Deepfakes, la nueva amenaza tecnológica imposible de solucionar

Fuente:
TyN Magazine

Imagine la siguiente situación: por fin existe un detector de ultrafalsificaciones (*deepfakes*) totalmente eficaz. El sistema inmediatamente coloca una gran etiqueta roja con la palabra *DEEPFAKE* a cada vídeo que ha sido manipulado con inteligencia artificial (IA), por muy realista que parezca. Eso podría ser exactamente lo que necesitamos para luchar contra los *deepfakes*.

La gente está cada vez más preocupada por esta técnica que por su potencial para acabar con la verdad y la democracia. Los vídeos manipulados “perfectamente reales” podrían llegar en tan solo seis meses, lo que sugiere que, a partir de ese momento, los comicios electorales de todo el mundo podrían convertirse en un campo de batalla de vídeos falsos.

Los tecnólogos están intentando resolver el problema con más tecnología. El Gobierno de Estados Unidos ha financiado un proyecto sobre el “análisis forense de contenidos”. Facebook y Microsoft anunciaron recientemente un desafío de detección de *deepfakes*, y Google publicó una gigante base de datos con vídeos ultrafalsificados para combatirlos. Pero aunque la técnica de crear *deepfakes* es nueva, una gran parte del daño que causa (desinformación y acoso) no lo es, según la profesora de información en la Universidad de Rutgers (EE. UU.) y coautora de un reciente informe de Data & Society sobre *deepfakes*, Britt Paris. En su opinión, ningún detector de *deepfakes* totalmente fiel sería capaz de evitar esos daños. Estas son las razones:

Problema 1: el detector de *deepfakes* no puede decirnos si un vídeo debe ser eliminado

¿Se acuerda del vídeo a cámara lenta de Nancy Pelosi en el que parecía que la política estaba ebria? Eso no era un *deepfake*. Aunque el clip contenía falsedades, Facebook decidió no eliminarlo. Un detector de *deepfakes* no hubiera contribuido de ninguna forma a esa decisión. “Cuanto más se automatice la tecnología, más probabilidades habrá de que haya imprecisiones o censura. Definir que es sátira, qué son noticias falsas y qué es ficción grandes cuestiones filosóficas”, opina la profesora de la Universidad de St. John (EE. UU.) y experta en la gestión de plataformas Kate Klonick.

Solución: mejorar el sistema de moderación

La sociedad debería resolver estas cuestiones. Pero hasta que no se haga, una opción podría consistir en dar más poder a quienes tienen capacidad de decisión: los moderadores de contenido. Con ese objetivo, estos trabajadores podrían recibir un mayor salario, una mejor formación y estar mejor valorados como parte importante del mantenimiento de la seguridad de internet. Esto es lo que propone la profesora de información de la Universidad de California en Los Ángeles (EE. UU). Sarah T. Roberts. Bajo este enfoque,

los equipos especializados de moderadores experimentados podrían evaluar el contexto de los vídeos, verificarlos y decidir si deberían permanecer en la plataforma. Es posible que no tengan la respuesta perfecta para ese vídeo de Pelosi, pero sí que tendrían una idea del impacto social y político de una variedad de *deepfakes* y sus objetivos. Podrían decir que los *deepfake* de parodias de Nicolas Cage están bien pero que las falsificaciones pornográficas no.

¿Quién más podría opinar sobre un *deepfake*? La víctima del mismo. Las empresas deberían facilitar la opción de denuncia de acoso mediante *deepfakes*, apunta la experta en ciberderecho de la Universidad de Boston (EE. UU.) Danielle Citron. Todos los usuarios deben estar informados sobre sus derechos, y los pasos que deben seguir para denunciar deben ser claros y accesibles, en lugar de estar escondidos en la política de privacidad.

Problema 2: es posible que la tecnología contra *deepfakes* no ayude a las personas que necesitan más protección

Ya se ha demostrado que, antes de convertirse en una amenaza generalizada, las nuevas tecnologías se usan más contra los grupos más vulnerables, como las mujeres, personas de color, la comunidad LGBTI y los activistas, según Paris. En la década de 1990, por ejemplo, ya había imágenes de Photoshop de cabezas de mujeres en cuerpos de actrices de cine para adultos. A las personas en el poder este problema no les afectaba lo suficiente como para hacer algo al respecto. La jurista en la Universidad de Miami (EE. UU.) Mary Anne Franks detalla: “Si hubiéramos prestado atención al problema de la explotación sexual de las mujeres sin su consentimiento, ahora estaríamos en una posición mucho mejor para tratarlo social, legal y culturalmente”.

La historia se repite. Los investigadores afirman que el mayor riesgo de los *deepfakes* no es que influyan en unas elecciones, sino que se usen para intimidar a ciudadanos individuales.

Solución: no decidir nada sin hablar con los más afectados

Hay que hablar con las personas más vulnerables a los *deepfakes*, según el director de programa de la organización sin ánimo de lucro que analiza los medios sintéticos Witness, Sam Gregory. Incluso si el objetivo es crear un detector infalible de *deepfakes*, hay muchas cuestiones sociales involucradas. ¿Estará disponible en otros países? ¿Será capaz de detectar los *deepfakes* políticos o la violencia de género y la sexual? “Cuando se establece la infraestructura, las personas y las poblaciones realmente marginadas quedan excluidas porque no tienen representación ni poder para cambiar esa infraestructura”, explica Gregory.

Problema 3: la detección de *deepfakes* no llega a tiempo de ayudar a las víctimas

Con las ultrafalsificaciones, “hay pocos recursos reales después de que se haya publicado ese vídeo o audio“,

asegura Franks. Las leyes existentes no son adecuadas. Las leyes que penalizan el intercambio de información privada como los informes médicos no se aplican a los vídeos falsos y perjudiciales. Las leyes contra la suplantación son “extrañamente limitadas”, afirma Franks. Estas normas se centran en el hecho de que es ilegal hacerse pasar por un médico o un funcionario del Gobierno. Las leyes sobre la difamación solo abordan las representaciones falsas que retratan negativamente a la víctima, pero Franks cree que deberíamos estar preocupados por los *deepfakes* que representan falsamente a alguien, aunque sea de manera positiva.

Solución: nuevas leyes

Texas (EE. UU.) aprobó recientemente un proyecto de ley para prohibir los *deepfakes*. El proyecto de ley sobre los *deepfakes* de California (EE. UU.) fue aprobado en ambas cámaras. Además, la congresista estadounidense de Nueva York (EE. UU.) Yvette Clarke presentó recientemente un proyecto de ley federal llamado Ley de Responsabilidad de DEEPFAKES. Esta norma obligaría a las empresas de redes sociales a crear mejores herramientas de detección en sus plataformas y permitiría penalizar o encarcelar a quienes publiquen *deepfakes* maliciosos.

Franks y Citron están trabajando en un proyecto de ley federal para penalizar los *deepfakes* maliciosos, a los que denominan “falsificaciones digitales”. En su opinión, una falsificación digital o ultrafalsificación es algo que una persona razonable pensaría que es real. También es probable que cause daño a una persona en particular o al orden público, por ejemplo, si muestra falsamente a una persona musulmana cometiendo un delito.

Si publicar vídeos dañinos se convirtiera en un delito, no solo disuadiría a las personas de publicarlos sino que las plataformas como Facebook tendrían que esforzarse más para mantenerlos fuera.

Disponible en:

https://www.tynmagazine.com/deepfakes-la-nueva-amenaza-tecnologica-impos... [1]

Links
[1] https://www.tynmagazine.com/deepfakes-la-nueva-amenaza-tecnologica-imposible-de-solucionar/