

¡Es broma! Deepfakes, sátira y las políticas de los medios artificiales



El reporte completo en inglés está disponible en wit.to/JustJoking

RESUMEN EJECUTIVO

Entre artistas, creadores satíricos y activistas de todo el mundo, es cada vez más latente el uso de los deepfakes. Pero, ¿qué sucede si no se toma en consideración el impacto que pueden tener sobre personas en condición de vulnerabilidad? ¿Qué sucede cuando intenciones malignas se disfrazan de humor? Este informe presenta y analiza la creciente convergencia entre deepfakes y sátira. ¿Quién decide qué es divertido, qué es justo y quién es responsable?

El informe analiza más de setenta casos recientes de una amplia gama de deepfakes. Algunos son ejemplos de potente sátira, arte o activismo, desde burlarse de los líderes autoritarios hasta resucitar sintéticamente a víctimas de la injusticia para exigir acciones. Pero otros demuestran cómo los malos actores utilizan la comedia como espada y escudo, para glorificar a poderosos y atacar a las comunidades marginadas, mientras buscan escapar de la culpa. Cada vez más, la sátira se utiliza como excusa defensiva: "¡Es broma!" - después de que un video haya circulado y causado daños.

¿Cómo podemos distinguir la desinformación intencional o el *gaslighting* de la crítica? ¿Qué imágenes o voces son fáciles de imitar sin consentimiento? El informe investiga los desafíos éticos de democratizar la producción de medios artificiales, incluido cómo etiquetar estos videos a medida que circulan en línea y quién es responsable de crear y hacer cumplir las mejores prácticas, protocolos y leyes. Considera las formas en que las plataformas multimedia intentan o no hacer estas distinciones en sus políticas y protocolos, así como los principios y herramientas relevantes en el derecho internacional de los derechos humanos y las leyes de los Estados Unidos.

El estudio concluye con una serie de preguntas de orientación adicional para la investigación y la acción en torno a la producción, circulación y recepción de medios artificiales. Sostiene que una amplia gama de voces deben ser parte de esta discusión, incluidos no solo corporaciones, personas tecnológicas, abogadas y políticas, sino también activistas de derechos humanos, artistas, periodistas y comunidades de todo el mundo.

El reporte *¡Es broma!* es parte de una colaboración continua entre el Co-Creation Studio en MIT Open Documentary Lab y la red de derechos humanos, video y tecnología WITNESS. La iniciativa 'Prepare. Don't Panic' de WITNESS está basada en un enfoque de liderazgo de derechos humanos globalmente inclusivo para los deepfakes, la autenticidad y la manipulación de los medios.

Preguntas clave, próximos pasos

Es necesaria una reflexión más sostenida y con matices sobre el uso y el mal uso de los medios artificiales y su lugar en un ecosistema de medios más amplio. Los siguientes grupos de preguntas ponen en primer plano las posibilidades cívicas y el daño potencial de los medios artificiales, apuntando hacia futuras discusiones y agendas de investigación.

En los próximos meses, nuestro objetivo es convocar a personas de todas las disciplinas, sectores, campos de especialización y ámbitos de la vida para ayudar a explorar algunas respuestas. ¡Únete a la conversación!

El consentimiento y las personas target de deepfakes satíricos

1. ¿En qué casos, si es que hay alguno, se necesita el consentimiento para hacer un deepfake de una persona en una posición de poder?
2. ¿Debería la naturaleza visceral y fotorrealista de los deepfakes requerir el consentimiento de una manera que no exigen otras formas de parodia o suplantación?
3. ¿Puede alguien dar su consentimiento de manera significativa para que una tercera persona utilice su imagen de la forma que elija?
4. ¿En qué circunstancias es aceptable hacer un deepfake de alguien que ya falleció?

5. ¿Bajo qué circunstancias debería ser completamente inaceptable hacer un deepfake de otra persona?
6. ¿Existen pasos prácticos para distinguir mejor entre contenido satírico y manipulación maliciosa disfrazada de humor?

Revelación, etiquetado e intención

1. ¿Debería exigirse a las personas creadoras que etiqueten todos los deepfakes? ¿Cómo se podría implementar y hacerlo cumplir? ¿Podría el etiquetado explícito de la sátira deepfake y de otras formas de arte habilitado por inteligencia artificial socavar su resonancia?
2. ¿Cómo pueden las personas creadoras de medios artificiales dar una idea a las consumidoras sobre el proceso de creación sin necesariamente neutralizar el poder retórico o estético del objeto?
3. ¿Cómo podría compartirse la responsabilidad de etiquetar y divulgar este tipo de medios entre las personas desarrolladoras de herramientas, creadoras, las plataformas en línea y otras partes interesadas?
4. ¿Los deepfakes de baja calidad que engañan solamente a algunas personas espectadoras deben ser tratados tanto por las plataformas como por legisladorxs como funcionalmente idénticos a los que tienen como objetivo explícito dañar o engañar?
5. ¿Cómo pueden las interfaces de las plataformas y la experiencia del usuarix hacer que sea más difícil o más fácil para las personas espectadoras comprender el origen, el estilo y la intención de los multimedia que encuentran?

Usos para atacar e impacto diferenciado

1. ¿Cuál es el impacto diferenciado de los deepfakes como herramienta de ataque en los países del Sur global, así como en las comunidades marginadas del Oeste/Norte global?
2. ¿Cómo se puede aplicar un marco de derechos humanos para la toma de decisiones prácticas en torno a medios manipulados y deepfakes a nivel de plataforma y gobierno?

Moderación de contenido y contramedidas

1. ¿Cuál es el enfoque de moderación de contenido apropiado para deepfakes y medios artificiales que funciona a nivel mundial y respeta marcos normativos claros como el derecho internacional de los derechos humanos?
2. ¿Cuáles son las formas en que las plataformas pueden garantizar que el arte habilitado por inteligencia artificial (incluidos los deepfakes) no sea desacreditado por la moderación automatizada o una errónea verificación de hechos?
3. ¿Cómo impactan los diferentes enfoques de las plataformas de redes sociales para moderar este tipo de medios en unas regiones globales específicas más que en otras, o difieren en el impacto según la comprensión culturalmente específica de la sátira?
4. ¿Debería exigirse legalmente a las apps de deepfakes que se rijan bajo un "código de práctica" común para minimizar el potencial de uso indebido? ¿Se podría exigir a las personas usuarias que firmen algo similar?

Ley y política pública

1. ¿Cuál es el papel del gobierno en cuanto a regular cómo las diferentes formas de medios circulan en línea? ¿Debería la legislación dictar protocolos para categorizar y examinar los medios habilitados por inteligencia artificial específicamente, y/o establecer estándares para la rendición de cuentas?
2. ¿Debería prohibirse el uso de deepfakes en campañas y anuncios políticos en determinados contextos o en todos los contextos?
3. ¿Cuáles son las formas en que las leyes de derechos de autor y difamación podrían utilizarse como pretexto para suprimir los usos creativos y socialmente cargados de los medios habilitados por la inteligencia artificial? ¿Existen barreras especiales o medidas legales anticipatorias que puedan tomarse para prevenir esto?
4. ¿Cómo pueden las sociedades democráticas evitar que el "pánico por deepfake" y el "dividendo del mentiroso" se usen en su contra, así como las políticas gubernamentales o comerciales que restringen demasiado la sátira, el periodismo veraz y la verificación de hechos?

Reporte completo en inglés disponible en wit.to/JustJoking