

## Multi-stakeholder engagement and corrective measures to contribute to a transparent and inclusive online ecosystem es (Spanish)

[00:00:18] Bienvenidos al módulo cinco, el último de este curso online masivo y abierto sobre Información y Elecciones en la Era Digital. Como ya saben, este curso está organizado por el Centro Knight para el Periodismo en las Américas con el apoyo de la UNESCO y del PNUD. Me llamo Albertina Piterbarg, soy experta electoral en la UNESCO y la instructora principal de este curso. Quiero darles las gracias a todos por estar aquí durante las últimas semanas. Estamos muy contentos de que este curso les resulte interesante y útil. Eso es muy gratificante y vamos a seguir en esta dirección. Muchas gracias por sus comentarios, por su participación y, nuevamente, por estar ahí.

Empecemos nuestro último módulo. Vamos a hablar sobre la participación de múltiples partes interesadas y las medidas correctivas para contribuir a un ecosistema en línea transparente e inclusivo durante las elecciones. En este último módulo, abordaremos más prácticas que pueden ayudar a abordar las prácticas nocivas durante las elecciones. Examinaremos la identificación, las medidas de verificación de datos y monitoreo para detectar contenido problemático, el monitoreo de los medios en materia electoral y la verificación de datos, las medidas correctivas –como la señalización, el etiquetado y las listas negras–, las respuestas técnicas y algorítmicas. Medidas reglamentarias y reglamentarias. Y contenido. Moderación. Empecemos por las medidas de identificación, verificación de datos y monitoreo. La verificación es una práctica tradicional del periodismo profesional. Consiste en comprobar doblemente las fuentes de información, buscar pruebas primarias de testigos presenciales o comprobar los hechos y las cifras. Es una forma de hacer un control de calidad del contenido antes de que se publique o difunda. Por otro lado, la verificación se realiza después de que el contenido se publique o difunda, y busca que las figuras públicas, las instituciones y los medios de comunicación rindan cuentas de lo que están difundiendo. Los verificadores de datos buscan fuentes fiables que puedan confirmar o negar las afirmaciones al público. Cada vez más, la verificación de datos también implica una desacreditación proactiva. Esto significa publicar las evidencias y las pruebas que demuestran las falsedades, a menudo explicando el proceso que implica pasar a estas conclusiones. El monitoreo de los medios de comunicación implica un programa para supervisar las noticias y el contenido relacionado con elecciones que se difunde en las redes sociales durante un período de tiempo determinado, generalmente adjunto al calendario electoral. Así que, como pueden ver en este gráfico, la verificación de datos, la desacreditación y la verificación son formas de corroborar la calidad de la información que circula en las redes. Y esto es muy útil para los procesos electorales. Veamos esos tres pasos de la verificación de datos. Como dijimos, este análisis de contenido se basa en una pregunta básica, que es ¿cómo sabemos que esto es cierto o falso? Y para verificar los datos de manera eficiente debemos seguir tres pasos: 1. El primero es determinar qué contenidos pueden, deben o deberían verificarse. Eso es muy importante; 2. Encontrar los datos/hechos analizando las mejores pruebas disponibles sobre la reclamación [de la información] en cuestión; 3. Corregir el registro de la información al evaluar la reclamación [la información reclamada] a la luz de las nuevas pruebas, por lo general sobre una escala de veracidad. Así es como suelen hacerlo los profesionales. Siguiendo estos tres pasos. La verificación de la fuente y de los orígenes del contenido visual puede resultar muy difícil –incluso más difícil que crear contenido–. Hay diferentes tipos de contenido visual falso o engañoso. Están las imágenes inexactas mal emplazadas en términos de tiempo y/o lugar; es decir, un tipo de contenido que se puede desacreditar fácilmente porque solo se trata de fotos que se utilizan de otros eventos y es fácil ir a los archivos. También están el contenido manipulado; esto es, contenido que se ha digitalizado y manipulado mediante software de edición de fotografías o de vídeo. Y el último contenido es el "escenificado"; contenido que se ha

creado y compartido con la intención de engañar. Es solo "contenido escenificado" [es decir, una "puesta en escena"], en el que la información que se graba o se fotografía es totalmente falsa. Hay muchas herramientas diferentes disponibles para desacreditar la desinformación y la información errónea. Está la búsqueda "invertida" de imágenes, que permite ver si esa imagen está en algunas bases de datos o si está archivada con versiones precedentes. Y eso es muy fácil de hacer, incluso con Google. Y luego está la geolocalización, que permite determinar dónde se capturó el vídeo o la imagen. A menudo, el borde de una ubicación requiere realizar referencias cruzadas en torno a las características visuales y las señales de referencia del contenido con imágenes de satélite, streetview y contenido visual disponible. Consiste en cruzar diferentes tipos de imágenes para obtener una geolocalización. Y luego están las imágenes forenses [o imaginología forense] que son algunas herramientas que permiten detectar inconsistencias en las imágenes, a través de los metadatos de las imágenes que sugieren manipulación. Como una radiografía la imagen. Como ejemplo de nuestra respuesta global, podemos mencionar la Red Internacional de Verificación de Datos en POINTER. Es una de las redes globales de verificación de datos más importantes y más comprometida con promover la excelencia en el ámbito de la verificación de datos a nivel mundial. Han creado y desarrollado un código de principios. Como pueden ver en la pantalla, hay cinco principios importantes para una verificación de datos profesional y confiable: comprometerse con la imparcialidad y la firmeza; la transparencia de las fuentes; la transparencia de la financiación y la organización, la transparencia metodológica, y una política abierta y honesta de corrección –incluso si se equivocan, necesitan ser transparentes y contárselo al público para no perder credibilidad–. Estos son los cinco principios más importantes para la verificación de datos. Veamos ahora la verificación de datos electorales y el monitoreo de medios. Analizemos más detenidamente este tema. El monitoreo de los medios en materia electoral tiene una larga historia [los antecedentes de monitoreo electoral]. Por lo general, se lleva a cabo durante períodos cortos de tiempo. Por ejemplo, durante la campaña electoral u otra fase del ciclo electoral, como el período pre- o post-electoral, o durante las elecciones. Pero, por lo general, se lleva a cabo durante un corto período de tiempo. También tiene el potencial de implementarse en otras fases, por supuesto. La metodología tradicional de monitoreo de los medios, sobre un conjunto predeterminado de medios, se puede combinar con herramientas automatizadas. Pero esto también es bueno para llevar a cabo ejercicios de monitoreo que involucren al componente humano. Y esa es una combinación de metodologías. El monitoreo de los medios de comunicación en materia electoral generalmente se traduce en una muestra que define qué medios serán monitoreados, cuándo, cómo y por qué. Y, generalmente, involucra indicadores cualitativos y cuantitativos. Lo cualitativo podría reflejarse en palabras clave y los indicadores cuantitativos podrían reflejarse en gráficos y datos que podrían procesarse para llegar a conclusiones. El monitoreo de los medios requiere financiación para su infraestructura; es costoso, el hardware y el software. Por supuesto, hay otras formas de monitorear los medios mucho más, digamos, fáciles y factibles. Pero, en general, es un proceso caro por lo general. Y, a veces, lo más probable es que haya un tercero que pueda participar en el monitoreo automático de los recursos humanos y las amenazas. También es muy, muy útil. El monitoreo de los medios de comunicación es muy útil y muy efectivo cuando se combina con mecanismos y plataformas de alerta temprana –tal y como lo hemos visto en muchos países–. Hay muestras e indicadores, lo ideal es que una serie de palabras clave identifiquen que funcionan como indicadores del nivel de hostilidad o discurso de odio presente en los periodos electorales. Los informes de monitoreo de los medios de comunicación son extremadamente útiles, especialmente cuando se combinan con mecanismos y plataformas de alerta temprana. Como dije antes, la verificación de datos puede formar parte de los ejercicios de monitoreo de los medios de comunicación,

complementando el conjunto de indicadores. Por lo tanto, existen nuevas tecnologías que se pueden combinar para obtener mejores resultados, identificar los resultados y poder abordar los problemas relacionados con las nuevas tecnologías de la era digital. Hasta ahora, todo bien. Vamos a hacer un descanso de 5 minutos. Y después de eso, continuaremos con las medidas correctivas. Nos vemos en 5 minutos. Bienvenidos de nuevo. Vamos a seguir ahora examinando las medidas correctivas. Como parte de sus esfuerzos para aumentar la transparencia y crear conciencia sobre las crecientes amenazas a la libertad de expresión online, las principales compañías de Internet han comenzado a implementar una serie de medidas correctivas. Estas pueden ser el resultado de la iniciativa unilateral de estas empresas, pero también pueden ser el resultado del marco legal vigente en una determinada región o país. Sin embargo, estas medidas correctivas son muy relevantes para los procesos electorales. Las plataformas sociales, como las más conocidas como Facebook, Twitter, Instagram, Snapchat, etc., han establecido mecanismos de quejas y apelaciones que funcionan en función del acuerdo que los usuarios firman al utilizar estos servicios. En teoría, estos mecanismos activan medidas correctivas internas por parte de estas empresas. Tres de esas plataformas en línea suelen permitir a los usuarios señalar los contenidos por ser falsos, violentos, ofensivos, etc. El etiquetaje permite a los usuarios identificar los anuncios oscuros, especialmente importantes cuando están relacionados con la propaganda política. Y las listas negras pueden aplicarse incluso para la eliminación de contenido en particular o hacia un usuario en particular de una plataforma de redes sociales. Sin embargo, todas estas prácticas no deberían afectar el derecho a la libertad de expresión y el derecho al acceso a la información, lo que deja un equilibrio muy delicado y muy difícil de medir y mantener. Esta es la cuestión clave de todas estas medidas. En la pantalla ven un ejemplo de señalización en las redes sociales, Twitter, que es muy popular para el debate político durante las elecciones. Puede ver en la pantalla cómo se informa del problema. Las respuestas algorítmicas y técnicas para la detención semiautomática de la desinformación en línea, y su utilidad práctica a lo largo del ciclo de las campañas de desinformación, incluidas las respuestas de contenido, el análisis de la credibilidad, la difusión de la red, la medición del impacto en las creencias. Estas respuestas técnicas pueden implementarlas las plataformas sociales y los propios motores de búsqueda, pero también pueden ser herramientas de terceros, por ejemplo, complementos de navegador o métodos experimentales de investigación académica. Pero en cualquier caso, la regulación de los contratos internos de las empresas privadas, que a menudo se automatiza con un número limitado de revisores humanos, no siempre es muy transparente ni fácil de entender para todos. La mayoría de las herramientas automáticas para la detección de desinformación actualmente tienen como objetivo proporcionar información para la toma de decisiones humanas. A nivel de contacto, proporcionan información que permite el análisis y la manipulación humanos. El presupuesto detrás de los enfoques técnicos y algorítmicos es que pueden frenar y reducir el intercambio de desinformación o información errónea. Las compañías de Internet implementan una variedad de modelos de detección automatizados para tipos de contenido y servicios. Estas incluyen herramientas para engañar a la difusión artificial de información, así como para identificar el contenido que cumple con los criterios de clasificación, etiquetado o eliminación. Pero esto es muy difícil, como mencioné antes, ya que el contenido del contenido circula en miles de idiomas y no siempre es fácil acceder automáticamente al contexto. Veamos ahora las medidas regulatorias y regulatorias, porque la combinación del contenido extenso durante el período electoral también es un tema muy complejo, ya que puede amenazar los derechos a la libertad de expresión y el acceso a la información. Y el proceso de desarrollo de leyes y regulaciones también debe seguir un enfoque abierto, transparente y participativo. Por esta razón, cada vez más países están tratando de hacer cumplir ciertas regulaciones durante los procesos electorales porque la situación

es muy complicada y para garantizar que las elecciones libres y justas no se vean socavadas. Pero, al mismo tiempo, estas medidas deben ser compatibles con el artículo 19 de la Libertad de Expresión y el Pacto Internacional de Derechos Civiles y Políticos. Sin embargo, la mayoría de las leyes y normas que rigen las elecciones y las herramientas de autorregulación de los medios relacionados no siempre se aplican a las plataformas digitales, las redes sociales y las redes sociales. Así que ese es otro desafío. Por ejemplo, las redes sociales no siempre respetan la jornada de reflexión –el período de "silencio" que muchos países imponen antes del día de las elecciones–. Incluso si un gobierno prohíbe la publicación de estimaciones electorales antes de una elección o los resultados del escrutinio durante el día de la votación, los usuarios de las redes sociales podrían hacer públicos estos resultados antes sin respetar las reglas electorales. Del mismo modo, la publicidad política a menudo se difunde en Internet, también mediante la mensajería social, incluso durante estos períodos de silencio. Y lo problemático aquí es que este contenido, como vimos en módulos anteriores, puede ser falso, estar manipulado e inventarse datos y resultados. Y esto puede afectar negativamente y desacreditar todo el proceso electoral. Además, como vimos en modelos anteriores, estas acciones también pueden formar parte de una campaña de desinformación más amplia orquestada para desestabilizar y socavar la credibilidad de las instituciones democráticas y de un gobierno en particular en un país determinado. La legislación contra la desinformación se ha ido desarrollando durante los últimos años y muchos países han adoptado leyes contra la desinformación, como la regulación del contenido mediante el bloqueo o la eliminación; la penalización por difamación o el discurso de odio; los marcos regulatorios o competencias legales para monitorear las redes sociales; las leyes de privacidad de los datos; las regulaciones sobre la financiación política en el ámbito de las redes sociales, y las regulaciones de empresas de tecnología o plataformas de redes sociales durante períodos de tiempo específicos. Sin embargo, las iniciativas regulatorias han sido muy criticadas en algunos casos porque a veces tienden a otorgar un poder de censura desproporcionado al estado o a un gobierno. E incluso eso significa a veces dar ese poder a un partido político en particular, en detrimento de los otros. Por lo tanto, cualquier intento de regular el contenido en línea debería equilibrar el respeto de los derechos a la libertad de expresión y el acceso a la información con la protección de otros derechos civiles y políticos, como la participación política, la privacidad y la ausencia de discriminación. Las restricciones a la libertad de expresión en línea deben cumplir con lo descrito en el artículo 19 del Pacto Internacional de Derechos Civiles y Políticos, que dispone que las restricciones deben estar prescritas por la ley, debe perseguir un objetivo legítimo y debe ser necesaria y proporcionada. La autorregulación es un mecanismo totalmente diferente, ya que es un [instrumento de] cumplimiento voluntario existente en sectores o industrias en los que la legislación no necesariamente desempeña un papel en la aplicación de las normas. Como hemos visto, hay empresas que ponen en marcha medidas por sí mismas. En otras palabras, se tratan de iniciativas privadas orientadas a la moderación de contenido. Pero, ¿qué es exactamente la moderación de contenido?. La moderación de contenido es el proceso de supervisar y monitorear si los contenidos que circulan en un sitio web cumplen con sus normas y directrices y si es adecuado para aparecer en el sitio y si es adecuado de acuerdo con el marco legal vigente del país o la región. Implica establecer reglas y pautas para todo el contenido que aparece en el sitio web, el cual debe cumplir con el requisito de filtrar cualquier cosa que se considere dañina, sensible o inapropiada. Básicamente, la moderación del contenido ayuda a garantizar que el contenido publicado en el sitio no sea ilegal, inapropiado, no incite al acoso, ni dañe o perjudique a otros. Se usa comúnmente en aplicaciones que dependen en gran medida del contenido generado por los usuarios, como: foros, plataformas de redes sociales, sitios de citas y comercio online. Principalmente, las redes sociales y otros intermediarios de Internet se han centrado en desarrollar reglas de autorregulación,

generalmente en forma de un código de normas de términos de servicio sobre o sobre cuestiones como la eliminación de contenido en el marco de las prácticas de procesamiento. Un ejemplo muy interesante de uno de estos mecanismos es una Meta Junta de Supervisión, un organismo independiente de moderación de contenido. En 2020, Meta creó una Junta de Supervisión para Facebook e Instagram, compuesta por personalidades muy reconocidas en materia de derechos humanos y libertad de expresión. La idea principal era tener un organismo independiente que se encargue de las decisiones de moderación más complejas; inicialmente, en el ámbito limitado del contenido que retira la empresa. La Junta acepta casos de usuarios que hayan agotado el proceso de apelación en Facebook o Instagram. Examinan el tema de forma independiente y ajena a la interferencia de la gestión [de la plataforma], y toman decisiones, por ejemplo, sobre eliminación de contenido. También ofrecen recomendaciones políticas para mejorar las reglas y actuar de una manera que sea equitativa, transparente y trate a todos los usuarios correctamente. Así que este es un ejemplo interesante. Y, finalmente, para finalizar nuestro MOOC tras estas cinco semanas, vamos a hablar un poco sobre el enfoque regulatorio híbrido para los procesos electorales. Este es un enfoque muy específico que no se puede aplicar en todas partes, pero que está dando muchos buenos resultados y ha sido analizado ampliamente. En este escenario, el marco legal existente no funciona ya que no impone requisitos restrictivos adicionales en las plataformas de redes sociales, sino que más bien se dirige a los resultados esperados. Por ejemplo, un código de prácticas para las empresas de plataformas de redes sociales –en consulta con estas empresas, así como con otros partidos políticos y el público más amplio– se podría desarrollar, revisar y hacerse pública de manera colaborativa, creando así un ecosistema más fuerte en el que cada actor clave participe en las decisiones y las medidas adoptadas contra los contenidos dañinos. Por lo tanto, todas las partes interesadas e implicadas en materia electoral tienen un papel que desempeñar en este enfoque, articularmente los organismos de gestión electoral; las autoridades de los medios de comunicación; las organizaciones de la sociedad civil; el gobierno, y también los partidos políticos. Todos estos compromisos son, por supuesto, voluntarios y dependen del marco legal existente. No podemos generalizar con estos, porque en muchos contextos este enfoque no sería posible. Un ejemplo de iniciativa co-regulatoria es la adopción de un código de conducta y de compromisos acordados, por ejemplo, entre las instituciones estatales, los partidos políticos y las principales empresas de redes sociales, las empresas de publicidad y marketing, los medios de comunicación y otros actores. Estos códigos de conducta suelen tener como objetivo garantizar la transparencia de la publicidad política y reforzar las acciones para cerrar cuentas falsas o detener actividades dañinas, etiquetar mensajes difundidos por los bots y colaborar con los verificadores de datos y el mundo académico para llevar a cabo campañas de desinformación orientadas a aumentar la visibilidad y el alcance del contenido verificado. Hagamos una breve recapitulación de nuestro último módulo de nuestro curso online, masivo y abierto sobre Información y Elecciones en la era digital. Durante este módulo, examinamos la participación de múltiples partes interesadas y de las medidas correctivas para contribuir a un ecosistema transparente e inclusivo. Muchas de las diferentes medidas disponibles hoy en día para abordar la desinformación, la información errónea y otras prácticas nocivas en las elecciones son: la identificación, la verificación de datos y las medidas de monitoreo, las medidas correctivas y las medidas regulatorias y no regulatorias incluyendo el enfoque híbrido de co-regulación. Esto es todo. Muchas gracias por estar ahí. Estoy muy orgullosa de haber podido ser su instructora principal durante estas cinco semanas. Una vez más, agradezco todos sus comentarios, todas sus sugerencias y sus contribuciones. Todo ha sido muy positivo y gratificante. Y muchísimas gracias, espero que haya sido bueno y útil para ustedes, y, ¡nos vemos pronto!