Daniel Andrés Birchner 18/1/2025

POSVERDAD Y POLÍTICA:
EL CRECIENTE IMPACTO DE LOS DEEPFAKES

Si bien el término *deepfake* no goza actualmente de una definición consensuada, con el objetivo de arribar a una noción transversal podemos acercarnos a ella mediante lo expresado por Nina Schick (2020), que lo define como: *"un tipo de media digital (audio, video e imágenes) totalmente o en parte generado por inteligencia artificial con una finalidad maliciosa o bien, des-informativa"*, [1]. De esta forma, un deepfake consiste en la creación o modificación (con fines maliciosos) de la apariencia física de una persona o incluso, su voz, y colocarla digitalmente en lugares en los que nunca estuvo o hacerle decir cosas que nunca dijo, simplemente haciendo que su imagen se mueva como se le ordene, mediando en todo este proceso, la IA.

Como fue mencionado anteriormente, no es posible hoy en día llegar a un consenso sobre la definición de *deepfake* en la literatura científica pertinente. Sin embargo, es posible encontrar ciertas ideas comunes a prácticamente todas las definiciones, las cuales se detallan a continuación:

- El término *deepfake* es una conjunción de los términos *deeplearning* (aprendizaje profundo)y *fakenews* [2] [3].
- El *deepfake* consiste en la alteración de videos, audio o imágenes mediante IA (a través de los procesos de *machine learning* -aprendizaje automático- y más precisamente, *deeplearning*) [4] [5].
- El *deepfake* se trata de un montaje intencional, es decir, en el medio digital alterado hay una intención humana manifiesta, que es realizada con asistencia de IA [6] [7].
- El *deepfake* puede tener una finalidad positiva (demostración de tecnología, bromas pesadas) como negativas (suplantación de identidad, difamación

[1] Schick, N. (2020) – *Deepfakes: the coming infoapocalypse*. Editorial Tamang Ventures, Nueva York, EUA. p. 9.
[2] *Ídem.*
[3] Botha, J., Pieterse, H. (2020) – *Fake news and deepfakes: a dangerous threat for 21st Century Information Security.* Consejo de Investigación Científica e Industrial (CSIR), Gobierno de Sudáfrica.
[4] Le Cunn Y. *et al*., (2015) - *Deep Learning* en Revista Nature, Vol. 521.
[5] Twang, T. (2020) – *Deepfakes, a grounded Threat Assessment*. Centro para la Seguridad y la Tecnología Emergente (CSET).
[6] Schick, N. *op. cit.*, p. 10.
[7] Twang, T. *op. cit.*

pública) [8] [9]. Esta última percepción negatia de la intencionalidad es la predominante.

Es importante, sin embargo, contener las dimensiones del concepto. *Deepfake* es, como se dijo anteriormente, un montaje intencional, asistido con IA, con posibilidades fines des-informativos o malintencionados por parte del creador (descartando aquellas finalidades ‘’positivas’’, como el mero entretenimiento o la demostración de tecnología) [10]. Esta conceptualización deja de lado los montajes que no requieran IA, o bien, aquellos que se hacen en películas para recrear rostros de actores que ya murieron (caso de la Princesa Leia en la película “*Star Wars: Rogue One*”). Quedan de lado, así mismo, aquellos montajes que resultan de una modificación con programas de edición de videos que no usen IA. Además, gracias a esta mencionada asistencia total/parcial de la IA, los *deepfake* no requieren un gran volumen de intervención humana [11], ya que la IA se encarga del trabajo más denso de forma automática permitiendo al ser humano plantear la intención del *deepfake*.

Existe también una sub-categoría de clasificación respecto al nivel de sofisticación del deepfake. El *deepfake* propiamente dicho, es una manipulación profunda de video y audio que emplea herramientas avanzadas de IA para sustituir rostros o discursos con alto realismo. Esta técnica se basa en algoritmos de *deep learning*, lo que dificulta su detección. El *cheapfake*, en cambio, se logra con ediciones básicas como alterar la velocidad de reproducción o los subtítulos, sin necesaria intervención de IA. Por su simplicidad, suele ser más fácil de identificar que el *deepfake*.

En cuanto a su finalidad, de acuerdo a las estadísticas de la tecnológica holandesa Deeptrace[12], un 96 % del *deepfake* (de video) actual se usa con fines de pornografía y en mucha menor medida, con fines cómicos (bromas) o de manipulación política. Adicionalmente, aunque aún se trata de un uso marginal, el Programa Nacional para la

[8] Mahmud, B. U. y Sharmin, A. (2020) – *Deep insights of Deepfake Technology* en Revista DUJASE, Vol. 5.
[9] Graber–Mirchell, N. (2021) - *Artificial Illusions: Deepfakes as Speech* en Revista Intersect, Vol. 14, N° 3.
[10] Botha, J., Pieterse, H. (2020) – *Fake news and deepfakes: a dangerous threat for 21st Century Information Security*. Consejo de Investigación Científica e Industrial (CSIR), Gobierno de Sudáfrica.
[11] Westerlund, M. (2019) – *The Emergence of Deepfake Technology: A Review* en Technology Innovation Managment Review, Vol. 9.
[12] Cafranc, P. (2019) - *Deepfake, cuando lo que vemos ya no es fiar*, pág 6.

Inteligencia Artificial de Estados Unidos, en su informe del año 2021, ya advertía el uso potencialmente peligroso (intencional) con fines políticos de *deepfakes*[13].

En 2020, el Centro por la Seguridad y las Tecnologías Emergentes, un *thinktank* ubicado en la Universidad en Georgetown, USA, emitió su informe *Deepfakes: a Grounded Threat Assessment* (Una evaluación fundamentada de amenazas). En el escrito, la entidad señala la creciente preocupación de la comunidad de seguridad nacional estadounidense en torno a los usos potencialmente peligrosos de los *deepfake*, sobre todo aquellos relacionados a la manipulación de video con figuras políticas y sus consecuencias tanto nacionales como internacionales. Destaca que la masificación de los *deepfakes* corresponden a una era donde "*we can no longer believe what we see*" (ya no podemos creer en lo que vemos), cuando gran parte de la difusión informativa se realiza por medio de plataformas sociales digitales, susceptibles por una lado de ser fácilmente manipuladas y por otro, pensadas para polarizar/compartimentar la opinión pública.

Es necesario tener en cuenta que los *deepfake* se difunden y contemplan en un contexto y universo definido, el de las *fakenews*. De acuerdo a la Federación Internacional de Periodistas[14], el término *fakenews* se usa para conceptualizar la divulgación de noticias falsas que provocan un peligroso círculo de desinformación intencional[15]. Pueden tomar varios formatos, como ser prensa escrita, de audio o videos, en internet o en forma de rumor.

Ahora bien, a guisa de exactitud conceptual, existe cierta propensión a usar como sinónimos de *fakenews* algunos términos: noticias falsas y propaganda. Las *fakenews* no son necesariamente noticias falsas, ya que estas últimas no son inicialmente concebidas con el propósito de desinformar, sino que pueden ser falsas en base a omisiones o a información interpretada erróneamente. Las noticias falsas pueden ser revisadas y corregidas en consecuencia, mientras que una *fakenews* desde su concepción parte con el fin explícito de desinformar, a la vez que son generalmente incontrastables, sin posibilidad de verificar su veracidad. Una noticia falsa puede aparecer en un medio

[13] National Program for Artificial Inteligence (2021) - *Deepfake Guide.*
[14] Federación Internacional de Periodistas - *¿Qué son las fakenews?*
[15] Ídem.

informativo reconocido, mientras que las *fakenews* suelen distribuirse en fuentes dudosas o tendenciosas.

Los deepfakes, cuando tienen una intencionalidad política, pueden ser considerados dentro de la órbita de las *fakenews*, es decir noticias intencionalmente falsas. Es importante resaltar que las noticias falsas o contenciosas no son nuevas, pero lo innovador de las *fakenews* radica en su medio de difusión principal (las redes sociales) y en el contexto social-informativo conocido como posverdad.

El concepto de *posverdad*[16] se refiere, de acuerdo al Diccionario de Oxford (1992), al marco simbólico en el que los hechos objetivos son menos importantes a la hora de modelar la opinión pública, priorizando las apelaciones a la emoción o a las creencias populares[17]. Es decir, una perspectiva modelada no por criterios racionales y objetivos de análisis de la realidad, sino interpretaciones basadas puramente en puntos de vista emocionales y volátiles, factibles de ser manipulados por noticias o información que apelan al sensacionalismo y no al pensamiento crítico. Esta definición también se verifica en la versión de la Real Academia Española, que define a la posverdad como "*distorsión deliberada de una realidad, que manipula creencias y emociones con el fin de influir en la opinión pública y actitudes sociales*"[18]. Esta definición será importante a la hora de entender el efecto sobre la opinión pública de los *deepfakes*, conjuntamente con los errores de procesamiento que se estudian desde la Psicología Cognitiva. Es un concepto decisivo a la hora de entender por qué el *deepfake* supone una amenaza para la Seguridad Nacional e Internacional, paralelamente al contexto tecnológico que lo sustenta.

EFECTOS POLÍTICOS (CASOS)

El uso político de los deepfakes es menos frecuente en comparación con sus aplicaciones en la pornografía o la parodia, pero cuando se emplean con fines de manipulación política pueden tener consecuencias profundas en la continuidad de un gobierno, en procesos electorales o incluso en el curso de un conflicto bélico. En esos casos, la clave radica en la manipulación de la opinión pública a través de imágenes o

[16] *Post-truth* es el término en inglés.
[17] Disponible en: https://www.oxfordlearnersdictionaries.com/spellcheck/english/?q=postruth
[18] Disponible en: https://dle.rae.es/posverdad?m=form

videos falsificados para sembrar dudas sobre la legitimidad de un gobierno, influir en el respaldo a políticas de Estado, inducir el voto a favor o en contra de un candidato, difamar figuras políticas o incluso dirigir percepciones económicas con fines de lucro.

En los ejemplos que se presentarán, se muestran casos donde los deepfakes produjeron o estuvieron a punto de producir cambios políticos relevantes. Los gobiernos son cada vez más conscientes del poder de estas técnicas, contando con expertos que buscan detectarlas y desmentirlas antes de que alcancen gran repercusión. Tanto los *deepfakes* como los *cheapfakes* pueden manipular eficazmente la opinión pública, a veces sin requerir grandes recursos técnicos. Sin embargo, cuando se apunta a un grupo más informado o reducido, se requiere un mayor realismo para lograr el efecto deseado. A continuación, se ordenarán los casos de menor a mayor impacto político.

En 2018, Estados Unidos presenció un giro sorprendente en la percepción del poder de los *deepfakes* cuando el actor y cineasta Jordan Peele difundió un video falso del expresidente Barack Obama. Bajo el título "*You won't Believe what Obama says in this video!*", la grabación mostraba al exmandatario realizando declaraciones ofensivas hacia Donald Trump, con un realismo inquietante. Dado el prestigio de Obama y la popularidad de Peele, el video atrajo atención mediática masiva y sirvió como una alerta temprana sobre cómo la manipulación digital de alta precisión podía usarse para difundir información falsa. Peele, en un gesto de transparencia, reveló rápidamente su autoría, explicando que su objetivo era concienciar a la sociedad acerca de la facilidad con que la tecnología *deepfake* puede engañar a la opinión pública si se emplea irresponsablemente.

Un año después, en 2019, la presidente de la Cámara de Representantes de los Estados Unidos, Nancy Pelosi, se convirtió en protagonista involuntaria de otro episodio de manipulación mediante un *cheapfake* (sin IA). En el video que circuló con rapidez a través de redes sociales, se la mostraba hablando con lentitud y aparentando estar ebria. Si bien la técnica fue mucho más simple que en el caso de Obama —bastó con ralentizar el metraje—, el impacto fue similar en términos de alcance, provocando que diversos medios y figuras políticas cuestionaran su estado y capacidad. Pelosi se vio obligada a realizar declaraciones oficiales para desmentirlo, y el incidente evidenció que incluso

las manipulaciones básicas, sin algoritmos de IA, pueden erosionar la reputación de líderes clave y alimentar narrativas que perjudiquen su imagen pública.

Ese mismo año, en Gabón, la ausencia prolongada del presidente Alí Bongo desató rumores sobre su salud y capacidad para ejercer el poder, acrecentados por un video en el que aparecía con dicción extraña y gestos poco naturales. Muchos ciudadanos y opositores llegaron a sospechar que se trataba de un *deepfake* destinado a ocultar la muerte o incapacidad del mandatario, lo que generó tal nivel de desconfianza que terminó alimentando un intento de golpe de Estado. Aunque posteriormente se demostró que el video era real y que la apariencia de Bongo se debía a los efectos de varias cirugías y tratamientos médicos (razones por las cuales e había ausentado), el caso puso de manifiesto la fragilidad de un gobierno frente a la mera sospecha de manipulación audiovisual. El temor colectivo y la falta de información oficial consistente hicieron que un rumor de *deepfake* casi precipitara la caída de un gobierno, subrayando la capacidad de esta tecnología para desestabilizar a países enteros.

En 2021, Letonia vivió un suceso que ilustró otro riesgo de los *deepfakes*: la suplantación de identidad en reuniones políticas o diplomáticas de alto nivel. Varios eurodiputados, entre ellos Richard Kols, creyeron mantener una videollamada con Leonid Volkov, opositor ruso y jefe de campaña del difunto Alexei Navalny. Pero en realidad conversaban con un impostor, que empleaba un *deepfake* en tiempo real. El engaño, atribuido a dos comediantes rusos conocidos como *Vovan & Lexus*, dejó al descubierto la vulnerabilidad de las instituciones internacionales ante estrategias de manipulación digital cada vez más sofisticadas, pudiendo llegar a alterar discusiones geopolíticas sensibles si no se detectan a tiempo.

Por último, en 2022, el conflicto armado en Ucrania sirvió como escenario para la primera gran muestra de *deepfakes* utilizados con fines psicológicos y militares. Surgieron videos en los que el presidente ucraniano, Volodymyr Zelensky, supuestamente pedía la rendición de sus tropas ante autoridades rusas, lo cual fue desmentido de inmediato por las autoridades ucranianas. Poco después, un supuesto alcalde de Kiev convocó a líderes europeos en otra grabación falsa. Ambas maniobras se atribuyeron a Rusia y reflejaron la creciente explotación de los *deepfakes* para desmoralizar a la población y socavar la legitimidad de los líderes en escenarios de

conflicto. La rapidez con la que se desmintieron demostró, no obstante, un avance en la concientización y la capacidad de respuesta para identificar y neutralizar estos intentos de manipulación, marcando así un nuevo capítulo en el uso de la información (y la desinformación) como arma.

Los casos mencionados son los más relevantes, pero no son los únicos. Desde 2019, los deepfakes han aumentado significativamente en el contexto de campañas electorales alrededor del mundo, destacando casos en Corea del Sur (2020) y Estados Unidos (2024), en este último, supuestamente las manipulaciones llegaron desde China, Rusia e Irán. El caso norteamericano requiere una atención especial, dado que en 2024, el uso de *deepfakes* durante las elecciones se ha casi normalizado en la sociedad norteamericana, alertando a expertos como Donie O'Sullivan. En una entrevista de CNN[19], afirma que los *deepfakes*, dada su facilidad de generación y su impacto, comienzan a transformarse en una herramienta de desinformación política muy poderosa, al alcance de cualquier individuo. También resaltó que para generar los *deepfakes*, ya sean audios o videos, los autores utilizan el mismo medio que los candidatos utilizan para promocionarse: su aparición en los medios. Las fotografías, combinadas con los masivos discursos, proporcionan todos los datos necesarios para el entrenamiento de los algoritmos, que resulten en un *deepfake*. Eso sumado a la difusión de propaganda política en las redes sociales, constituye un terreno fértil para la propagación de estas producciones artificiales.

Hablar de *deepfakes* ya significa hablar de una realidad que, de acuerdo a las tendencias actuales, no hará mas que crecer. Será fundamental para su mitigación no sólo el mejoramiento de los sistemas de detección, sino también, la educación de la población en cuanto a los posibles efectos y a las vulnerabilidades psicológicas que los *deepfakes* buscan explotar.

[19] Disponible en: https://edition.cnn.com/2024/01/24/politics/deepfake-politician-biden-what-matters/index.html