cool with their mates.
LISA LAZARD: There's also communities. You mentioned celebrity deepfakes, where celebrities are deepfaked. But it seems that the intention isn't to spread these deepfakes. It's more about fandom. So they're kind of isolated communities, and they usually put up notices that these are fake, and that they're really to do with imagination fan play.
ROSE CAPDEVILA: They align it with fan fiction, don't they?
LISA LAZARD: They do. Yeah.
KSENIA BAKINA: Oh, wow. I was not aware of that. Actually, talking about fandom deepfakes, so if there are deepfakes created, let's say, of celebrities, do they actually cause that much harm? Like, say, if it's a fan creating a deepfake just to fuel their imagination and it's not what we see being spread on Twitter, it doesn't go viral, it just stays within that community, is there a harm to the person who has been deepfaked in that instance, do you think?
LISA LAZARD: Well, that's the question, isn't it? I think it's quite a complex one and it's certainly something that's been raised in the literature. But I think that if it's non-consensual, then there is a harm.
And it's not just about that particular celebrity, it's about the perception that it's OK to generate non-consensual images. And using celebrity status as an excuse really doesn't help us challenge those kind of power relationships that underpin the creation of deepfakes in the first place.
ROSE CAPDEVILA: The assumption that celebrities don't have a right to privacy, or control of their own body is problematic.
AROSHA BANDARA: I think there's also an element in terms of that motivation, where people are trying to understand and learn about this technology, and even if they are trying to learn about it for benign purposes, to try and use this generative technology in a positive way. It seems that the communities that are engage with and sharing information about how to use these technologies have a bias towards generating sexual deepfakes.

So the Reddit channels where this originally started, a lot of the content was focused on, well, how do you generate nude images of people, or how do you generate videos that are based on pornographic videos, but superimposing other people into those scenes and so on. And that means, yeah, so you have a legitimate and benign reason to want to use these technologies, until more recently, your sources of knowledge and expertise was in this community, and that had an additional effect of you couldn't learn this without being exposed to this content, basically.

ROSE CAPDEVILA: Sorry, I was just going to add that it doesn't only harm the people who are victims of it. The whole idea of deepfakes and the creating pornography actually isn't great for the people who are making it either.

I mean, there's something about how it impacts the way they view the world, it impacts the way they view women, the way it creates communities around toxicity is problematic. And the impact it has on society more broadly, it creates societal harms as well by also mitigating those relationships in ways that are quite negative.

KSENIA BAKINA: Yeah, absolutely. I completely agree with you. And I think it's really interesting how the harms, first of all, the harms that this online image that's been created, causes to its victims are very much in the real life. They're very much felt from physical harm through physical symptoms of anxiety and psychological harms, of course.

And then, there's financial harms that could be involved and reputational harm to the person targeted specifically. And it's interesting that you've highlighted that even if that deepfake doesn't necessarily spread online, there's still a chance that it will be harmful because you take away the control of that woman to consent, and you take away her autonomy.

So the viral aspect isn't necessarily a decisive factor, but then we go further to see what harms it causes on a societal level beyond that individual victim and how it can catch software developers who might have no intention of viewing deepfakes and no intention of participating in creation deepfakes, but they are so almost pushed into that by this maybe potential toxic masculinity ideals that in order to learn a deepfake, you should learn sexual deepfake.

AROSHA BANDARA: I guess that's the important distinction, what you said at the end of that statement, Ksenia, that these software developers are interested in learning about how to make deepfakes, as in, images that are not real, that are synthesized using other content and images, but they're not interested in the sexual aspect of these deepfakes. But yeah, as I said earlier, the route to that learning, until more recently, was in these communities that were predominantly engaged in generating sexual deepfakes.

LISA LAZARD: And just following on from that, I think one of the things that has been found within the literature is that the more technical communities that are generating these deepfakes are often also performing a kind of toxic masculinity. So although it is a kind of masculinity that's presenting itself as interested in technological prowess, it's actually also doing some really problematic toxic masculinity behaviors, which again, contributes to that overall promotion of toxic environments online.

ROSE CAPDEVILA: And it is the objectification of women. I mean, it is that par excellence, is it?

KSENIA BAKINA: Absolutely. And I just wanted to explore a little bit more. We've touched on this already, but in terms of motivations, we've talked about some of the motivations for the creators of deepfakes and why they might share it, deepfakes.

But do we know anything about people who are bystanders, for instance, who might just come across these images through social media why they don't report it, for instance, straight away, or why they might participate in sharing it? Is it misogyny, toxic masculinity, or are there any other motivations behind it that we might be able to pinpoint?

ROSE CAPDEVILA: Well, I think an interesting thing about deepfakes is that you don't necessarily know they're fake. So you don't know when you see one if it's fake or not. So you would just assume. I also think there's a normalization, a sense of disempowerment in terms of social media, where people really don't think if they report things it makes any difference, so why bother. And that's something we need to work on.

LISA LAZARD: Small amount of evidence, I think. I think that's fair to say on bystanders, partly, because of the factors that Rose has identified. But there has been some experimental studies that suggest that people are more likely to intervene in some way if they feel angry about what they're seeing, rather than if they're feeling guilt for seeing it. So there are, I think, some interesting avenues to explore in the future.

KSENIA BAKINA: I think, Rose, you made a really interesting point about bystander behavior that it's quite difficult to judge and test, because a lot of the times, when you see a deepfake, you might not necessarily know that it is fake.
And what's interesting is that a lot of laws and regulatory measures that are now being implemented are both at EU level and beyond. We have Ofcom, who's issuing guidances for various platforms to ensure that any AI-generated material is labeled, that it is made by AI, with this focus on transparency.
So I want to come to Arosha, who has done some research, specifically on this issue of labeling AI-generated material. So Arosha, could you tell us about your research?
AROSHA BANDARA: Yes, absolutely, Ksenia. So yeah, together with some colleagues based in Sri Lanka, at the University of Colombo and here at the Center, we came to the realization that even though there were these legislative and regulatory measures being introduced, we didn't really understand what makes an effective label.
And what we were observing in the online ecosystem was that platforms and content providers were choosing designs of their making and putting them out there to see, I guess, in an attempt to be ahead of the curve and demonstrate that they were doing something to address this challenge. But again, with very little evidence behind to justify the choices they were making in how they were presenting these labels.
So their utility is to be determined. I think we're still trying to understand their effect. I think the reason for them comes from the earlier work and imperatives to somehow identify misinformation content.
So the kind of clickbait headlines and content that goes with that tries to mislead people who read or engage with it, and various efforts to try and control that through mechanisms like on Twitter, now X, you have this kind of community notes feature where people can add additional information and context to highlight if something may be being misrepresented.
In other platforms, you have fact-checking links and labeling to indicate whether something has been fact-checked and found to be true or not, and the extent to which that is. But with deepfakes, I think there are two main challenges to those approaches.
So one is about the fidelity of them. And as Rose highlighted earlier, it's very hard for anyone to just look at something and determine if it's fake or not. And then the other one is the mechanization of the technology. The mechanization as a result of the technology developing means these things are being produced at such a rate that it's impossible for people to look at these, make those judgments.
So I think, at the moment, it's really an intuition hypothesis that these labels are useful or could be useful. But our study was exactly trying to look at how useful might they be, and how might different designs of these labels affect that utility.
So what we did was looked at the ecosystem of labels that were out there and identify some dimensions of design. So the language being used to describe whether something is AI-generated or not, or whether they explicitly use the term "deepfake" or fake in some form. The colors and icons that are used. So some of these labels use exclamation mark, warning triangle, and red. So things that are indicative of danger and so on. And then also the positioning. So some labels cut across the content and make it very, you can't ignore it. Other positionings put the label unobtrusively, either outside the image but nearby it, or on the image, but still not obscuring the main content. So there's a whole range of things, therefore, that we wanted to look at.
KSENIA BAKINA: Yeah, and I looked at your study and I found it really interesting how you collected these different types of labels and tested them on the users. Some of the findings were quite surprising for me. So could you tell us more about which labels you found were more successful in communicating the message and being believed from these sort of samples that you've tested?
AROSHA BANDARA: Yeah, so let me give a little bit more context about the study so that listeners can understand what we did. So we picked eight images, where there were four real ones and four fake ones. And in each pair of real and fake images, there was some that we're representing political figures, and others that were just representing celebrity/entertainment type of topics, basically.
But we labeled all of these images with different forms of label, and we asked people, in an online survey, to look at the image with the label and tell us the extent to which they believed that it was AI-generated, how they might engage with that image, whether they would like it, share it, comment on it, and so on. And we also had a control group who was shown the

same eight images and were asked the same questions, but those images were not labeled. So that was our comparison group that we used.
And, for me, the most surprising part of the finding was that any label had an effect on making people believe that the content was fake. Maybe that's too strong, that it was AI-generated, because that's how we phrased the question as well, that somehow AI was implicated in the generation of this image.
People really believed the label, even if it was a real image that they may or may not have seen elsewhere. But even the real images, if we attach a label to them, people think that there's AI implicated in the generation of them.
And in terms of the effectiveness. So there wasn't a lot to choose between the labels. All of them had very similar kind of effects on that belief, but labels that provided additional information. So there's a particular label design standard that's being looked at in the industry called content credentials. And we found so that form of label had a slightly higher level of belief and trust in AI-generated content being correctly interpreted as such, and therefore, acted on with more moderation by users.
KSENIA BAKINA: No, I think that's really interesting. It's great to know that labels can work and make the users believe that the content is AI-generated. What I found interesting was that amongst the different labels as well, there were some that marked in red with "splashed across".
And my intuition was like, well, the ones that are most visual, the most vivid, they're going to be most effective. But the reality of it was the opposite, was that actually it was the labels that were much more subtle that were placed either at the top of the image or somewhere on the sides that did not interfere with the image in itself were more likely to be believed. And I found that quite surprising.
And what I also found really surprising from your study is that the one issue was, will the users believe that it's AI-generated, so if there is a deepfake? But the other question is, even if they believe that it's a deepfake, will that affect their subsequent behavior? And I think your study found that there's a limited correlation to that.
AROSHA BANDARA: That's right. Yeah, broadly, the presence of a label didn't really change their interaction, engagement with it. I mean, that's a limitation of the study in that we didn't ask them what form that engagement might take. But even between liking, reacting to it. So again, we phrased the question as whether they would react to it. So it was the form of interaction rather than a specific type of interaction.
Some respondents might have said they would react to it, but their reaction might have been thumbs down or angry or what, something that indicated they recognized it was fake and they wanted to denote their dissatisfaction. So we didn't have sufficiently granular data to separate those different types of interaction. But broadly, the presence of the label didn't affect people's interacting, engaging with it.
This is important, I think, from the point of view of the current models of how social media platforms and their algorithms operate, because any form of engagement is seen as an indicator that content should be promoted to others.
And so the presence of a label actually might have this unintended consequence that unless the algorithms are adapted to reduce the prevalence of this content being shown, it might actually promote it and make it go even further around the world virally than had the label not been there because it encourages this engagement with it.
So yeah, I mean, similarly with commenting and sharing, people might have said they would share it because they intended to say, look at this fake image. Don't believe it, but they'd still share it. But again, it would have that same effect on the current model with the algorithms. The fact that it's being shared multiple times, even with a negative comment, would only promote it further.
ROSE CAPDEVILA: Yeah, sorry not to go hugely off topic, but I think the point about promotion is a really important one for all of our work at the Center for Protecting Women Online, because it's not just what people do, it's also what's being promoted online and why it's being promoted. And that's having a huge impact on how these things are playing out in society as well. And we need to recognize that. So there's all that element of it that I think is really important.
KSENIA BAKINA: Absolutely. And I think what I'm finding really interesting about this discussion, and what we're highlighting here is that, on one hand, I think what comes clearly from Arosha's study is that software developers or those people creating deep fakes, need to be very careful about what kind of standards they impose for labeling to begin with.

And at the moment, there isn't a single standard like a universal standard that everyone knows and trusts, and some may be believed less rather than more. But ultimately, there's so much emphasis being placed on transparency as the solution to the problem of pornographic deepfakes the ones that affect women the most, the ones that cause so much harm that are non-consensual.
But the reality of it is that potentially these labeling exercises aren't going to make the problem go away. And, in fact, through promotion, either positive or negative, they are going to exacerbate the problem even further.
AROSHA BANDARA: Yeah, and I think the other issue and challenge with labeling is our study just looks at the design of labels. It doesn't look at the decision process that sits before a label is actually applied to some content. But what it does show is that if that decision process has flaws or shortcomings, which mean real content is mislabeled as being fake, people will believe the label and interpret that real content to be fake, which I think has real challenges and notes of caution for areas where trust is really important, like politics, for example.
If images related to political topics are mislabeled as having AI-generated content in them, and that erodes trust when that's not a valid erosion of trust, that has some serious implications that we need to think about. So it's just adding to that point that labels are not a panacea, because there's a whole range of other things that go around them that need to go around them to make sure that we can be confident that we're labeling things, the correct things in appropriate ways.
ROSE CAPDEVILA: And just to pick up what Arosha was saying about political images, I think we could argue that all of these images are political, because they're all performing a specific political function in the way that feminists have always argued that the personal is political, and that has impacts for the way we function in society. I think it's really important, that element of it.
KSENIA BAKINA: Absolutely. I couldn't agree with you more on this aspect. So we've discussed what harms deepfakes cause to society, and we've touched upon what motivations those who are sharing deepfakes and creating deepfakes might have.
And we also talked a little bit about these measures that are being implemented as a first response to deepfakes, such as transparency requirements and labeling of AI-generated content, that they might miss the mark and may potentially actually cause the images to spread and be promoted further.
So I want to look to the future and have a think about, is there any information that we're missing before we can understand deepfakes better? Is there any research that still needs to be carried out in relation to deepfakes to increase our understanding of this phenomenon?
AROSHA BANDARA: A starting point, maybe to just be able to clearly define what is in scope as a deepfake versus other uses of AI to generate media, whether that's images, videos, or audio. Because I think part of the challenge with deepfakes is around when they're used to deceive versus when they might be used to educate, entertain, and engage in positive and legitimate ways.
And at the moment, this is a challenge. I think we have the language to use and how to clarify what's included under different categories, and to be able to communicate that effectively to the wider world.
We have the same challenge with "AI" as a term, as in, it's used as a catch-all, but predominantly, now refers to generative AI, even though there's predictive AI and different forms of AI that do different things, very different things, and have different capabilities and limitations.
And so deepfake is another one of those terms, which is kind of overloaded now because it's kind of prevalence in media. And part of the research, I think, is about looking at the dimension and scope of AI-generated content and what clearly does fall in the category of deepfake versus other forms of AI-generated content.
ROSE CAPDEVILA: And I think in our recent research, that's certainly what we found as well, that there is a kind of conflation between image-based sexual abuse, which can be photoshopped images and deepfake. And that's happening in research, but it's also happening in wider public arenas, where I think deepfake is now being used as a way to describe online fake material, which is obviously quite a wide remit of content that could fit into that category.
AROSHA BANDARA: And exactly, I mean, that thing about Photoshop as one particular image editing product. But many of these products now integrate these AI-based tools, AI-

trained models to allow effective merging of images or removing backgrounds, and so on. So all of that based on different forms of artificial intelligence, and therefore, the overuse of this terminology just confuses rather than clarifies how we need to deal with the really harmful and problematic uses of AI in the form of sexual image generation or other forms of harmful content.

ROSE CAPDEVILA: So I think that one of the things that our review identified is the lack of research around the way in which deepfake is used to control women, to put women in their place, so to speak, to make them feel unsafe in online spaces. And that's a really important element of it. So it's the way the deepfake can do that in a way that maybe other types of posting can't.

And I think that's something that we were wanting to explore further, because that's something that's just not there, that research around those processes, aren't there yet, and importantly, how the platform providers can mitigate that, and what they can be doing, which they're not doing to alleviate that particular danger.

LISA LAZARD: Just picking up on that, Rose, I think that element of control is really important because I think that also when we discuss it, we sometimes miss that actually posting is an everyday practice. It's not uncommon, particularly with younger generations, to post all sorts of material, and it's part of normal everyday behavior now.

I think we also probably need to attend to and recognize that a lot of sexual relationships do happen online. And actually, it's quite normal for young people to do that. And I think part of dealing with the problem of deepfake is embedding an intervention that recognizes that properly, rather than undermining young people's practices, because it is just how it is now.

KSENIA BAKINA: I think that's really interesting. And I also think that there should be more work done surrounding these ideas, the idea of control, because from what's also come up in our discussion is that, well, what happens if that deepfake is not shared? If it's just used to control the victim, then all of that labeling and transparency measures, they don't even come into play because it's not widespread. And so I think definitely more research needs to be done about what harms creation alone can cause to women.

I agree that when I did my research on image-based sexual abuse and when the image-based sexual abuse first came out, there was a lot of messages focusing on what I call these victim blaming attitudes, where they say, well, just don't send your nude images to your partners.

But this is paternalistic and unrealistic advice because, as you've said, so many relationships are formed online. There are so many long-distance relationships. And I remember coming across articles in like magazines, such as Cosmopolitan, where they would say, light up some candles and have Skype sex, because the Skype was the thing at the time.

And so when you have, on one hand, advice to engage in these kind of, if you're in a relationship with a partner and he's weighing in these activities, but then, on the other hand, send the message, well, don't send your image, because if you do, you're the one to blame.

But I think with deepfakes, and particularly with sexual deepfakes, I think the scary aspect of it is that you don't even need to send anything. And I think that's what's so disempowering about it, is that it can be done from one image of you, taken from LinkedIn, or even taken of you going out about your business, because the technology that phones have at the moment, you don't even need to be active on social media.

Look at the case of 15 schoolboys and using pictures of their classmates. So I think it can affect, irrespective of whether or not you do send anything. So that makes it a much harder to prevent because anybody can be a victim and to control as well.

And so that kind of brings me on to another question, which I want to explore. So then, how should we deal with deepfakes in a society to counteract these harms? Should we just ban them outright? Is there a case for that?

ROSE CAPDEVILA: I mean, we could do, but that wouldn't really solve the societal problem that produces the abuse of deepfakes. So I think that dealing with deepfakes requires us dealing with wider societal problems about how we engage with each other, how we regulate social media providers, the kind of behavior that's expected, what's considered acceptable behavior in our society. I mean, those are much, much bigger issues, but they come into play with this.

AROSHA BANDARA: Yeah, and absolutely, I think it does require thinking about and addressing the social and human factors much more. Technology, although maybe has enabled and exacerbated those behaviors, it's not the root cause, if you like. There are other

factors to consider. But it has created also, well, maybe a need for online platforms and image and content sharing.
Maybe we need to think differently about how we engage with and interpret online content, and have to flip the narrative of one from trust to one to distrust, have mechanisms to verify and validate content, which personally I think that's a sad state of affairs if that's the future we have to live in, that you have to mistrust everything by default, unless it's actively and positively validated to be true. That's not a world I want to live in. But it may have been-- it may be the world we've inadvertently created with some of these technologies.
ROSE CAPDEVILA: I think it's also really important to bear in mind that deepfake is a new manifestation of gendered sexual violence, and that we do know quite a lot about and we already know a lot about how we can stop that.
And there's an inordinate amount of literature that has a very clear pathways to ending violence against women and girls. And I think that's probably where we need to start. Rather than seeing this as a completely new problem, we need to go back to the older problem and actually pay attention to the research that's already been done and enact it.
KSENIA BAKINA: And I think that's a really interesting point. As we've said that potentially deepfakes are just a symptom. They're not the cause. And especially the cause of sexual deepfakes is gender-based violence, and generally societal attitudes to normalization of this kind of content. So then the bigger question is, Lisa, were asking for it when you said that there are already clear pathways. So could you please tell us what do you think needs to be done to tackle the cause of that?
LISA LAZARD: Normalization is something that we really do need to tackle. And we do know that as soon as something is not normalized quite clearly, then we do see a change in behavior quite consistently.
So that's where we need to tackle. And to do that, we need governments on board. We need bigger organizational support. It can't just be done by a few people trying to tackle the problem of deepfake violence.
ROSE CAPDEVILA: I don't want to sound trite, but I think that's why the Center for Protecting Women Online is really nice, because it has the legislation and policy. It has the technology. It has the human behavior element. Bringing those things together is, I think, where I think that's a good place to be looking for these ways forward.
KSENIA BAKINA: Thank you for this discussion. It's been illuminating, enlightening, and super interesting, and I hope that our listeners enjoy listening to it too. In today's episode of The Cyber Armour, we discussed issues in relation to deepfakes, and the way they're being tackled to make the internet a safer space for women.
I hope that you enjoyed listening to this episode and follow the Center for Protecting Women Online on our LinkedIn page. You can also stay tuned for the next installment of this podcast.
[MUSIC PLAYING]