ISOJ 2020: Day 5

How to fight deepfake and cheapfake videos: The challenges of verifying authenticity of visual content

---

Chair: Claire Wardle, U.S. director, First Draft

- **Christina Anagnostopoulos, senior producer, Reuters**
- **Rhona Tarrant, U.S. editor, Storyful**
- **Matthew Wright, director for research, Global Cybersecurity Institute, Rochester Institute of Technology**

Rosental Alves Hi there. Hello, hello. I am I'm kind of becoming sad because it is the last panel. I can't imagine how I'm going to live next week without ISOJ online, but anyway, it's a grand finale because this is a great panel with a very important issue. So thank you very much, everybody, for joining us again in the last day, last panel of ISOJ online. The conference has been a great success. I'm very happy. If you would like to watch in Spanish, remember that you have to click the interpretation globe on the bottom of the Zoom screen and select the Spanish-language channel. And we are live streaming on the ISOJ YouTube channel in English and in the Knight Center YouTube channel in Spanish, in case you have any tech issues with Zoom. Also, please remember to follow and use hashtag ISOJ to stay connected with the conference on social media. Remember that after this panel at 5:30 Central, we are going to have a fantastic virtual happy hour on the virtual campus of UT and the Moody College of Communication.

Now I would like to introduce you and to welcome Claire Wardle, who is going to present this panel on How to Fight Deepfake and Cheapfake Videos: The Challenges of Verifying Authenticity of Visual Content. Go, Claire.

Claire Wardle Hello, my name is Claire Wardle, and I'd like to introduce this panel, which is entitled How to Deepfake and Cheapfake Videos: The Challenges of Verifying Authenticity of Visual Content. I'm the U.S. director of First Draft, and we're a nonprofit that help newsrooms tackle exactly the problems we're going to talk about today.

Now, we have a stellar lineup. We have Matt Wright, who is director of research at the Global Cybersecurity Institute at the Rochester Institute of Technology. We have Rhona Tarrant, who is the U.S. editor of Storyful. And we have Christina Anagnostopoulos, who is a senior producer at Reuters. And basically, we just found out she's based in Mexico City, but soon to be moving to D.C. just as the U.S. election kicks off.

And I've said a few times that, "I'm not losing sleep over the threat of deepfakes right now. I'm actually more worried about the threat of people denying something that actually happens by saying, 'oh, that's a deepfake.'" But the threat is real. And as the technology gets more sophisticated, we could actually be in really serious trouble by 2022. And luckily, we have people like Matt to keep us on top of the technology. But the one very good thing about this obsession with deepfakes is that this sexy name has actually made everybody suddenly take manipulated images of videos seriously. Many of us have been working around on this topic for about a decade, but we could never get people to take it seriously.

And now people are like, "Claire, can you come and talk to us about deepfakes?" And I sneak in and then I tell them really they should be worried about cheapfakes, which is why we have Rhona and Christina who are going to talk about the work that they do every single day, which is doing solid journalistic work to make sure that newsrooms don't run with false or misleading imagery.

So you're in excellent hands with this panel. Just to let you know, all the Q&A will be at the end. As the speakers go through their presentations, please add your questions into the chat box to make sure that when we get to that Q&A session, there's a whole long list of excellent questions as opposed to comments that our speakers can respond to. So I'm going to kick off with Matt, who's going to probably scare us rigid about the threats of deepfakes, but also what we can do about it. So over to you, Matt.

Matthew Wright OK, thanks a lot, Claire. So, yes, I am going to talk about the things that might scare you about deepfakes as well as hopefully what we can do about it. I'll talk about technical progress and limitations on things that we've been working on. Don't worry, there will be lots of lines of code so that you can follow along easily. Alright.

So what are deepfakes, of course, is a portmanteau of deep learning and fake videos, and this includes several different types of deepfakes. One of those that you can see here is where there's a face swap, and this is the one that you probably see most often in entertainment based videos or if you've seen some of the deepfake porn that's been out.

But there's also this type, so here we have President Obama. And here I'll take a second and maybe we can really test the capabilities of doing these types of online conferences. If we're all doing Zoom, then we have a participant window, and if you click on that, there's a little button for a yes and a little button for no. And we'll use that. If you think the Obama on your left is the fake one, click yes. And if you think the Obama on the right is the fake one, click no. I'll give you just a second to do that. And it turns out, it's the one on the right, so I have no idea how well you all did, but usually some people definitely get this wrong.

The MIT Media Lab is running a site where they have so far found that 73% of people get their responses correct on different deepfake videos, by the way, not on the latest state of the art stuff. So that tells you definitely this is a thing that people can look at the video and definitely be fooled.

Here we have a video of an Indian politician, which I'll play for you just a little bit of.

This is an Indian politician who wanted to be deepfaked. So this is a video of him speaking English, but he didn't speak in English originally. They made a deepfake of him in English and a number of other languages so that they can spread that out to the Indian public without having him actually learn a bunch of languages he did not know. And so deepfake politicians are here. They have been made. They aren't being made surreptitiously, at least not yet, or at least not as far as we know.

Given that, what can we do to detect it? There's actually quite a bit of work going on. There's two basic types of methods. One is on using biological signals. So, for example, with the video of Obama, you can actually learn, well, what are the behaviors of Obama on video as opposed to other people? And can you recognize that, not just to distinguish it from other people, but in this case to distinguish it from a fake video of Obama that's not moving exactly the right way and not tilting his head at exactly the right time. Then there's another group of work that looks at the specifics of video frames and is looking for weird

inconsistencies with what's going on. And this is some of the work that we've been doing, along with a number of other people.

One of the things that I think is worth pointing out here is that you can see the years at which this work is done. Normally when I put up a slide like this on recent literature in the field, I've got literature from five years ago, four years ago, and then a couple maybe more recent. Here you can see 2019, 2019, 2020, 2020, so you can see that this is a really fast moving area.

In fact, Facebook and Microsoft have recently joined the fray and have created this contest to detect deepfakes, the Facebook Detection Challenge. And this is really interesting. They put up a lot of new videos, brand new deepfake videos that had never been tested before, as well as some non-fake videos so that we can see whether we're getting the right kind or not. It turns out that no team, no technical team was able to break past 70% on the private data set, which is pretty bad. I mean, 30% of things getting through. Unfortunately, our team submission actually failed, so for technical reasons, we weren't able to get it going. It makes me feel a little bit better because Facebook's team submission also failed, so they couldn't even get their own software to work on their own data set. In any case, there's a lot of deepfakes that are getting past even the best state of the art systems.

But the tool that I do want to tell you about is our deepfake tool. Where it is effective and where we've been, at least in terms of publishing papers on data sets of deepfakes, we can say we're pretty effective on forward-facing videos where there's like a single face, just like we are now. And where it runs into problems are where there's different poses or multiple faces. On the one hand, there are things that we think we can do to improve this. We can add more new videos to the data set and also enhance some of the tool's capabilities. But then there are also some open questions. Well, what are the real uses of a tool like this for journalists? And, in particular, if we have a tool like this, but we know the accuracy isn't perfect, what is it that we can do that would make the tool most useful for you?

For example, here's a visualization that we've developed that you can see kind of highlights that this is one of the fake videos, and it's highlighting where in the video it is that it thinks that it's fake. And so it may be something like that could be useful in cases where you're not sure whether it's a deepfake. The tool is giving you an answer, but you want to know more about what it's looking at.

Another thing we might be able to do is provide better uncertainty values. So, for example, in conditions where we aren't sure, where the tools shouldn't be confident, if we just told the tool to run right now, it will not only give an answer, it will say that it's very confident in its answer, no matter what kind of situation you put it. In fact, the weirder the situation often it will even think it's supposed to be more confident. What we might be able to do is detect situations like that and say, "Well, you know what? Here we have, for example, bad lighting or multiple faces, or just maybe it's a video that's too creepy, even for the tool." And in situations like that, the tool could throw up its hands and say, "Look, I don't know, I don't think I can give you some kind of answer, but you should really be careful with this answer." So maybe something like that could be helpful.

Alright, well, with that I'll wrap up. I just want to say thank you to all of the members of the team, as well as the ethics and governance of AI initiative and the Knight Foundation for their support. And so with that, I'll go back to you, Claire.

Claire Wardle Thanks so much, Matt. I think it's really important that we have a sense of the types of things that we must be most worried about. And in fact, there was a story last week about somebody in the U.K. that purported to be a University of Birmingham student at Wilson and his face that was being used was a deepfake. So those kind of like head-on profile pics, clearly there's an issue with that. The question is, how long will it take for those more sophisticated versions to become a real problem?

But while we're waiting, it's not as if we don't have a very serious problem when it comes to so-called shallowfakes, deep cheapfakes manipulated or misleading videos and images. And so I'm going to turn it over now to Rhona Tarrant who does this every single day as the U.S. editor of Storyful, and she's going to talk you through some cases, I expect.

Rhona Tarrant Thanks. So, as Claire mentioned, I'm the U.S. editor for Storyful. Just to briefly explain what we do first, and we work with newsrooms in two ways. So first is we verify footage from social media, from breaking news stories. So that's everything from conflict zones to terror attacks to the extreme weather in the U.S. And most recently, that means verifying videos around the coronavirus and the anti-racism protests across the U.S. And the other way we work with newsrooms is through online investigations, so we partner with a lot of newsrooms to do visual investigations, to piece together what happened using user-generated content or open-source imagery, or tracking how misinformation or disinformation spreads across social media platforms.

And so those of you who work in the states will be familiar with this graph. But I think it's a good way to think about different categories of platforms and to think about how information and in this case footage flows between them. When we're gathering footage from breaking news stories, we're mostly looking at the top tier here. But often footage is emerging from private communities or closed networks, which is a challenge for journalists because they're really hard to track, and tracking back is really important.

So there's been much discussion about deepfakes as something of a technological issue, I think. And that is improving technology makes it easier for bad actors to create footage that appears real, but it's fake. To be honest, though, that actually hasn't factored into my day to day yet, I would say. The majority of deepfakes I see online are created by people who are warning about them, which is kind of ironic.

But from my perspective, as Claire mentioned, it's the so-called shallowfakes or cheapfakes that are more problematic. And that's a piece of content that's altered in a way that's not sophisticated, but it has the power to really mislead. Or actually the biggest problem that I see in a daily basis is a piece of video taken out of context. And it's important to say as well, the majority of footage that we see coming out of breaking news stories don't fit into any of these categories, but when they do emerge, they have the potential to do a lot of damage.

So I suppose kind of what I'm saying is you don't need a high degree of sophistication to spread disinformation. And in fact, the success of those seeking to spread disinformation is actually just creating doubt. So just making you go, "Did that actually really happen?" Right, then it completely convinces you of another narrative. Just inserting the doubt in your mind, or exploiting people's suspicion and distrust, our tendency to believe narratives that fit into their own worldview when it comes to breaking news story.

And so news organizations will depend on us to monitor conversations in realtime and verify footage of news events that are happening, and verifying in real time presents its

own challenges. But we generally rely on the first principles to trace origins. And we use the same approach for manipulated content and deepfakes that we would for all of our videos. And that is the fundamentals. Verify the time, verify the date, verify the source. If you can answer those three questions, you're on your way to figuring out if a piece of content is authentic or if it is what it says it is. Also an important step as well, I would say, is tracing back and sourcing the earliest version of the footage online, which often provides additional clues. And you could, for example, some of the steps we would take is comparing things frame by frame, analyzing the video on its own to spot inconsistencies or evidence of manipulation.

And so just to bring up this example, you might remember this really famous one. I think it was 2018, CNN's Jim Acosta had his White House pass revoked after he refused to go up to the microphones asking a question. So Sarah Sanders had tweeted this video claiming that it showed him sort of kind of chopping his hand down on a White House intern arm after she tried to take the microphone. And so it turns out the White House had actually tweeted a video that had extra frames in it, and which made Acosta's hand movements look sharper than it actually was. And so this video was picked up online. It was online before it was tweeted by the White House. And I want to use this one as an example, I just give it here our own analysis.

And what we did was actually pretty manual. I know there's a lot of talk about what tools journalists can use, but this was a case of us finding the original, which was C-SPAN, and then lining up the two versions beside each other frame by frame. And we were able to detect extra frames. So it's a good example of going back to basics and figuring out what you can figure out. I think sometimes journalists get a little bit worked up about technology, and how do we know? And sometimes it's just a matter of taking a break and kind of going, "OK, what can we figure out here?".

Using old footage and claiming it shows current events is one of our biggest problems. And this instance that I'm showing here, no editing required. It's just you recycling what's already out there to push your own narrative.

So in this case, several videos and images falsely came to show strikes on U.S. bases in Iraq were widely shared across social media earlier this year. It seems like another year. But essentially a lot of videos came out that night saying this obviously shows Iran attacking U.S. bases, and it's something we see a lot, particularly when it comes to conflict. Old images and videos being recycled.

In terms of verification, there are a lot of ways you can approach these. In the case of these examples, it was just simple reverse-image search because these images and videos had been widely circulated already. But I will say nighttime footage is generally really hard to verify, and that's one of our biggest challenges, too. And sometimes we simply just can't confirm or debunk footage, because if we really can't say if it's fake, we just know that there's nothing really we can say about it. And often the metadata has been scraped because it's been shared in closed networks like WhatsApp. And so that's another challenge.

And then just here are two other examples of videos out of context from the past few months. On the left, it's old footage being circulated claiming to show one thing. And on the right, it's groups online claiming that a video shows an Asian woman with COVID spitting on fruit in an Australian store. And this is at the height of when it was just emerging. I think it was March, April, particularly in Australia and the U.S. And in the case

on the right, it was a matter of us getting a statement from police about what actually happened. So, as I said, just trying to figure out. Sometimes we just have to go back to the basic principles of actually calling someone up and asking them, "Did this actually happen?".

So how does this all fit into the conversation we're having about deepfakes? And so what I'm showing here is a screenshot of a remake of "Home Alone," using Sylvester Stallone's face. So I've got a running search for deepfakes on my computer every day, so I can spot these as they come out. And the majority of the time, I'm just seeing people making hilarious remakes of movies with Sylvester Stallone's face or somebody else's face.

And so what I would say is there are a lot of alarmist headlines that we see about deepfakes wreaking havoc on society. Currently, from my perspective as a journalist, there is no need to panic, I would say, currently. For me, content out of context is the biggest issue, and it's trickier and trickier to deal with. And also worth noting as well, in the light of COVID, the cuts to the journalism industry give me a lot of cause for concern, as we've got fewer people to give the context for these news events and fewer people to correct the record.

And the second thing I'd say is I'd encourage people to see this as part of one aspect of the disinformation problem that we have. And the insertion of doubt is often as effective as convincing someone of a different narrative. And that's a catch 22, I think, of talking about disinformation, as people know, because it makes people doubt what they read, what they hear, what they see. You think it's a bigger problem than it actually is. But for those in the journalism industry, I think we have to be really careful about how we talk about disinformation, and while it is a problem, we're also possibly unintentionally making people believe it's a bigger problem than it is or further lowering trust. So just be very careful about how we talk about deepfakes and cheapfakes and everything else, and how big a problem it actually is.

And finally, for journalists as well, it's also important to understand fakes and manipulated content as part of the misinformation ecosystem. So we're constantly monitoring all corners of the web because spotting coordinated campaigns early is kind of key to dampening their impact sometimes. So for all these examples I spoke about, technology isn't solely the answer to verification. It's relying on the basic principles of journalism that you've learned all the way along. Reporting on the facts and letting the audience know what you can find out, and what you can't. And that's it. Back to you, Claire.

Claire Wardle Thank you. That was great. You said so many things. I found myself just nodding along, nodding along. So let's go to our third, but also great speaker, Christina. I'm not going to even try again to say your last name. I'm so sorry. I'm one of those terrible British people that can't speak any other languages. Christina works for Reuters. And similarly, Reuters has a great course online, which I'm sure Christina's going to talk about in terms of how to spot deepfakes. But again, I know she has some great case studies to help us understand this issue. So over to you, Christina.

Christina Anagnostopoulos Thanks, Claire. Hi, everyone. Thanks for having me on this panel. I'm very excited to share the stage with the others here. So I currently work at Reuters leading our fact-checking operation, so that's verifying social media posts and content, video, photos based in the U.S. Before that, I was working on Reuters UGC team, that's user generated content, so that was verifying pictures and videos in breaking news situations.

So what I hope to bring to the panel is some thoughts from the experience verifying visual content from the front lines. And both our UGC work and our fact-checking work requires us to verify visual content every day. In UGC, we encountered a lot of cheapfake videos. They've certainly popped up in fact-checking work as well. In answer to the title of this session on the challenges of verifying cheapfakes and deepfakes, I'd say the key challenge facing us as an agency is both UGC and fact-checking are really time sensitive. And at the same time, we require immense precision and accuracy as a news agency, where not just our reputation is at risk, but thousands of agency clients around the world who rely on us for accurate information that has already been verified once it gets to them.

So a lot of you might be aware that Reuters did a course on video verification. The course was published last year. It's available in 16 languages now, and it's an excellent dive into how Reuters tackles verification of visual content and what we've learned so far.

We identified five types of media manipulation when we talk about video. So that's lost context, edited, staged, CGI-modified and synthetic. So the deepfake is the last one, and the first four would be examples of cheapfakes. In our daily work, we've mostly encountered all of these except for the deepfakes, that we know of. The vast majority of the videos would fall within the first two categories, so the lost context and edited categories.

To talk a bit about deep fakes in our line of work. The 2019 reports showed that 96% of deepfakes are still used in pornographic content. Right, so it still remains scarse in breaking news, misinformation landscapes. And so people might ask "Why bother putting all this energy into deepfake studies?" And I think it's because they might start appearing. And when they do, we want to be ready for that. Practice is really important when we prepare, so the more we familiarize ourselves with deep fakes, and Reuters made a deepfake as part of our manipulated media course, the easier it will be for us to identify them if and when they come up.

So for now, I think our policy to verify deepfakes if they show up would be very much the same way we do with other video content. So that's a human verification, you know, asking about the source, asking to get the original file, asking the user for the motive for sending out the video, etc..

So another interesting trend that I think you touched upon, Claire, is the potential we're seeing with people using deepfakes as a way to discredit actual content. So we've seen people jumping to label something as a deepfake when in reality it isn't, just with the purpose of trying to discredit that content. So that's also something that we're keeping our eye on, that kind of plausible deniability over something that's actually true, but that people are labeling as a deepfake.

To talk about cheapfakes a little bit. So cheapfakes are something we encounter very often. These videos are challenging because they're really easy to make, and there are so many of them. And as users around the world tech literacy increases, it becomes increasingly easy for anyone to bring manipulated video into the internet world, into the social media world, and it's quite harmful and easily scalable.

So I want to share a few examples that we've come across in the fact-checking work recently of these kinds of cheapfakes. So the two videos that you'll see here are two examples of lost context video. One on the left is missing context. So we saw, for example,

a terrifying trailer for the movie Contagion being shared as an upcoming new film. So this was to prove somehow that there was a conspiracy that Hollywood and the government knew that this pandemic would hit this year. And the movie is actually from 2011. Another example that we saw was imposter content because of lost context. So we saw a video and screenshots from the apocalyptic movie World War Z being shared as if it were MSNBC footage, and posts where accusing MSNBC of manipulating viewers and so trying to limit trust in mainstream media. But the video was actually doctored. It was never actually aired on MSNBC.

A couple of good examples that we have of edited video recently, as well, you'll see now on the screen. One was a mixture of staged and edited video featuring Donald Trump. So a video of a voice actor, comedian, imitating Trump's voice to perfection was used over actual Fox and Friends clips. It was a super realistic voice. I definitely fell for this one. But it was saying totally fabricated things, and it was viewed over a million times on Twitter. The example on the right was a video that we saw of Joe Biden on the popular daytime talk show The View. This one had over seven million views on Twitter and loads of iterations on Facebook. When Biden is asked about a recent Me Too allegation, his answer is confused, jumbled. It makes very little sense. The View host can be seen staring at him, super awkwardly speechless. In reality, the video was completely edited, so Biden's answer was cut to appear more ineloquent and separate clips of The View hosts were placed into the video to make them look more bewildered than they actually were. If we look at Biden's answer in the actual original interview, it's way more natural, and he actually answers the question that he's asked. But there was very little indication on people's comments and people posting this video that they understood that it had been edited. So that's where we would come in with a fact-check on social media.

So to finish off, I just want to talk about a few trends that we're seeing in manipulated content and visual misinformation. So especially as a fact-checker, I think, is that these examples are often rooted in some level of truth, and then taken further into harmful territory. So they stay within what might be believable so that users fall for it, but they take it a step further. So a fake video will hinge itself onto some level of truth.

So, for example, Biden has been known to stutter when he answers interviews, and they take this a step further into modified territory. So let's make Biden's answer in this interview appear so confused as to damage his perception among potential voters on a topic we know matters to this potential group of voters, which is Me Too, in this case.

So I think it's important for us to step back and look at these trends, to prepare us for the challenges in verifying. For example, some patterns we've seen amidst this pandemic in the last few months is there might be some truth in vitamin C helping cure coronavirus, but the post will say "vitamin C cures coronavirus completely." And so people might not go to the hospital. Some other posts we saw were saying "don't trust Bill Gates" in the middle of a global pandemic, where the Gates Foundation is very much involved in a global health effort, because he flew to Epstein's island 17 times. So there's some truth in that. Flight logs did show that Bill Gates flew on Epstein's plane once from New Jersey to Florida and that he met with Epstein to discuss philanthropic opportunities. But there's no evidence in all the flight logs that we reviewed that he flew to Epstein's island, and certainly not 17 times. But in the meantime, thousands and sometimes millions of people have viewed this misinformation.

And these impressions have a power to really influence society. So whether it's someone's health care decision, like whether to get vaccinated or thinking that COVID isn't real, or

impacting elections by fueling fake news on a certain candidate, or through conspiracy theories, which we're seeing a lot of. The misinformation, the manipulated content, whether it's targeting global health decisions or political elections, the importance of verification and fact-checking, I think, is more important than ever. And for all of us, even just questioning the videos that we see online, especially during a socially or politically tense moment like this year is and has been and will continue to be, I think, is more important than ever.

So thanks. That's all from me.

Claire Wardle Great. Thank you. And thank you also for reminding us all that actually the challenge right now is around pornography. And so people think, "Oh, I'm worried about Trump's deepfake. This actually is mostly about women having their images used and taken. In many cases, they don't even know that their images have been taken and used in pornography. So it's just a really important reminder that when we're flippant about "oh, deepfakes aren't a problem," they are a huge problem for women, and the way that they're used to harass people.

And the second point, thank you for also talking about audio. If there's anything that keeps me up at night, it's about impersonation. And in the Brazil election in 2018, and it was an impersonation of Bolsonaro that really caused problems. And when you watch Sarah Cooper do her impersonations of Donald Trump, I'm like, she could really cause some problems if she wanted to.

Anyway, thanks for your three presentations. They were excellent. And we're going to open now for Q&A, so thanks again.

Hello. I think those of you who have been watching ISOJ know that this is a trick that we actually recorded it a week ago. Therefore, I have not had a massive haircut. I have my hair up. But it was wonderful to watch those presentations and to hear the conversation. And thank you to those of you who have put a couple of questions in. Please add more. We've got a good half an hour, which is a great amount of time for Q&A.

So I just wanted to start with a question that came in quite early. It's quite technical, so I'm going to give it to Matt. But it's actually talking about this question that is a consideration to the point that with the technology that we have around detection, how relevant for the verification of deepfakes are like microexpressions? So when there's tiny movements in somebody's face, what does that mean in terms of when we're looking at deepfake content?

Matthew Wright Right. So that's a great question. What I would say first is that in terms of our tool and most of the detection techniques that have been proposed so far, those wouldn't necessarily matter. Sometimes what we're looking at might be more about the video manipulation itself. So, if the face swap case is easier to understand, that basically you're taking someone else's face and you're swapping it onto someone else's head, and then we've got some merging that needs to be done at the edges of the face. And that itself can be detected to a degree. So in those cases, we don't need to look at microexpressions in particular.

Where micro expressions come in handy is when you've got the other techniques that look more like the biological signals. So looking like the model has actually learned something about Obama's microexpressions, essentially, and patterns of movement, and then looked

for ways in which those are being violated. The other thing the microexpressions are used for is if you are looking for whether this is a deepfake, so if you can look at the expressions and say, "Well, looks to me like something might be off here." Comparing it to original videos of the person speaking might give you a good sense for whether that might be a deepfake or not. But that's hard to rely on, of course.

Claire Wardle Right, thank you. And I'm not going to name people's names because these things go on the internet, and you don't know where they'll end up. But we do have a question from somebody who is in Lahore, Pakistan. So goodness knows what time of night it is in Pakistan. But thank you for joining us. And the question is a great one, which is what is the role of users in fighting against deepfakes and cheepfakes, and how can journalists educate the masses on this?

So I'm going to start with Rhona on this one, because you mentioned this, which is part of the challenge here is that it allows journalists to say, "Oh, I didn't say that. That must be a deep fake." So it feels like we have to talk about deepfakes to educate people, but by talking a lot about deepfakes, it makes people think, "Oh, that was probably a deepfake." So I don't know, Rhona, if you have any thoughts about that? What should the role of users and journalists be in this space?

Rhona Tarrant Yeah, I mean, for users, it's probably practicing skepticism without being too cynical about everything you see. But I do have strong feelings about how newsrooms could probably up their game on this. And I think there's a massive knowledge gap in newsrooms, as you know yourself, you work with newsrooms. You know there's a big legacy of certain types of reporting, but online reporting is quite new. And I do feel there is a gap in editors knowing how to deal with this sometimes. Some newsrooms are fantastic. Some are not so good. But the thing that drives me mad is newsrooms instead of calling something out or doing the legwork, they say, "You decide. So here's a video. It appears to show this. We haven't confirmed it, but we're putting that out there for you." In my opinion, I don't think that's very responsible. There's certain language. "You decide. This footage appears to show." When none of the journalist legwork has been done. So I think newsrooms have a long way to go in educating themselves on this type of work.

And the knowledge gap in newsrooms is a big problem, especially when there are huge cuts to journalism. And so I'd say and like you say, it's an excellent point, not overblowing deepfakes and not overblowing cheapfakes either. And another kind of annoyance is the alarmist headlines that you see. So "it's going to wreak havoc on society." And I think it's a great line. It might get some clicks, but it's not necessarily serving your audience. So I'd say really there's a big onus on newsrooms to up their game on this stuff.

And there was an example during the week where a reporter was accused of saying something to the press secretary in the U.S. in a press briefing. And that story, what we did was we just slowed down the audio and we went to the official transcript, and you could tell she actually hadn't said that. I don't want to repeat it, but she actually had not said that. But a lot of newsrooms went with "you decide," which is not doing the journalism. So it's really on the newsrooms.

And in terms of people, it's practice and skepticism and knowing who to rely on, knowing the reputable newsrooms to rely on.

Claire Wardle That's great, and I think, Matt, wasn't it right that the poll was 50/50 with the people who did it? So if it's like "you decide," it's not necessarily that we are all talented

enough to do that. And so, Christina, you mentioned the course from Reuters, which is excellent. And it's been translated into a number of languages. But what was the kind of feedback about that course that you guys did? And do you think that there is more training that we could be doing?

Christina Anagnostopoulos I think it was generally a really useful tool for newsrooms, but again, it's a course that's available for anyone online. So I think it's about having the conversation around news include fact-checking and verification as well in the more daily conversations about news in general. Pointing people towards these resources that they might not know exist, I hope that does that today a little bit. So I think it's about highlighting the work that fact-checkers and people who verify video like us are doing. And it's good that it's a very interesting topic. People love to talk about deepfakes and cheapfakes and election meddling and all these kind of thing. So I'm glad that that's happening because that means there's a conversation going on that's going to hopefully lead people to resources like this conference, this course that we have, all kinds of things, articles about this.

I think the question was also talking about how users can help fight the misinformation. And it is worth noting that both in UGC and fact checking, we get a lot of help from people online. As many bad actors as there are, there are a lot of good actors. There's a lot of people under videos that say, "Wait, no, I was there. I was on the scene. This didn't happen." Or "here's my video showing a different angle," which then helps us geolocate if we talk to that user. So I think that community can be harmful, but the social media community can also be really beneficial for us as fact-checkers. So I think it's good that, you know, we have to filter out what's fake, but I think it's generally good. And we often get tips.

I mean, the the Trump video that I talked about with the Fox and Friends interview, very shortly underneath that tweet, it had gone pretty viral already, I think, like half a million views, and one of the first tweets with a lot of interaction was, "Wait, is this real?" And then people started debating if it's real or not, and what they thought. And I think this is really fascinating as well as a fact-checker, because they often give us amazing tips.

So I like that user-based feedback mechanism, and Facebook has that as well actually. If people don't know, you can you can report something the same way you would report it for nudity or obscenity, you can report something as possibly false. And that is an amazing tool, I think, because it sends that content straight to fact-checkers. So I think it's quite a nice thing to be able to get the community of users involved.

Claire Wardle And Christina, somebody actually asked whether the course is still available, which I think it is, so you might want do a little bit of marketing and drop the link into the chat so that people can take the course.

And we also have a question from India, which I'm also astonished by, which is a great question. Of course, in many countries we have relatively low levels of literacy. And so in those spaces, WhatsApp or kind of things that pass on a mobile phone, it's hard to see the detail. We can talk about deepfakes, but people are really going to struggle to see the clarity. What kind of ideas do you think we can have about reaching communities that really literacy is lower, yet they are potentially spreading this kind of misinformation?

Christina Anagnostopoulos I'm happy to answer. I have a quick answer. So it is tricky. I know that WhatsApp. I don't want to speak for WhatsApp, but I remember hearing recently

that they have done a lot to help fight misinformation with chat response. I think they're chats that users can sort of message with possible fake information. So if they're wondering if X state in India is actually doing certain COVID testing, they can send a message and these sort of bot replies have put together a bunch of accurate information on COVID testing, for example.

So I think social media platforms are each doing a little bit because they realize that it's hard to communicate the fact-check message across. I can say that on Facebook. I mean, the impact that fact-checkers have is that our fact-check gets applied to a piece of fake content. So if any user anywhere that's using Facebook, anywhere in the world, that scrolls past a piece of fake content will see a false news warning, and they will be directed to our fact-check if they want to read it. So I think that the social media platforms are doing a relatively good job in terms of applying warnings and fake news warnings onto fact-checks.

I think what might be harder is sharing resources that you need a high-speed connection or access to certain publications to be able to read them. That might be harder.

Rhona Tarrant Right, and there's a difficulty there with WhatsApp for journalists to actually figure out what's going on in these closed networks, and that's been a big problem, obviously, over the last few years. I'm sure you've been dealing with that, Claire, as well. And I spent quite a lot of time in India speaking to local journalists, and that is a huge problem, actually, even tracking this stuff. At Storyful, I know during Indian elections, we had set up WhatsApp groups where people could just forward us what they were seeing, and that gave some insight into what people were actually looking at. In terms of reaching these communities, it was actually a case of working with the local journalists and getting them trained up so that they could work with people in their communities. But still, especially if you're talking about low literacy areas, it's really, really challenging.

Claire Wardle And then there's a question actually here that says this kind of content seems to spike around elections, are there any statistics around that? And I'll just answer that very quickly, is that we have really, really poor data, because it is actually very difficult to define this. What I would say is that the media cares more about this in the lead up to an election, so therefore it feels like there's more of it. But it's a great research question actually, if we were to look globally. And going back to that point that we made in the recording, which is most of this stuff, unfortunately, is women's images being used in pornography. So if we added all of that, we would realize that that's mostly the kind of stuff that we're seeing.

Rhona Tarrant Right. And I don't know, Claire, if you've seen this as well but especially around the midterms or any sort of big event like the debates, everybody is on high alert. So like Facebook is on high alert, the fact-checkers are on high alert, the journalists are on high alert. And we're glued to our computers. And it's actually a bit of a damp squib. It's when you're not expecting it. It's like, fine, there might be more around that time. You might expect it. But actually the campaigns, and the platforms, and everyone's expecting it. It's when they're not expecting it, and there's a breaking news story, and that's when it causes damage, I find myself.

Claire Wardle This is a great question. I'm going to ask all of you to come up with one tool. So if somebody has watched this and says, "oh, I'm certainly much more worried about this than I was half an hour ago," what's the one tool that you would say that people could use for free with their naked eye it would help them with the verification of facial imagery. So I'm going to start first with Matt, then I'm going to go to Rhona, and Christina

is like, "oh, those are two tools that I was going to say." So she's got the hardest one. But, Matt?

Matthew Wright Well, I mean, I suppose I would just start with the reverse image search, right? That would be the default answer, and I get to say it because I get to go first.

Claire Wardle Do you just want to explain a little bit? Where do I find this magic reverse image?

Matthew Wright So you can go to Google image search. Then you can not just search for a particular image, but you can say here's an image, please search for it. And it will look for similar looking images on the web. Sometimes it really comes up with things that are exactly what you're looking for, and sometimes it struggles a little bit and finds things that look broadly similar because AI is looking for things that look kind of like that and hasn't found the exact thing. So it's not 100%, but it can be helpful.

Claire Wardle Great, Rhona?

Rhona Tarrant Are you talking about an image or a video?

Claire Wardle Yeah.

Rhona Tarrant Right, I would say, first of all, and this is not a tool, but I would Google it. Google what you can see in English, but also in other languages, because very often that will actually bring the video up for you. So like "child falls from balcony," and Google that in a few languages. And that's actually a quick trick we would use if we're trying to figure out if it's an old video or not, aside from reverse image search. Or searching Twitter natively or are searching Facebook natively with those words. That can really, really help, and it's a quick and easy one.

Claire Wardle And also Googling signs or telephone numbers that you can see. So earlier we were back channeling on WhatsApp, and I was like, I want to geo-locate Rhona's apartment because of that red awning. And you can't really see enough, so I feel OK saying it. But often, there are a million clues in an image. You can often see numbers, signage. So that can be one of the best things. Christina, any other tools that you would recommend?

Christina Anagnostopoulos Aside from the Reuters course, I think quality. Just if something seems low definition, that's always a good warning sign. And also just like, what's your gut feeling? We always did this on UGC. It's, you know, does this video seem too good to be true? Is there something that just doesn't feel right? Like use your instinct when you're looking at something because there's a reason people are tweeting "this can't be real." Like there's often something, so using that kind of instinctual thought, if something is too good to be true, then it's worth looking into it more.

Claire Wardle Great. And the next question is something that I said in the recording that I'm really worried about, which is audio. So if I'm a reporter and somebody sends me an audio file, like, what's the reverse audio search? Like, what do I have to do if it's audio? Matt, maybe you can start us off talking about audio deepfakes, and then let's just talk generally about the challenge of audio.

Matthew Wright OK, so in terms of the technology, we are fortunately, there is a gap between what can be done technically and what we actually see in practice. So we don't, as far as I know, see a lot of computer-generated audio that sounds really good right now. But I'll tell you that the technology exists. I've heard it. I've seen it used in videos. So it is a future danger. Hopefully further in the future, but we'll see. And then, yeah, but I don't know about actually what tools you folks would actually use for that, so.

Claire Wardle Rhona, I think I know what you're going to say, but maybe you'll surprise me.

Rhona Tarrant In terms of audio, well, I know just during the week we were looking at a particular piece of audio altered from the original, and we used Adobe Audition. We just went back to lining up audio on Adobe Audtion, and like watching the wav files really, really closely. But in terms of audio, I really have no tricks on that. Like, it's it's an extremely difficult one. We're often asked by newsrooms, as well, about what would you do to verify a piece of audio? And it's extremely difficult. Even the metadata can also be changed on us, so that's not something we would rely on. It's more of an indication. And really there are no tricks around that unless you have the original or unless you know the source really, really well.

Claire Wardle Christina, do you have anything to add?

Christina Anagnostopoulos Yeah, so I think it's a bit easier, similar to what Rhona said. It's a little bit easier when there is a video and someone's doing a voiceover because you can look at if the voice is matching the lip movements. But when it's a WhatsApp audio message, for example, or it's a voiceover like the Trump interview on Fox and Friends, that was just an actor that was so, so believable, we had to look for that voice actor, and reach out to him, and confirm that it was his audio. I mean, and it did make sense after a while, because then we listened to his other audios, and it did match his voice. But it was so good, it was such a good impression of Trump. So it's really, really hard. I think you kind of need to find what the source actually is for you to debunk it.

Claire Wardle Yeah, and I would say, and I'm sure you see this, too, but in countries like Brazil, or Nigeria, or India where we've worked, the number of WhatsApp audio messages is significant, particularly in countries where literacy levels are low, and we know WhatsApp has no metadata because it's encrypted. And it's just impossible. And I think that there will be many, many more audio messages in the U.S., and I just don't think people are prepared for them whatsoever. That's the way we miss the spreading. So, yeah, we need more tools and unfortunately we don't have them.

Christina Anagnostopoulos I remember that we had seen this audio. We had heard this audio of this football player, I forget his name, that had died in this tragic plane accident in the English Channel, something like that. And we ended up, I think, contacting people that were part of his team and people who knew him to make sure that that was his voice because we weren't really confident. Right. It was really hard to know. It was a WhatsApp voice note that had been sent so many times.

Claire Wardle Yeah, and the next question I'm going to just answer, because I get pretty passionate about this, which is are universities teaching these kind of verification skills? And the answer, maybe Austin, Texas, is different, but my frustration is these skills are often not embedded in core reporting classes. So people like Rhona, Christina and I will be

asked to come in and do a guest lecture as if it's like "oh this fun special thing." And it is insane.

And I mean, just a little story, I used to teach a little bit at Columbia Journalism School, and the day after the Pulse nightclub shooting, one of my old students emailed me and said, "Sorry, Claire, just to let you know, I never really listen to anything you said in class. But I'm now on the Miami Herald breaking news desk, and I'm asked to verify all this footage that's coming out of the nightclub. Can you send me all your notes?" I was like, "ugh!" Because one of the first jobs that many young journalists get are these kind of breaking news desks, and it's all about how do you verify footage quickly? How do you verify the account?

So if there are any journalism educators on this live stream, please, please, please, it's 2020. Can we start integrating this into core curricula? Because it could not be more necessary. And it's frustrating to me that there's not enough of it. OK? And that was someone from Ecuador, so I'm talking about the U.S. context. Maybe international journalism schools are better, but I just think that we're not there. And just a quick plug for First Draft as well as the Reuters course. We do a lot of training at First Draft. It's all on our website. It's free. We do it in multiple languages to try and fill this gap because a lot of this training just isn't out there.

OK, I'm going to get to this question. It's really interesting, which is this issue of corrections. So everybody famously knows this line, which is by the time the line is halfway around the world, the correction hasn't got out of bed, or whatever the phrase is. What do we all think about this new move by Twitter to add labels to manipulated media? How effective do we think these kind of debunks are, these corrections? And I'd be interested, Matt, maybe starting with you, somebody who's not directly in the journalism space. What's your sense of some of these attempts to do corrections?

Matthew Wright Well, I think it's great that they're doing something. I can't really speak to the effectiveness. I haven't seen any studies about that. I think it'd be really interesting to examine that. But I think it's really important that they are at least trying because eventually some of these things are going to get out that are really more dangerous. And we've got to have some some kind of protections. Typical users, I mean, you can only have so much media training and expect so much of the general public to really be trained to look for potential falsehoods and misinformation, especially as it gets better, as the quality of the misinformation really goes up. Then having some kind of warning for users at minimum is helpful.

Claire Wardle So do you have anything to add, Rhona or Christina, about what's the most effective? I mean, Christina, you're part of the Facebook fact-checking program, right? So your work ends up being flagged on Facebook, so maybe you can't say whether you think it works or not.

Christina Anagnostopoulos I think I'm a big proponent, obviously, but I also think that I'm encouraged by the information we've gotten from Facebook about how it actually affects information sharing. So a big amount of people, I don't remember the exact percentages, when they see a false warning label, they don't click through to the article or the post, and they're also much less likely to share it. So that's huge for me. Because I do believe that, you know, there are some malignant actors sharing this information, but there are also people that just, you know, they share because they think something is real. And so it

helps propel these ideas that they might already hold. But I think most people aren't malignant sharers, malignant users.

So I think that most people, when they would see this, would think twice before sharing it. I'm sure there are people that will share it anyway, because whatever, I don't care if the fact-checkers disagree. But I think a lot of people have a pause, and the numbers that Facebook has shared with us show that people click through less. And that's really encouraging to me.

Claire Wardle And Rhona, this is slightly kind of on the same lines, but this challenge that reporters have, which is how can you correct the misinformation without giving oxygen to the misinformation? Well, I mean, obviously, you work with newsrooms every day. I mean, are there some times when you like, "Oh, I just feel like actually we've highlighted some misinformation, and if we hadn't highlighted it, then nobody would have known about it."

Rhona Tarrant Right, and I think that's probably a challenge as well in that, like, misinformation became very involved to report on. And of course, because it's a massive problem. But then you also have the issue of falsehoods being amplified when a reporter is really just trying to say one thing but then amplifying the other thing.

And I think for us, I mean, we work directly with newsrooms. So, like, we kind of have a luxury in that, like we can provide information to them about what's out there, and it's up to them then to report out to their communities how they want to report that. But it's a huge problem in journalism. How do you report on this stuff in a responsible way without amplifying it? And I think newsrooms have been pretty, you know, have made good progress in this. I think there's a lot more awareness. And it's like any area of journalism, like you have with, for example, mass shootings, where they would report everything. They would report on the manifesto. They would report on the name. And you see that less and less. People are not reporting on the name. They're not repeating what's in the manifesto because they're kind of learning over time. And I feel that's happening with misinformation as well, and possibly a bit slower than it should. But it is a real issue, and I don't think there's a silver bullet to it, to be honest.

Claire Wardle Yeah, yeah, I agree, and I think it's a perpetual challenge, and what that means is that you have media manipulators who are desperate to get in contact with the reporter. So that's a challenge again in a breaking news event. If somebody comes to you on Twitter with information, are they an eyewitness or are they somebody who's hoping to give you the wrong name? I mean, there's a lot written now about media manipulation, and it's a real challenge for journalists. Again, something that we should be teaching. And somebody has just sent us a message to say that she works at Iowa State University and that she will be teaching this to her students, so that's great to hear.

So we've got five minutes left. I just want to end to talk about the human element of all this. It's very easy to focus on the kind of the technological aspects of how they're created and how they are verified. But how does a human fit into this? The humans who create it, but also the humans who share it. How much do we need to understand this through the lens of kind of what it means to be human, and why we click on these things, why we share these things? Because I think that gets back to where we started, which is what's the role of users in all of this? So you just gave me a little smile, Christina, so I'm going to start with you.

Christina Anagnostopoulos Yeah, because I look at really, really horrible things all day. So I'm like, why? Why are you doing this? I think what's what is worrying is the real-life harm potential of things. We've seen threats to journalists become actual gun violence in a newsroom. We've seen fake news that can lead to. I think, for example, a lot of fake news that we've seen around the protests recently, the Black Lives Matter protests, sharing falsehoods about protesters can lead to people being more harsh with them, harming them. We actually did a fact check on "no, it is not legal to plow through protesters," which to me is of course, it's not legal, but like the fact that they were posts on Facebook insinuating that. Like when I saw that, I thought this is more than just a fake headline like people can die. I thought of that Charlottesville incident right away. So I think the real-life harm of, like, this polarizing, harmful content is really real. And so it just gives us that push to want to keep fact-checking it.

I think what I love to read about, and I think it's just crazy and interesting is the conspiracy theories that are coming out recently. And I'm sure a lot of people may have seen the Wayfair conspiracy and all the QAnon conspiracies that are kind of flooding Twitter and Facebook recently. And they are just crazy. And it's difficult because we don't want to amplify when something isn't, as Rhona said, already in the public sphere. But these things are super viral, and they're just crazy. So I never thought we'd be having to tackle conspiracy theories, but they're not fringe anymore. They're quite mainstream. So I try to consume everything I can about QAnon and understanding the psychology behind that I think is scary, interesting, and something to watch, basically.

Claire Wardle Rhona, why is it that the falsehoods travel faster than the truth?

Rhona Tarrant I think it's the old thing of people wanting to believe something that fits into their own worldview. And you recognize that that fits in my worldview, so I will believe it. The problem with online is it becomes much, much easier to only be exposed to things that fit your own worldview, if that's what you're looking out for, with who you follow, or what you're looking at, or what news outlet you look at. And I think it's it's a great question looking at the human impact of all of this. And I suppose like what we're talking about, what platforms are working towards, is making sure that people, when they are online, are not being radicalized or are not not getting into a place where democracy is undermined. And I mean, there's a much bigger picture to all of this hand-wringing that we do.

And I think that's a great point, particularly about QAnon. The platforms this week are taking steps to crack down on that stuff. And we often see kind of belated action to a lot of harmful and dangerous content. And again, I think it's because everybody's learning this for the first time, particularly newsrooms and also for journalists themselves. Like you're saying, "how do you monitor this stuff all day long and not get completely disheartened, disillusioned or go down one of those rabbit hole yourself?" So it's a great question.

Claire Wardle Matt, do you want to add anything?

Matthew Wright Well, one thing that has come to mind through this discussion that gives me a little bit of hope is that it is really difficult to create a good deepfake, not in terms of the technical side, but in terms of you need to actually design what's the script. What's the lie that you're trying to get out and craft all of that in a way that actually makes it so that it spreads virally and gets accepted because it has to fit into someone's worldview, and in a lot of people's worldview, just enough and yet not be true. And finding that, like actually tuning that, is challenging. So that gives us a little bit of hope.

On the other hand, they're making enough of the two different types of cheapfakes and so on, that we see that they're getting lots of practice at it, too. So smaybe they're learning how to do that.

Claire Wardle Yeah, and I think I'll just for a second, I think the other disconnect is bad actors, media manipulators, they understand emotion, and they understand that that's what drives this. And that's what makes humans click and share. But those of us who work in the journalism researching and fact-checking space, we like to think that people have a rational relationship to information, so we feel very uncomfortable about emotion. And so you've got this disconnect, which I think is part of the problem here, which is how do we fight a really emotive meme with an 800-word fact-check? And the question is how how do we balance that? And I think we're still working that out, which I think to flatten the curve, the like that beautiful image, was one of the most effective messages this year, and it meant that it was shared virally. It wasn't 800 words, and it explained a really complex concept.

And so I keep thinking, what's the flatten curve for all sorts of different things. But the other thing about, I think you're absolutely right, Christina, I really, really fear about the rise of conspiracy theories. I've never seen it go mainstream like we've seen in the last four months. But that's because people's lives have been turned upside down. They don't know what's happening. We don't have any clear explanations for where did COVID come from. We still can't really agree on how it spreads, and we haven't got a treatment. So everybody wants explanations, they want to feel like they've got agency. So in the absence of that, sharing a conspiracy theory, that is a simple, powerful explanation is giving them that. And so if we don't, I think because those of us who work in this journalism fact-checking research space, we have to understand those dynamics. Otherwise we'll keep having these conference presentations being like there is a problem. I just I think it's the thing we have to grapple with, otherwise we're not going to get much further.

OK, so we're nearly at the end here, but I just wanted to give everybody just a final comment. We are obviously, we're in the U.S., so everybody's a little bit focused on the election. But any predictions about what you think might happen between now and the 3rd of November 2020 when it comes to deepfakes or cheapfakes? And I'll start with Matt.

Matthew Wright OK, I will. Making predictions is always a fool's game, but I predict that there will be a reasonably high quality, deepfake of either Joe Biden or Donald Trump or both. But I predict that it will be either detected or just seen as clearly being not true, because there's just enough context around what those two people are doing all the time, that it won't really get out as like into the mainstream media and it won't be taken up as real.

Claire Wardle I promise I won't come back and say you were wrong, so this is a safe space, everybody, to say whatever you want. Rhona?

Rhona Tarrant So I also hate predictions, but I would think that in the coming few months, the election is going to get wrapped up with the protests. And in terms of deepfakes and cheapfakes, possibly content taken out of context, I think is a big problem for us and that we're keeping an eye on. And the other thing is coronavirus and misinformation, including vaccine hesitancy, is the next big one that we are gearing ourselves up for. But in terms of manipulated content, it's really the content out of context that I'm looking at for over the coming months, particularly as everything's online.

Christina Anagnostopoulos I can't wait to see the deepfake that Matt's predicting. I'm going to be ready. I think that we're going to see mostly cheapfakes around Joe Biden and Trump. I think they're both unique personalities that lend themselves to very many iterations. There's such public people. I think we might see that with whoever Biden's V.P. is, some kind of smearing, possibly kind of thing. And then separate to the candidates, but election related, I think mail in voting. I'm seeing a lot of stuff that is trying to discredit mail votes. And then that's COVID related. In the middle of a pandemic, it's safety versus whatever. So I think mail voting and manipulated video about Trump and Biden.

Claire Wardle Yeah, and actually on that topic, and I just realized we didn't really go over it, which is a lot of people talk about fact-checking and debunking after the fact, but there's a lot of academic research now that we should be doing pre-bunking and innoculation. So exactly to your point, Christina, we should be talking to audiences now about the likelihood of things they're going to see. Because if you talk about tactics and techniques to say "it's very likely you're going to see a box of ballots in a place that you wouldn't expect them," and that will be done to discredit the integrity of the election. And it's probably going to be from another country, or it will have been Photoshopped, but that will be a tactic that you will see.

And so I think that newsrooms potentially should be doing that because ultimately misinformation thrives when there's a vacuum of information. And so to your point, exactly, the absence of clear information about the COVID vaccine, the absence of and confusion around mail in voting is allowing this to thrive. So, yeah, I couldn't agree more that we should be doing more of that.

And actually, this is a story from a newsroom that was talking about, and you will see often a lot of these at Storyful, but during an election day, you often see a video of somebody clicking the video screen of somebody's name and it flips. And I saw somebody in a newsroom about six months ago say we've done an investigation and the reason that is, is because a lot of the machines are old, and the glue has dried, so between the screen and the computer, that's why that happens. And I said, "Oh, what are you going to do the explanation now?" He said, "No, no, no. We're just going to wait until election night to see if that happens. And if it happens, we'll have something ready." And I was like you should tell people now that that's what they're going to expect. So I think that concept of pre-bunking is interesting. I don't know whether Christina or Rhona if you're doing any of that kind of stuff?

Rhona Tarrant No, just on just on that. For the night of the midterms, actually, I remember we were on high alert, and we saw some things that we were like, "Is that a deepfake? It that a fake sign?" But actually it was the traditional journalism of literally calling the polling stations to say, "Did this happened there?" And they said, "yeah" or "no." So we're not actually using any big technology or anything. It's just the traditional journalism. But like just diving in and just figuring out what we can.

In terms of pre-bunking, I think it's just for us, we alert newsrooms to the trends and narratives that we're seeing, and we're trying to do that early, if we can, particularly to keep them up to date for, for example, like I said, with the vaccine stuff. So what are the trends online? When was that happening? But I think that's a great idea. And I think particularly for readers, I think it's a great service that newsrooms could do, explaining certain narratives and what the intention of those are.

Claire Wardle And yes, and another little plug, actually. We've just started a 14-day text message course, which is aimed at the public, and every day there's a little like just text that you get at the time that you want it. And we're about to do a randomly controlled trial to see if it makes a difference. But it's all about pre-bunking. It's about teaching people the tactics and techniques of media manipulation. So, yeah, I'll drop the link in. But it's interesting to try and get people to think about those things.

And the last thing I wanted to say before we wrap up is that I feel like my talking about universities has made people get a little bit like, "Oh Claire!" And you're absolutely right. It's very easy for me to sit here on my Zoom and say, "This is what universities should be doing." And the point that somebody made is absolutely right, which is journalism is moving so quickly. How can you suddenly have a faculty that are experts deepfakes, experts in verification, experts in augmented reality? And I completely understand that, and we are still very happy to do guest lectures.

But if Rosental is listening, I do think that there is a need for a train the trainer course, that we could bring on a number of academics, one from each of the major journalism schools, and run a week-long course on the things that we're talking about here. So that people feel confident to teach it, because deepfakes is a little bit niche. But the stuff that Rhona, and Christina, and I do every day, is not so niche anymore. And we should ensure that there are enough journalism professors that can teach it. I would absolutely help to put something like that together, because I think it's really necessary.

So thank you for pulling me up on that point because I didn't want to be flippant. I know it's really, really hard to teach journalism. But I've been doing this work for 10 years, and I just think there's been a lag in the journalism education space around a topic that's so important. So Rosental was listening. Look he's popped up. His ears were burning.

Rosental Alves I mean, I love the idea. So we have a project to work together after this ends, so do you want to wrap up?

Claire Wardle I think we're done with a great panel. So I'd just like to say thank you to everybody for their insights and their great presentations, which will be available afterwards.

Rosental Alves Exactly. Thank you so much. This was amazing and so important. I think, you know, with these predictions, et cetera, you put this in the proper context. That it's super, super important. I'm so happy with ISOJ, and I think this is a grand finale in all aspects of the importance of this.

I want to take a moment to give an incredibly huge thank you to our sponsors, the Knight Foundation, Google News Initiative, Microsoft, Univision, JSK fellowships at Stanford, the Trust Project and the Moody College of Communication. Huge thanks also to everyone involved. We had 84 speakers like you. They were so dedicated. We had the pre-recording then come out live. So thank you so much to all the speakers. Thank you very much to Veritas Group, who B.A. Snyder and Grace and Suzy. This team is fantastic. And without them, this would not have been possible. Also, the Texas Student Television that did all the professionally edited part of it, the Moody College of Communication tech teams, the people who were behind it are also. The interpreters led by Steve Mines and the Knight Center team.

I mean, this is so complex as an operation. Much more difficult than the organization of an in-person conference. So I hope you enjoy this marathon, this week long conference. All those videos are going to be available on our YouTube channels. And I hope we can all come back to Austin next year.

But we are not done completely. We have a party. We have a party that is very special, and that's sort of a 3-D party that you get an avatar. So we are creating this. This is going to be an experiment. We all are going to be with our avatars going around our beautiful building of the Moody College of Communication. We are going to walk in the lawn of the University of Texas Boulevard. So in just a few minutes, we're going to be there, but you can just go to ISOJ.org, go to the ISOJ 2020, and you're going to have the link there, and also some instructions that you can find.

I know everybody is going to be learning how to use this 3-D environment, but I think it's going to be a lot of fun. So I hope to see you, or your avatar, there on our party. So thank you very much, and I hope to see you in Austin next year in April. Bye bye. Thank you so much.