Multi-stakeholder engagement and corrective measures to contribute to a transparent and inclusive online ecosystem

[00:00:18] **Speaker 1** Welcome to the model FUT. The last one of these open online course or information and elections in the digital era. You know this course is organized by the ninth Center for Journalism in America with the support of school. And even my name is Albertina Peter. I'm an electoral expert at Unite School and the lead instructor of the Stream. I want to thank you all very much for being here for the last weeks. We are very happy that you're finding this course interesting and useful and this is very rewarding and we are going to continue in this direction. Thank you very much for your comments, for your participation and again for being there. Okay. Let's start on your last one. We're going to talk about the multi-stakeholder engagement and corrective measures to contribute to a transparent and inclusive online ecosystem during the elections. In this final module, we will address more practices that can help to tackle harmful, harmful practices during elections. We will examine ID, check checking and monitoring measures for problematic content, electoral related media monitoring and track checking. Corrective measures like flagging labeling and blacklisting. Technical and algorithmic responses. Regulatory and regulatory measures. And content. Moderation. So let's start by identification, fact checking and monitoring measures. Modification is a traditional practice of professional journalism. It consists in double checking sources of information, looking for primary evidence from eyewitnesses or double checking facts and figures. It is a way of doing quality control of the content before they are published or disseminated. On the other hand, fact checking happens after the content is published or disseminated, and fact checking looks to make public figures and institutions and news outlets accountable of what they are disseminating. Fact checkers look for trustworthy sources that can confirm or negate claims to the public. More and more fact checking also involves proactive debunking. This means publishing debates and the evidence demonstrates falsehoods, often by explaining the process involved in switching to these conclusions. Electoral related media monitoring implies a program for surveying news and the related content disseminated on social media during a certain period of time, usually attached to the electoral calendar. So you can see it's crazy that fact checking, debunking and verification are three ways of corroborating. The quality of information circulating on the networks are very useful for electronic processes. Let's see those three steps of fact checking. As we said, fact checking, this content analysis driven by one basic question, which is how do we know that this is true or false? And and for fact checking, for efficiency, fact checking, we need to follow three, three steps. So one is determining which quantum scan must ought to be structured. That's very important. Finding the facts by looking on the best available evidence regarding the claim at hand and correcting the rate when evaluating the claim in light of the new evidence, usually when it's less stressful. So that's how usually professionals are taking this done. Following these three steps, the verification of the source and of the source and the origins of visual content can be very challenging. Even more challenging than creating content. There are different types of common false or misleading visual content. You have the wrong time from place images that are really common, which is the content that can be easily, very easily debunked because it's just photos that are used from other other events and it's just easy to go to the archives. Then you have manipulated content. That means content that has been digitalized and manipulated using photo editing or video editing software. And the last piece of stage content, this is content that has been created and shared with the intention of misleading. It's just a stage content like acting or preparing or doing a set is setting where the information that is recorded or photographed is totally false. There are many different tools available to debunk the usual misinformation and misinformation. There are various image search that allows to see if that image is in some databases or where it has been. Archive

in other versions. And that's very easy to do even with Google. And then you have to look and you look at location, which allows determining where the video or image was captured. Often edge a location requires cross-referencing visual characteristics and landmarks from the content with satellite image imagery, street view, visual content available. So it's like crossing different kinds of images and can give you a look at any location and then you have image forensics. Some tools allow to detect inconsistencies in image metadata suggesting manipulation, I guess of running actually of the image as an example of our global response, we can mention the International Fact-Checking Network and at Poynter, and that is one of the most important fact checking global networks and very committed to promoting excellence, in fact, checking for what they have created and develop a code of principles that you can see on the screen. There are like five important principles for professional and reliable check checking. One is to be committed to nonpartizanship and firmness, transparency of services, transparency of funding and organization, transparency of medical, voluntary and open and honest corrections policy. Even if they if they get it wrong, they need to tend to the public in order not to lose credibility. So these are the five most important principles for fact checking. But let's see. Électorale fact checking and media monitoring. Let's take a closer link to this issue, to electoral media monitoring. It is a long story, a long history, three long history of electoral monitoring and usually takes place for a short period of time. For instance, during electoral campaign or other, the other phase of the electoral cycle, like being post-electoral, post-electoral period or during elections, the independence events. But usually for a short period of time it has the potential of being implemented in other cases too, of course. For instance, let's say during the also the registration process, the methodology of traditional media monitoring followed, the predetermined set of media can be combined with automated tools. But this is also good for to carry out monitoring exercises involving the human component. And that's that's a combination of mythologies. Electoral related media monitoring usually translates in a sample which defines which media is going to be monitored when, how and why, and usually in both qualitative and quantitative indicators. The qualitative might be reflected in keywords and quantitative indicators might be reflected on and determine graphics and data and can be processed to to arrive to conclusions. Media monitoring requires funding from infrastructure for infrastructure. It's very expensive to turn on the hardware and software. Of course, there are other ways of doing media monitoring much more, let's say, easy and doable. But in general, it is it's an expensive process usually. And sometimes most likely you are going to have a third party that might be involved for automatic monitoring on human resources and threatening. It's also very, very useful. The media monitoring is very useful and very effective when combined with early warning mechanisms and platforms. And we have seen that in many countries. There's samples and indicators. Ideally, a series of keywords identify that function as indicators of the level of hostility or hate speech present in the electoral periods. Media monitoring reports are extremely useful, particularly when combined with early warning mechanisms and platforms. As I said before, and fact checking can be part of the media monitoring exercises, complementing the set of indicators. So there are new technologies and multiple analogies that can be combined to get to get better results, to identify results and be able to address the issues related to the new technologies and the digital data. So so far, so good. We're going to make 5 minutes break. And after that, we will continue with the corrective measures and see you in 5 minutes. My. Welcome back. We are going to continue now taking a look. Two corrective measures as part of their efforts to increase transparency and raise awareness of the increasing threats to freedom of expression. All nine major Internet companies have started putting in place a series of corrective measures. These can be the result of the unilateral initiative of these companies, or it also can be the result of the legal framework in place in a certain in a specific country or region. But these corrective measures are very relevant to electoral

processes. Social social platforms, including the very well-known ones like Facebook, Twitter, Instagram, Snapchat, etc., have put in place complaints and appeals mechanisms which function based on the agreement individual signed when they use, when they apply, when they decide to use these services. These mechanisms theoretically trigger internal corrective measures from the side of these companies. Three of those online platforms usually allow users to flag almost as follows falls, violent, offensive, etc. The labeling allows users to identify dark advertisements, particularly important when it is related to political propaganda and blacklisting can apply even removing content or a particular user from a social media platform. Nevertheless, all these practices should not affect the right to freedom of expression and the right to access to information, which leaves a very delicate balance and very challenging to measure and sustain. This is quite the key issue of all of these measures. If you will see on the screen an example of flagging in social media, Twitter, which is very popular for political debate during elections. You can see on the screen how the issue is reported. Colony Issue three Technical algorithm responses for semi automatic detention of online disinformation and their practical utility across the life cycle of disinformation campaigns, including content answers, credibility analysis, network spread, measuring impact on citizens beliefs and actions and debunking methods. These technical responses can be implemented by social platforms and search engines themselves, but can also be third party tools, for instance, browser plugins or experimental methods from academic research. But in any case, internal contract regulation of private companies that are often automated with limited human reviewers are not always very transparent or easy to understand by everyone. Most automatic tools for disinformation detection are currently aimed at providing input to human decision making. At the contact level, they provide information to enable human analysis and manipulation. The assumption behind technical and algorithmic approaches is that they can curb and reduce the sharing of disinformation or misinformation. Internet Archive Council Contents. Internet companies deploy a range of automated detection models for content types and services. These include tools for tricking the artificial spread of information, as well for identifying content that meets criteria for ranking, labeling or removal. But this is very challenging, as I mentioned before, as the content of the content circulates in thousands of languages and context is not always easy to access automatically. Let's see now the regulatory and regulatory measures, because combination of length content during the electoral period is also a very complex issue as it can threaten the rights to freedom of expression and the access to information. And the process of developing laws and regulations should follow an open, transparent and participatory approach also. For this reason, one of the more countries are trying to enforce certain regulations during the electoral processes because the situation is so complicated to ensure that free and fair elections are not undermined. But at the same time, these measures need to be compatible with the Article 19 of the Freedom of Expression and International Covenant on Civil Rights and Rights. However, most legislation and rules governing elections and related media self-regulatory tools do not always apply to digital platforms, social media and social media. So that's another another challenge. For instance, the silent period that many countries enforce immediately before Election Day and not all before or even after are not always respected by social networks. Even if a government prohibits the publication of polling results before an election or during polling day. Social media users might make these results public earlier without respecting the electoral rules. Likewise, political advertising often spreads online and via social message and even during the silence periods. And what is problematic here is that this content, as we saw in previous models, may be formed within the data and the polls and made up results and can negatively affect and discredit a whole electoral process. Also, as we saw in previous models, these actions can also be part of a wider disinformation campaign orchestrated to destabilize and undermine the credibility of democratic institutions and a particular government in a particular country. Legislation

gave. This information has been developing during the recent years, and many countries have adopted legislation against disinformation such as regulation of content through blocking or removal, criminalization on the grounds of defamation or hate speech, regulatory or legal mandates to monitor social media and data privacy laws, and also political finance regulations that pertains to social media regulation of a specific technology companies or social media platforms during the specific periods of time. Nevertheless, regulatory initiatives have been very criticized in some cases because sometimes they tend to give disproportionate power of censorship to the state or to a government. And even that means sometimes to give that power to a particular political party in detriment of the other ones. So any attempt to regulate online content should be balanced to balance the rights of freedom of expression and access to information with the protection of other civil and political rights, such as political participation, privacy and freedom from discrimination. Restrictions to online freedom of expression should meet the three parties outlined in the Article 19 of the International Covenant on Civil and Political Rights that consists in. Laws and regulations should be prescribed by law. All attention to a legitimate aim and may need to be necessary and proportionate. Self-regulation is a mechanism that is totally different. It is voluntary compliance sector, industry level, where legislation does not necessarily play a role in enforcing the standards as we saw. There are companies that put in place measures by themselves. So in other words, these are private initiatives for content, moderation. But what is exactly content? Moderation. Content. Moderation is the process of monitoring. Whether content submitted to a website complies with its rules and guidelines and is suitable to appear on the site if it's suitable according to the legal framework in place, the area in the country or in the region, it involves setting rules and guidelines of all content. Appearing on the site must abide by filtering out anything that is deemed harmful, sensitive or inappropriate. Essentially, content moderation helps to ensure that the content published on the site is not illegal, inappropriate, not harassing or harmful or harming others. It is commonly used in applications that rely heavily on content generated by users such as forums, social media platforms, dating sites and online marketplace. Mainly social media networks and other Internet intermediaries have focused on developing self-regulatory rules, usually in the form of a code of the standards of of terms of service on or on issues such as content removal under the processing practices. A very interesting example of one of these mechanisms is a major oversight board that is an independent content moderation body. In 2020, it created an oversight board for Facebook and Instagram, composed of highly renowned human rights and freedom of expression personalities. The main idea was to have an independent body to handle the most complex moderation decisions. Initially, in the narrow remit of what content is taken down by the company. The ball accepts cases from users who have exhausted the appeals process on Facebook or. They examine the issue independently for them, make management interviews and make decisions, for instance, of content. And they also deliver policy recommendations that to that Bush made that improved the rules and to act in a way that is equitable, transparent and and treats all users correctly. So this is an interesting example. And finally at the end to finalize our nuke. After these five weeks, we are going to talk a little bit about the hybrid regulatory approach of electoral process for electoral processes. This is a very specific approach that is not possibly to apply everywhere, but are taking having a lot of good results. And it's been very much analyzed. And this is a scenario. In this scenario, the existing legal framework did not does not work. It doesn't impose additional restrictive requirements on social media platforms, but rather said the expected outcomes. For instance, a code of practice is a practice for social media companies in consultation with these companies as well with other political and political parties. A wider need might be developed, review and made and made public and in a collaborative way, creating an ecosystem of much more a stronger ecosystem where every actor is being part of the decisions of the measures taken against harmful contents.

So electoral all electoral stakeholders, we have a role to play in this. We have a role to play in this approach, particularly in electronic monitoring by these media authorities, civil society organizations, the government, and many, many, many times also political parties. This implies also all of these commitments are, of course, voluntary and depend on the existing legal framework. We cannot generalize with these because in many contexts this approach would not be possible. An example of a regulatory regulatory initiative is the adoption of code of conduct and commitments agreed, for instance, by the state institutions, political parties and major social media companies, advertisers, media outlets and other actors. These codes of conduct usually aim at ensuring transparency for political advertisement and strengthening actions to close fake accounts or to stop harmful activities. Labeling messages disseminated by the bots and collaborating with fact checkers in academia to conduct disinformation campaigns and increase visibility and reach to verified content. Let's do a brief recap of our last more of these open online course information and elections in the digital era. During this model, we examine the multi-stakeholder engagement and corrective measures to contribute to transparent and inclusive ecosystem. Many of the differing measures to be able nowadays to tackle disinformation, misinformation and other cultural practices under elections are ID fact checking and monitoring missions, corrective measures and regulatory and regulatory measures, including the hybrid curricula to implement this. Thank you so very much for being there. I'm so proud of being able to have been able to be your lead trainer during these five weeks. Again, I appreciate all your comments, all your suggestions, your contributions. Everything has been very positive and worthy. And thank you so very much. I thought it was good for you and useful to and see you soon by.