El reto casi imposible de ofrecer seguridad y privacidad en el **metaverso**

La ONG Oasis ha lanzado una serie de recomendaciones para que las empresas desarrollen espacios virtuales seguros. Los comportamientos tóxicos digitales podrían resultar aún más agresivos en estos entornos y necesitamos formas de proteger a los usuarios sin sacrificar su privacidad.

MIT Technology Review
Publicado por Opinno

TANYA BASU
TRADUCIDO POR **ANA MILUTINOVIC**
24 ENERO, 2022

Internet puede parecer un agujero negro repleto de los peores aspectos de la humanidad. Y, de momento, hay pocos indicios de que el metaverso, ese mundo digital virtual imaginado donde podemos trabajar, jugar y vivir, vaya a ser mucho mejor. Como ya hemos informado, una mujer beta tester en la plataforma de realidad virtual de la red social Meta, Horizon Worlds, ya se ha quejado de haber sido acosada sexualmente.

Tiffany Xingyu Wang cree que tiene una solución. En agosto de 2020, más de un año antes de que Facebook anunciara que cambiaría su nombre a Meta y el enfoque de su principal plataforma de redes sociales a otros planes para su propio metaverso, Wang creó la organización sin ánimo de lucro Oasis Consortium, un grupo de empresas de videojuegos y compañías *online* con la idea de desarrollar «un internet ético donde las generaciones futuras sientan confianza para poder interactuar, crear y existir sin el odio y la toxicidad *online*».

¿Cómo? Wang piensa que Oasis puede garantizar un metaverso mejor y más seguro al ayudar a las empresas tecnológicas a autorregularse.

A principios de este mes, Oasis lanzó sus Estándares de seguridad del usuario, un conjunto de recomendaciones que incluyen la contratación de un responsable de fiabilidad y seguridad, la moderación de contenido y la integración de las últimas investigaciones en la lucha contra la toxicidad. Las empresas que se unen al consorcio se comprometen a trabajar para lograr estos objetivos.

Wang, quien pasó los últimos 15 años trabajando en inteligencia artificial (IA) y moderación de contenido, explica: «Quiero darle a la web y al metaverso una nueva opción. Si el metaverso va a sobrevivir, debe tener en cuenta e integrar la seguridad».

Tiene razón: el éxito de la tecnología está ligado a su capacidad de garantizar que los usuarios no resulten perjudicados. Pero ¿de verdad podemos confiar en que las empresas de Silicon Valley lograrán autorregularse en el metaverso?

De momento, entre las empresas que se han unido a Oasis figuran la plataforma de juegos Roblox, la empresa de citas Grindr y el gigante de los videojuegos Riot Games. Juntos cuentan con cientos de millones de usuarios, muchos de los cuales ya utilizan espacios virtuales activamente.

Sin embargo, cabe destacar que Wang aún no ha hablado con Meta, posiblemente el jugador más importante en el futuro metaverso. Su estrategia es acercarse a las Big Tech «cuando vean los cambios significativos que estamos consiguiendo con nuestro movimiento». Cuando se le preguntó sobre sus planes de seguridad en el metaverso, Meta señaló dos documentos: un comunicado de prensa que detalla las asociaciones con grupos e individuos para «construir el metaverso de manera responsable» y una publicación de blog sobre cómo mantener la seguridad en los espacios de realidad virtual. Ambos fueron escritos por el CTO de Meta, Andrew Bosworth.

Wang espera garantizar la transparencia de varias maneras. Una es creando un sistema de calificación para que la sociedad sepa cuál es la posición de cada empresa en el mantenimiento de la fiabilidad y la seguridad, similar al que siguen muchos restaurantes, donde se muestran las calificaciones de la ciudad por cumplir estándares de sanidad y limpieza. Otra consiste en exigir a las empresas miembros que contraten

desarrollar un internet ético donde las generaciones futuras sientan confianza para poder interactuar, crear y existir sin el odio y la toxicidad *online*.

a un responsable de la fiabilidad y seguridad. Este puesto se ha vuelto cada vez más común en las empresas más grandes, pero no existe un conjunto de estándares acordados que deba cumplir cada responsable de la fiabilidad y seguridad, señala Wang.

Pero gran parte del plan de Oasis sigue siendo, en el mejor de los casos, idealista. Un ejemplo es la propuesta para utilizar aprendizaje automático para detectar acoso y discursos de odio. Como informamos el año pasado, los modelos de IA o dan demasiadas posibilidades de propagación al discurso de odio o se exceden. Aun así, Wang defiende la promoción de la IA por parte de Oasis como una herramienta de moderación. «La IA es tan buena como los datos. Las plataformas comparten diferentes prácticas de moderación, pero todas trabajan para lograr una mejor precisión, una reacción más rápida y la seguridad mediante la prevención en el diseño», explica.

la IA es tan buena como los datos. Las plataformas comparten diferentes prácticas de moderación, pero todas trabajan para lograr una mejor precisión, una reacción más rápida y la seguridad mediante la prevención en el diseño.

El documento de Oasis, de siete páginas, describe los futuros objetivos del consorcio. Gran parte parece una declaración de misión, y Wang explica que, en los primeros meses, el trabajo se ha centrado en la creación de grupos asesores para ayudar a definir los objetivos.

Otros elementos del plan, como su estrategia de moderación de contenido, resultan demasiado vagos. Wang afirma que le gustaría que las empresas contrataran un grupo diverso de moderadores de contenido para que puedan comprender y combatir el acoso de las personas

«La ética se suele relegar a un segundo plano, pero [Oasis] fomenta pensar en ella desde el principio».

racializadas y de todas aquellas que se identifican como no hombres. Pero su plan no ofrece más medidas para lograr este objetivo.

El consorcio también espera que las empresas miembros compartan datos sobre los usuarios que cometan abusos, lo cual es importante para identificar a los infractores reincidentes. Las empresas participantes se asociarán con otras organizaciones sin ánimo de lucro, agencias gubernamentales y fuerzas del orden para crear políticas de seguridad, indica Wang que también planea que Oasis tenga un equipo de respuesta policial, cuyo trabajo será notificar a la policía sobre el acoso y abuso. Pero no está claro cómo la colaboración del grupo de trabajo con las fuerzas del orden diferirá del *statu quo.*

EQUILIBRAR PRIVACIDAD Y SEGURIDAD

A pesar de la falta de detalles concretos, algunos expertos creen que el documento de estándares del consorcio es, al menos, un buen primer paso. «Es bueno que Oasis se dedique a la autorregulación, empezando por las personas que conocen los sistemas y sus limitaciones», opina la abogada especializada en tecnología y derechos humanos Brittan Heller.

No es la primera vez que las empresas tecnológicas colaboran de esta manera. En 2017, algunas acordaron intercambiar información libremente con el Foro Global de Internet para Combatir el Terrorismo (GIFCT, por sus siglas en inglés). Hoy en día, GIFCT sigue siendo independiente y las empresas que se adhieren a él se autorregulan.

La investigadora de la Escuela de Computación y Sistemas de Información de la Universidad de Melbourne (Australia) Lucy Sparrow cree que lo bueno de Oasis es que ofrece a las empresas algo con lo que trabajar, en vez de esperar a que ellas mismas inventen los términos o esperar a que un tercero haga ese trabajo.

Sparrow añade que la integración de la ética en el diseño desde el principio, como propone Oasis, es admirable y que su investigación de sistemas de juegos multijugador muestra que eso marca la diferencia. «La ética se suele relegar a un segundo plano, pero [Oasis] fomenta pensar en ella desde el principio», destaca.

Pero Heller cree que el diseño ético podría no ser suficiente y sugiere que las empresas de tecnología modifiquen sus términos de servicio, que han sido fuertemente criticados por aprovecharse de los consumidores sin conocimientos jurídicos.

Sparrow está de acuerdo y duda de que un grupo de empresas de tecnología actúe en el mejor interés de los usuarios. La experta señala: «Surgen dos preguntas. Una, ¿cuánto confiamos en las corporaciones impulsadas por el capital para controlar la seguridad? Y dos, ¿cuánto control queremos que tengan las empresas tecnológicas sobre nuestras vidas virtuales?»

Es una situación complicada, especialmente porque los usuarios tienen derecho tanto a la seguridad como a la privacidad, pero esas dos necesidades pueden entrar en conflicto.

Por ejemplo, los estándares de Oasis incluyen consejos para presentar quejas ante la policía si algún usuario sufre acoso. Pero, presentar una denuncia suele ser difícil porque, por razones de privacidad, las plataformas a menudo no registran lo que sucede en ellas.

Este cambio marcaría una gran diferencia en la capacidad de disciplinar a los reincidentes, que actualmente pueden salirse con la suya al practicar abusos y acosos en distintas plataformas, porque estas no se comunican entre sí sobre qué usuarios son problemáticos. No obstante, Heller cree que, aunque en teoría es una gran idea, resulta difícil ponerla en práctica, porque las empresas están obligadas a mantener la privacidad de la información del usuario de acuerdo con las condiciones del servicio.

Y se pregunta: «¿Cómo se podrían anonimizar estos datos para poder compartirlos de manera efectiva? ¿Cuál sería el umbral para compartir los datos? ¿Cómo hacer que el proceso de compartir información sea transparente y que las eliminaciones de usuarios sean recurribles? ¿Quién tendría la autoridad para tomar tales decisiones?»

«No hay precedentes de empresas compartiendo información [con otras empresas] sobre los usuarios que violan los términos de servicio por acoso u otro mal comportamiento, aunque sobrepase los límites de la plataforma», agrega la abogada.

Una mejor moderación del contenido, por parte de personas y no máquinas, podría cortar el acoso en la raíz. Sin embargo, Heller no tiene claro cómo Oasis planea estandarizar la moderación de contenido, especialmente entre un medio basado en texto y otro que es más virtual. Moderar en el metaverso tendrá su propio conjunto de desafíos.

Y detalla: «La moderación de contenido en las redes sociales con IA que capta el discurso de odio se basa principalmente en el texto. La moderación de contenido en RV deberá rastrear y monitorizar el comportamiento principalmente, en teoría y los actuales mecanismos de informes de RV y RA [realidad virtual y aumentada] son, en el mejor de los casos, cuestionables y, a menudo, ineficaces. No puede ser automatizada por IA en estos momentos».

«Quiero darle a la web y al metaverso una nueva opción. Si el metaverso va a sobrevivir, debe tener en cuenta e integrar la seguridad».

Eso pone la carga de denunciar el abuso en el usuario, como experimentó la víctima de acoso de Meta. El audio y el vídeo tampoco se suelen grabar, lo que dificulta encontrar pruebas. Incluso entre las plataformas que graban el audio, Heller destaca que la mayoría retiene solo fragmentos, lo que hace que el contexto sea difícil, si no imposible, de entender.

Wang enfatiza que los Estándares de seguridad del usuario han sido creados por un consejo asesor sobre la seguridad, pero todos eran miembros del consorcio, algo que inquietó a Heller y Sparrow. La verdad es que las empresas nunca han tenido un gran historial de protección de la salud y la seguridad del consumidor desde que existe internet; ¿por qué deberíamos esperar algo diferente ahora?

Sparrow no cree en eso: «Lo importante es tener un sistema establecido para poder hacer justicia o señalar qué tipo de comportamientos se esperan y que hay consecuencias para esos comportamientos que están fuera de lugar». Eso podría significar involucrar a otras partes interesadas y ciudadanos, o algún tipo de gobernanza participativa que permita a los usuarios testificar y actuar como jurado.

Pero no cabe duda de que la seguridad en el metaverso requerirá algo más que un grupo de compañías tecnológicas que prometen cuidarnos. **</>**

Reportera senior que cubre la intersección humanos/tecnología en *MIT Technology Review*. Fue editora científica en *The Daily Beast e Inverse*.

El artículo original «El reto casi imposible de ofrecer seguridad y privacidad en el metaverso» pertenece a la edición digital de *MIT Technology Review.*