Repercusiones éticas sobre el uso indebido del *deepfake* en el ámbito de las TIC mediante un análisis cualitativo documental

Ethical impact on the misuse of deepfake in the field of ICT through qualitative documentary analysis

Maldonado Arcila, Maribel Guisella, Ovalles Pabón, Freddy Oswaldo

Resumen

La presente investigación analiza las repercusiones éticas sobre el uso indebido del *deepfake* en el ámbito de las TIC. A través de un análisis cualitativo de revisión documental de información relacionada con problemáticas de índole ética, en la que se han visto perjudicadas figuras públicas y personas en general. Esto ha generado una preocupación tanto en presente como a futuro, puesto que, gracias a la inteligencia artificial con la que se emplea, cada vez será más fácil realizar vídeos realistas que emulan la realidad y permiten perjudicar y dañar la buena imagen de cualquier persona. Lo cierto es que es necesario recolectar y analizar las fuentes de información que traten el tema del *deepfake* en cuanto a sus principales problemáticas y usos indebidos; puesto que se requiere de la implementación de más normas y leyes que prohíban y regulen tanto el uso como la creación de software, para frenar el daño en el ámbito político, social, legal y religioso. Así, gracias al análisis de las consecuencias del *deepfake*, se pueden brindar estrategias que pueden servir de guía para crear mejores softwares de detección, y pautas éticas en la comunidad de las TIC.

Palabras clave:

Deepfake, inteligencia artificial, repercusión ética, aprendizaje profundo.

Abstract

This research analyzes the ethical repercussions on the improper use of *deepfake* in the field of ICT through an exhaustive search for information and a qualitative analysis of documentary review of information related to ethical problems that have harmed public figures and victims in general. ,thus evidencing a concern both in the present and in the future since, thanks to the artificial intelligence with which it is used, it will be

Fecha de recibido: 30/09/2021 Fecha de aceptación 20/12/2021
Correo electrónico: mgmaldonado00@misena.edu.co, fovalles@misena.edu.co
Investigación concluida, Proyecto: Repercusiones éticas sobre el uso indebido del Deepfake en el ámbito de las TIC. Abril 2021 a septiembre 2021. Semillero TESLA, Grupo de Investigación GINDET del Centro CIES SENA Reg, Norte de Santander. Maldonado Arcila, Maribel Guisella, Ovalles Pabón, Freddy Oswaldo.
Citar como: Maldonado, M.G y Ovalles, F.O (2021). Repercusiones éticas sobre el uso indebido del *deepfake* en el ámbito de las TIC mediante un análisis cualitativo documental. *Revista RETO, 9.*

increasingly easier to make realistic videos that emulate reality and allow damaging and damaging the good image of any person, therefore, it is necessary to collect and analyze those sources of information that deal with the subject of *deepfake* in terms of its main problems and improper uses, whose main authors defend the idea that said subject needs more control and later more rules and laws that prohibit and regulate both its use and the creation of software, which would stop damaging in the political, social, legal and religious sphere, thus, thanks to the analysis of the consequences of *deepfake*, strategies can be provided that can serve as a guide to create better screening software and ethical guidelines in the ICT community.

keywords:

Deepfake; Artificial intelligence; Ethical impact; Autonomous Learning.

Introducción

La presente investigación analiza las complejidades que se presentan hoy en día con el uso inadecuado de algunos elementos presentes en las TIC, específicamente, el uso indebido del *deepfake*. Gracias a la inteligencia artificial con la que se realizan, se da el espacio para la reproducción de videos falsos con el fin de entretener, engañar, difamar o chantajear. En tanto que son difíciles de detectar como información *fake* o falsa, este fenómeno audiovisual expone problemáticas éticas que apenas en algunos países se están regulando.

Uno de los principales problemas de un software que esté diseñado para suplantar imágenes, formas y hasta voces, es la gran repercusión ética que se tiene sobre la suplantación de identidad de una persona. Con todo lo cual se puede llegar a afectar la reputación social, posición política y hasta religiosa, además, de afectar, gravemente, la vida personal, económica y social, esto sin contar con que la persona suplantada puede ejercer un rol en el campo político de un país o ser una gran influencia social, con lo que se afectarían ambientes más allá del campo personal. Una de las finalidades del software es distorsionar la imagen pública de figuras reconocidas. Con la ayuda de las múltiples imágenes de referencia que se pueden encontrar, se pueden crear discursos falsos, videos falsos, difamaciones, entre otros. Gracias a los algoritmos RGA (Redes Generativas Antagónicas) se trata de minimizar al máximo cualquier falla en cada cuadro de imagen, para que el espectador al ver el video final lo crea y por ende, lo difunda.

Marco teórico

Actualmente, se tienen pocos estudios sobre el *deepfake* y fenómenos tecnológicos relacionados, debido a su reciente nacimiento. Sus principales exponentes tratan este tema como una problemática que requiere más control, tanto de índole ética como legal. La mayoría de estudios se encuentran en inglés, y tratan temas de repercusión social y política, que afectan gravemente a los involucrados si no se logran controlar. Para empezar, el doctor Jacob Bañuelos, en su artículo *Deepfake: la imagen en tiempos de la posverdad*, afirma "[...] los *deepfakes* minan la credibilidad de los documentos audiovisuales, principalmente videos, como medios de información o certificación de hechos, poniendo en entredicho su veracidad o generando riesgos de desinformación, difamación o chantaje" (2020). Así las cosas, se puede ver que existe una preocupación frente al ámbito ético y legal, en el que la credibilidad de los videos está cada vez más en entredicho.

Los autores Thanh Thi Nguyen, Cuong M. Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, Saeid Nahavandi, en su artículo "*Deep Learning for Deepfakes Creation and Detection*", brindan bases para la comprensión de cómo detectar *deepfakes* y la importancia de esta actividad para saber qué de lo que consumimos es real y qué no, para evitar así, tensiones políticas o sociales. A través de su estudio se muestra que esta tecnología no sólo sirve para la suplantación de identidades, sino también, para la modificación de mapas o terrenos, con lo cual se pueden perjudicar gravemente misiones militares.

> "Los *deepfakes* [...] podrían causar angustia y efectos negativos a los destinatarios, aumentar la desinformación y el discurso de odio, e incluso podrían estimular la tensión política, inflamar al público, la violencia o la guerra. Esto es especialmente crítico hoy en día, ya que las tecnologías para crear *deepfakes* son cada vez más accesibles y las plataformas de redes so-

ciales pueden difundir esos contenidos falsos rápidamente". (2020)

Frente a los temas que abarcan la alfabetización moral o ética sobre el mal uso de los *deepfakes*, los autores, Víctor Cerdán, María Luisa García y Graciela Padilla, en su artículo *"Alfabetización moral digital para la detección de deepfakes y fakes audiovisuales"*, muestran la importancia de generar una cultura ética frente al fenómeno, así como la problemática de no tener una frente al uso indebido de dichos videos. Estos autores afirman que estos videos tan realistas pueden ser, como lo mencionan, armas de acoso o incitadores de conflicto social. En caso de no contar con un software adecuado para la detección de estos, es necesario lo que ellos denominan como la alfabetización moral ciudadana, basándose en el propio humano, este, "como máquina compleja e inteligente que es [...] puede divisar las pruebas de lo falso. La inteligencia humana, por ahora, se erige como más útil y rápida que la inteligencia artificial, mientras no haya software libre que ayude" (2020).

Los autores, Marissa Koopman, Andrea Macarulla Rodríguez, Zeno Geradts, en su artículo, Detection of *Deepfake* Video Manipulation, brindan una mirada sobre la gran consecuencia que tiene la manipulación de vídeos reales y suplantación de identidades, en el que algunas veces, los videos *fake* se usan en salas de audiencia y en investigaciones de tipo policial, lo cual perjudica a víctimas con evidencias falsas en su contra.

Las pruebas fotográficas y de video se utilizan comúnmente en la sala de espera y en las investigaciones policiales, se consideran tipos de pruebas fiables. Sin embargo, con los avances en las técnicas de edición de video, la evidencia de video se está volviendo potencialmente poco confiable. Es probable que, en un futuro próximo, se requiera que las pruebas de video se examinen en busca de rastros de manipulación antes de que se consideren admisibles ante los tribunales. (Koopman, Macarulla, Geradts, 2018)

Para entender cómo funciona el *deepfake*, es necesario entender cómo funciona un vídeo, este consiste en "la tecnología de captación, grabación, procesamiento, almacenamiento, transmisión y reconstrucción por medios electrónicos digitales o analógicos de una secuencia de imágenes que representan escenas en movimiento" (Sorrivas, 2015), por lo que en el *deepfake*, se puede realizar un montaje o reemplazo de un objeto en específico en cada una de las secuencias de ésas imágenes, para reemplazarlas por aquello que se quiere que el público observe como real, por ejemplo, como se menciona anteriormente, los videos *fake* en situaciones legales.

Según, Koopman, Macarulla, Geradts, Cerdán y Padilla (2018), el *Deepfake* permite reemplazar el rostro de cualquier persona en un video con la cara de otra, con las condiciones de que se disponga de suficientes imágenes de ambas, ya que, al existir diversos ángulos, se puede reemplazar ese rostro con mayor facilidad, entre más información fotográfica se tenga de una persona, más realista será el resultado del video.

Para una sola persona, este trabajo sería un proceso largo y tedioso, si se realizara manualmente, se tendría que hacer básicamente un montaje de imagen sobre imagen, editando cada cuadro del video. No obstante, actualmente, se cuenta con la tecnología de inteligencia artificial, con la cual se pueden realizar procesos mucho más ágiles y precisos. Así, en el ámbito del *deepfake*, se tienen en cuenta las denominadas GAN o RGA, que hacen posible su realismo y exactitud.

> "Las GAN consisten en dos modelos [...] el generador y el discriminador, entrenados simultáneamente para desafiarse uno al otro [...]. Por un lado, el generador es entrenado para generar datos falsos lo más parecidos posibles a los ejemplos reales [...] Por otro lado, el discriminador es entrenado para ser capaz de discernir los datos falsos producidos por el generador de aquellos que corresponden al conjunto de entrenamiento [...]. Sucesivamente, los dos modelos tratan continuamente de superarse" (Calcagni, 2020)

Gracias a esta tecnología es posible crear videos que sean totalmente creíbles ante el público que consuma el producto final, sea con fines humorísticos, comerciales, fraudulentos, entre otros. En muchas ocasiones, difíciles de discernir o categorizar como *fake*. Ocaña-Fernandez. y, Valenzuela-Fernández, L. y Garro-Aburto, L, comentan que la inteligencia artificial, "es parte de las ciencias de la computación que se ocupa del diseño de sistemas inteligentes, esto es sistemas que exhiben

características que asociamos con la inteligencia en las conductas humanas" (2019), así, se entiende que se emula el comportamiento o pensamiento humano.

Al contar con una tecnología que puede ser más precisa que un trabajo realizado de forma manual, se tiene como inconveniente el hecho de hacer creer a cualquier persona que algo es real, cuando se presente ante los medios de comunicación, es por ello que la ética juega un papel importante a la hora de realizar *deepfake*, ya que se presentan situaciones como la difamación, uso no autorizado de imágenes, derechos de autor, engaño al público y alteración de videos en asuntos legales, violando así, los derechos de las personas involucradas, tanto de quienes reciben información falsa, como del actor cuya imagen se ve alterada.

Metodología

Al analizar las principales repercusiones éticas del *deepfake*, se implementa un tipo de investigación exploratoria, ya que es un tema que está tomando un auge en la comunidad tecnológica de habla hispana, se tienen entre sus principales exponentes autores mayormente de habla inglesa, por lo que se recolectan y analizan aquellas fuentes de información que traten el tema del *deepfake* en cuanto a sus principales problemáticas y usos indebidos.

Así mismo, se tiene en cuenta información documental que evidencie el éxito de los mecanismos y software para combatir prácticas incorrectas e ilegales en la comunidad del ámbito tecnológico respecto al *deepfake*. Por lo tanto, el enfoque de investigación es cualitativo, con un tipo de investigación exploratoria de tipo documental, con una técnica de recolección y análisis de información, sean documentos, artículos, tesis, libros, reflexiones, ensayos, capítulos de libro, entre otros, evidenciando los resultados en una serie de mapas conceptuales, con los cuales se demuestra el desarrollo de los siguientes objetivos: identificar la información existente sobre el *deepfake* con base en el estado del arte. Seguido de ello, contratar los principales usos del *deepfake* a través de un análisis cualitativo. Luego, describir las actividades realizadas con el uso incorrecto de tecnologías audiovisuales y materiales producto de inteligencia artificial que se han usado para difamar y engañar. Y, por último, orientar sobre el uso de diverso software para la detección del *deepfake*.

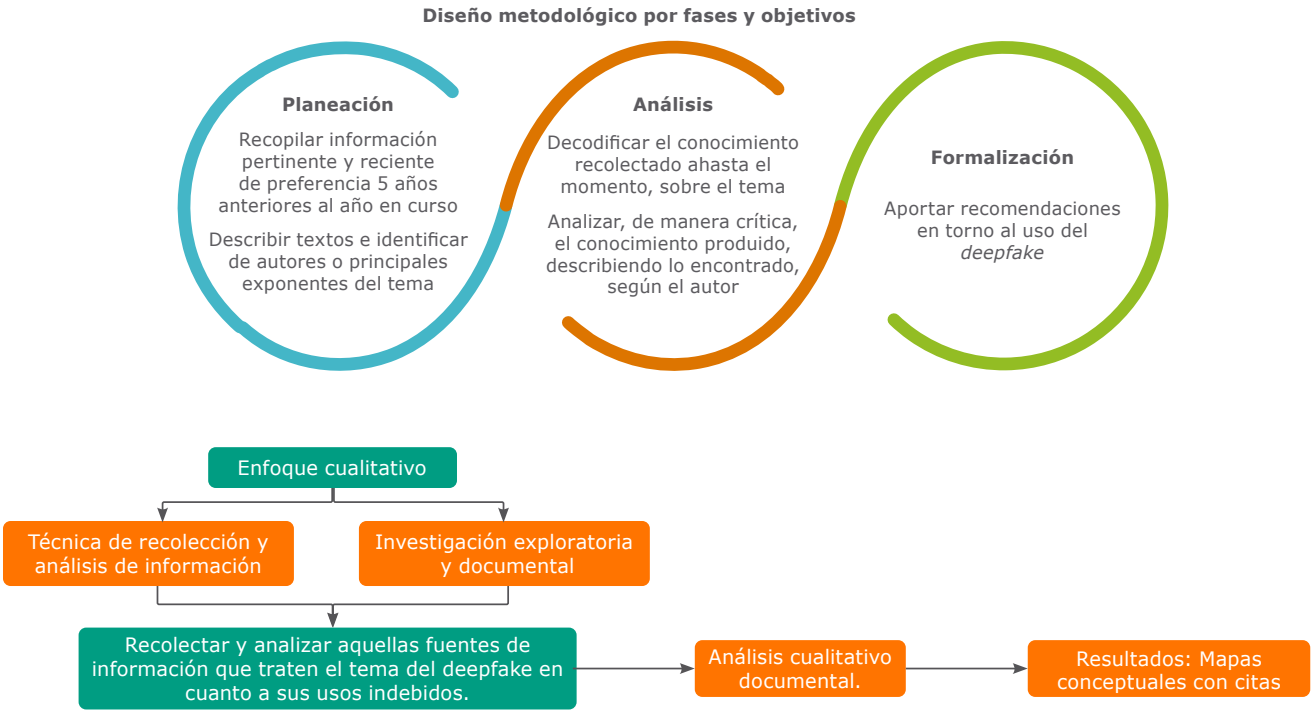

Figura 1. Metodología y Diseño Metodológico

Resultados

M. Westerlund, en su artículo de "The Emergence of *Deepfake* Technology: A Review", presenta una breve explicación del concepto, en el que la palabra engloba los conceptos de "deeplearning" y el concepto "*fake*", creando así la palabra *deepfake*, ya que este proceso "implica introducir imágenes de dos personas en un algoritmo de aprendizaje profundo para entrenarlas para intercambiar caras (...) usan tecnología de mapeo facial e inteligencia artificial" (Westerlund, 2019). Así, el *deepfake* está creado con software de alta calidad; distinto del denominado cheap*fake*s. Este último, creado como falsificaciones baratas, ya que su resultado es menos eficiente, creíble y con más imperfecciones, dando como resultado videos de baja calidad.

Para entender cómo funciona el *deepfake*, hay que entender que esta tecnología funciona como funciona un cerebro humano. Es decir, se hace uso de redes neuronales para actividades automáticas, como el reconocimiento de un rostro o de fallos en la lógica, esto se hace a través del aprendizaje continuo y la experiencia del día a día, así, "el principal ingrediente tecnológico en la creación de *deepfakes* es el aprendizaje profundo, una técnica de aprendizaje automático de IA que se puede utilizar para entrenar redes neuronales profundas" (Westerlund, 2019). Así, el software reconocería el rostro de una persona humana en específico, por lo que "[...] esta tecnología se basa en algoritmos sofisticados en los que una IA genera imágenes de personas y una segunda IA adivina si las imágenes son reales o falsas. De esta manera, las IA mejoran cada vez más en lo que están haciendo" (Gonzalez, 2020)

Por ejemplo, este procedimiento funciona como si se tratase de retratar un rostro humano, si se quiere dibujar el rostro de alguien, primero, el dibujante observaría una serie de diversas fotografías existentes, ángulos, expresiones, etc., con el fin de recolectar información para ilustrarlo, El *deepfake* recoge múltiples imágenes y se le entrena al software para reconocer características del rostro, para que, posteriormente, este pueda recrear imágenes únicas, "este proceso de reconocer primero un número comparativamente pequeño de características faciales en la entrada y, a partir de ahí, generar caras de aspecto real, como salida se logra en tres subpartes de codificadores automáticos: un codificador, un espacio latente y un decodificador" (Kietzmann, J., Lee, L. Mccarthy, I. & Kietzmann, T, 2019)

De acuerdo con lo anterior, dentro de este ámbito se pueden diferenciar dos tipos de categorías, el primero es el deepface y el segundo, el *deepvoice*. El primero, ya explicado anteriormente como aquel en el que se superpone un rostro sobre otro para falsificar sus gestos, y el segundo, "*deepvoice*: en este otro caso se trataría de unir frases y palabras sueltas utilizadas por una persona para crear un discurso. Incluso, es capaz de clonar la voz original a partir de estos fragmentos" (OSI, 2020) ambos, en la mayoría de los casos, van de la mano, ya que permite así, que haya más similitud con la realidad, por lo tanto, un resultado más creíble para el espectador.

Teniendo en cuenta cómo funciona el *deepfake*, se tiene en cuenta una de las principales problemáticas existentes en este ámbito y es la tendencia a usarse para la desinformación, donde la desinformación abarca no sólo el ámbito del entretenimiento, sino también, el ámbito militar y de seguridad en general. Debido a ello, existen métodos para la detección y explicación de los mismos, ya que entre más avancen las técnicas para la creación de *deepfake*, más tendrán que avanzar los métodos para combatirlo, "hay una guerra fría entre los métodos de detección y generación de *Deepfake*. Como hay mejoras en uno, provoca desafíos para el otro." (Masood, M. Nawaz, K. Mahmood, A. Javed, A. & Irtaza, 2021) esto permite, que cada vez haya mayores indicios sobre cómo contener la creación de vídeos falsos y brindar en la comunidad de las TIC avances en cuanto al uso de software.

El autor Barbas, D. abarca la información desde la preocupación sobre la generación de noticias falsas gracias a la alteración de lo que vemos como real y de calidad, ya que, gracias a los rápidos avances tecnológicos, la población tiene, cada vez más, acceso y facilidad de aprender sobre el uso de software para alterar información. "El uso incorrecto de estas tecnologías permite la creación de perfiles falsos, saltarse mecanismos de reconocimiento facial o comprometer a individuos haciéndoles aparecer en actos en los que no han formado parte" (Barbas, 2020), y si cada vez es más sencillo crearlo o aprenderlo, más inconvenientes con el *deepfake* habrán.

A continuación, en la Figura 2, se resumen, de manera conceptual, los resultados del primer objetivo de esta investigación, basado en identificar la información existente sobre el *deepfake* con base en el estado del arte, así, se tienen en cuenta los autores anteriormente mencionados para su desarrollo.

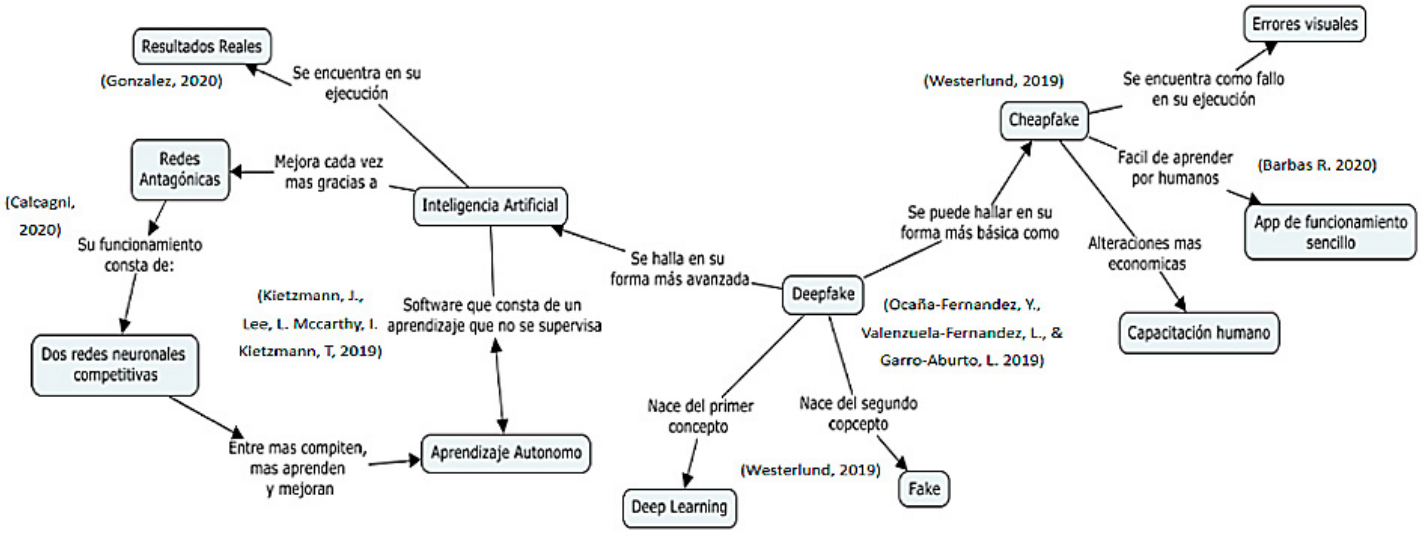

Figura 2. Información existente sobre el *deepfake* con base en el estado del arte

En el artículo de Government Accountability Office (2020), se expone que el *deepfake* es un video que ha sido manipulado con inteligencia artificial, reemplazando completamente un rostro y hasta las expresiones del mismo, exponiendo que estos representan básicamente lo que un usuario ve en un video *fake*, como hacer parecer ver algo que nunca se hizo y escuchar algo que nunca se dijo, lo cual permite que se puedan realizar los siguientes usos: dentro del entretenimiento, en el comercio electrónico y en la comunicación. Por ejemplo, se puede reemplazar a un actor cuando este no está disponible o no hay elementos suficientes en la escena de grabación, también se puede hacer publicidad con una celebridad fallecida de la cual se tengan suficientes imágenes, o bien, hacer una manipulación facial para que parezca que una persona está hablando otro idioma o diciendo algo diferente a lo que en realidad está expresando.

El *deepfake* se ha usado en diversos ámbitos, estos han evolucionado acorde con su usabilidad y complejidad. Según el Centro Belfer de Ciencias y Asuntos Internacionales (2020) estas son algunas aplicaciones más populares del *deepfake*: sátira política y social, imágenes publicitarias, reconstrucción de voz, desinformación política, pornografía, chantaje y extorsión, entretenimiento amateur, fraude financiero y ciberdelincuencia. De ahí que, "los *deepfakes* se han utilizado para difundir desinformación y desinformación sobre funcionarios públicos y cuestiones políticas, alterar sin consentimiento el contenido pornográfico y desarrollar entretenimiento amateur en aplicaciones de redes sociales" (Centro Belfer de Ciencias y Asuntos Internacionales, 2020). Teniendo en cuenta lo anterior, la información será cada vez más difícil de distinguir; si se observa un video de humor de un personaje importante, así se vea real, se podría discernir de lo falso a lo real por su contenido, pero si se ve al mismo sujeto realizando actos ilícitos y es catalogado como real, la percepción de lo real o ficticio puede variar y colocarse en duda, sólo se podría confiar, en este caso, de la fuente de información.

La autora Andrea Hauser (2018) brinda una perspectiva más amplia en cuanto a los diversos tipos de aplicaciones que se utilizan, más allá de mencionar los campos en que esta tecnología es usada, brinda un claro ejemplo de que el *deepfake* no se limita solo al intercambio de caras, sino que cualquier cosa, integrada en el video, puede ser reemplazada, fácilmente, por cualquier elemento que se desee integrar si cumple con característi-

cas similares. Esto puede ser de gran ayuda en el campo del entretenimiento, pero una gran desventaja en otros campos de índole social o de seguridad, pues alterar no solo a una persona sino el escenario donde se encuentra, dificulta el discernir la realidad y puede complicar una evidencia en un caso criminal, por ejemplo.

Dentro de los campos en los que la tecnología del *Deepfake* puede ser usada para que causen impactos positivos en la sociedad, además del estudio del comportamiento de la inteligencia artificial, son algunos ámbitos que mencionan en AECOC como: la protección de identidad, que gracias a la IA, se limita la identificación de entrevistados en casos como los que se han presentado en Rusia con la comunidad LGBTQ, también, el ahorro de costos económicos en vídeos de publicidad y videojuegos, reemplazo de personas como en los videos de entrenamiento y deporte, cambiar diálogos y escenarios en escenas de películas para evitar nuevas grabaciones y, por último, aplicaciones que aún se están desarrollando como "en la medicina, donde la tecnología generativa ya se está utilizando para crear escáneres cerebrales «falsos» basados en datos reales de pacientes. Estos escaneos falsos se utilizan para entrenar algoritmos para detectar tumores en imágenes reales" (AECOC, 2020)

De acuerdo con lo anterior, se presenta un resumen de lo previamente mencionado, esta Figura 3, corresponde al segundo objetivo de esta investigación, el cual consiste en contrastar los principales usos del *deepfake* a través de un análisis cualitativo, por lo que se presenta de forma concisa los usos de esta tecnología sean positivos o negativos.

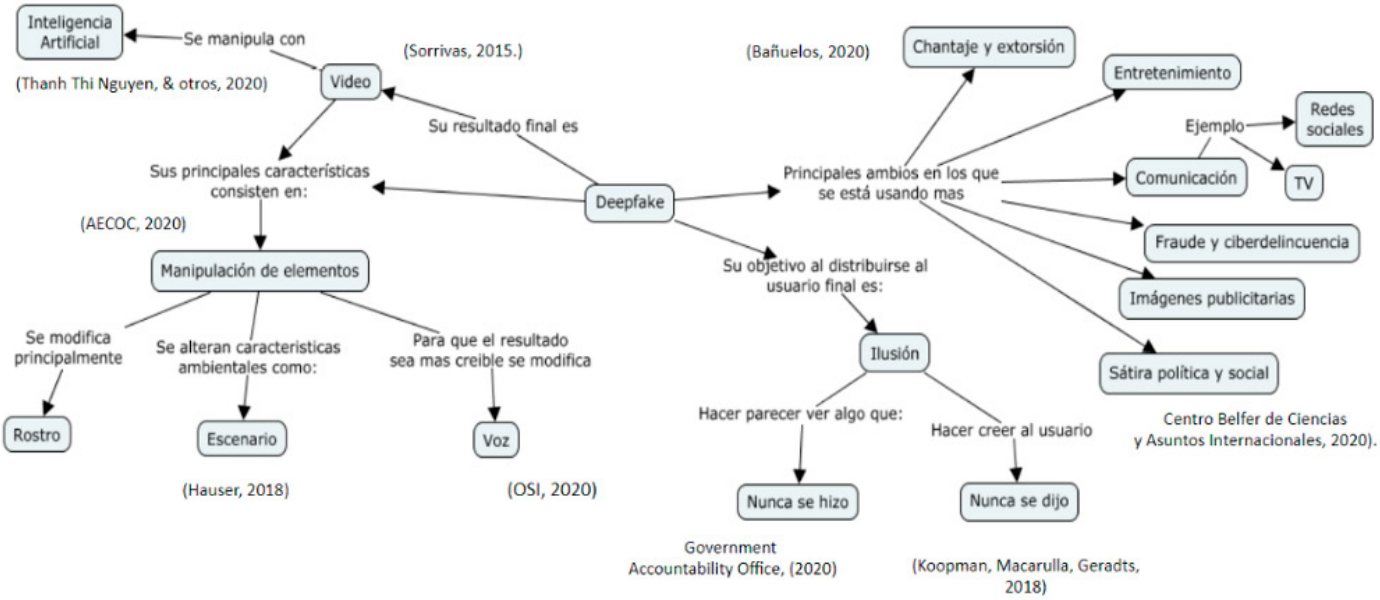

Figura 3. Contraste de los principales usos del *deepfake* a través de un análisis cualitativa
Fuente. Elaboración propia

La autora Samantha Cole brinda una mirada a uno de los graves problemas que puede enfrentar cualquier persona inocente que es víctima de alguna venganza, como sería el caso de una app creada para crear la imagen de una mujer desnuda, esta app, DeepNude toma como base la foto de la víctima y la anexa a otras fotos de mujeres desnudas tomadas de internet, creando así el resultado e ilusión de que ese cuerpo le pertenece a la víctima. Al respecto, la autora comenta que "la mayoría de casos, los *deepfakes* han sido utilizados como un arma contra las mujeres y sin su consentimiento, y gran parte de los titulares y el miedo político que tienen se centran en el potencial de las noticias falsas". (Cole, 2019).

Cada día y gracias a los avances tecnológicos salen más noticias falsas que, dependiendo de cómo las vea el público, se tornan creíbles, actua-

les y verídicas, cada día al usuario de Internet se le aborda con gran cantidad de noticias que, puede que no filtre a razón de buscar de lo que observó qué es real y qué es falso, "a más cantidad de información no garantiza la calidad de esta. Por tal razón, promover acciones de alfabetización digital siempre es una idea pertinente, vigente, y socialmente responsable" (Amorós, 2018), si se capacita digitalmente, menos incrédulos de noticias falsas se hallarán difundiendo y haciendo virales más *fake* news o videos *deepfake* que afecten la vida de alguna persona o cree un malestar en la comunidad, así, "con la aparición de las redes sociales, el ecosistema comunicativo ha cambiado [...] ya no es una responsabilidad de los medios, como en el siglo XX. Ahora muchas mentiras se mueven por Internet con el peligro de que otros las copien". (Alsina, Cerqueira, 2019).

Así, el ámbito ético recae sobre el usuario que, crea o difunde la información que se comparte en la internet, sea como en los casos anteriormente mencionados, creaciones de fotos desnudas de mujeres inocentes, al intercambiar sus rostros con el cuerpo de otras mujeres, o sea la difusión de noticias falsas gracias a que algunos creen en los *deepfake* como algo verídico y contribuyan a su difusión, la gran preocupación es que gracias a "las redes sociales y las noticias digitales ha aumentado la responsabilidad ética de la gente que opera en este campo, especialmente, teniendo en cuenta el alcance global y el poderoso impacto de estos nuevos tipos de medios de comunicación" (UNODC, 2019). Antes, la responsabilidad recaía netamente en los medios de comunicación, ahora, es deber del internauta no difundir o recrear la noticia, es una responsabilidad y deber ético, no contribuir con la afectación de la difamación de hechos o personajes a través de sus propias redes sociales.

Para distanciar al usuario de difundir noticias o videos falsos, La Federación Internacional de Periodistas (2018) brinda una serie de pautas que el internauta puede seguir para no caer en la difusión de las mismas, como lo son, el analizar la sección "Acerca de" o "About" del sitio, ya que en este, se incluyen datos que se podrían verificar, así como del autor de la noticia, también, revisar los links y las citas, ya que estas son el respaldo o fuentes de la noticia, y si estos redirigen a otros sitios, probablemente la noticia sea *fake*, por último, también se recomienda realizar una búsqueda inversa de imágenes, con el fin de desestimar una información que no sea real y cuyas imágenes sean copiadas de otros artículos. Estas pautas son necesarias ya que "las personas son los principales culpables de la propagación de las noticias falsas, ya que difunden un mayor número de bulos que los bots" (Servimedia, 2018).

Una de las soluciones parciales que pueden existir hoy día frente a la difusión de información falsa y videos que atenten contra la dignidad de las personas en las redes sociales, son estas mismas, al respecto, "para detener la propagación de videos falsos, Facebook, Google y Twitter tendrían que demostrar que pueden cumplir sus recientes promesas de vigilar sus plataformas" (Pierson, 2018), así, aunque la ética de la persona no frene la propagación de algún video que esté perjudicando a una víctima, la misma red social se encargará de poner un alto a su difusión, siempre y cuando, esta cumpla con su deber de regular las publicaciones. Así, García, plantea un ejemplo de Facebook, una de las redes sociales más populares en Occidente, en cuanto a qué regulan respecto al *Deepfake*, cumpliendo, dentro de sus políticas eliminar contenido del *Deepfake* siempre y cuando cumpla con los criterios de: "ha sido editado (...) de una manera que no quede claro para una persona promedio y que probablemente sirvan para engañarle (...), si es producto de una inteligencia artificial (...) para que parezca que es auténtico." (García, 2020) aunque hay una excepción no sólo en esta red social, sino en la mayoría, y es permitir y no borrar aquel contenido que sea de naturaleza humorística o bien, se resuma el video sin cambiar su mensaje final. Aún con estos criterios, se usan por el momento, verificadores humanos, para descartar contenido que no cumpla las normativas de la red social, "si uno de estos verificadores califica una foto o vídeo como falso o potencialmente falso, se reducirá significativamente su distribución en el news feed, o será rechazado si se trata de un anuncio" (Galeano, 2020) lo cual propicia que menos usuarios observen contenido falso además de impedir su viralización.

Pese a que algunas redes sociales se esmeran en regular su contenido, algunas páginas no consideran de este, un problema, un ejemplo de ello

son las páginas de contenido pornográfico, que, gracias a los *Deepfake*, cuentan con más videos y usuarios que esperan ver a sus celebridades favoritas en videos o imágenes que esta nunca ha creado.

"Hasta julio de 2019 había menos de 15.000 *deepfakes* circulando por la web. Un año después, la cifra creció a casi 50.000. El 96% son pornográficos y en lo que va de 2020 ya se subieron más de mil *deepfakes* por mes solo en sitios de pornografía, donde aparecen con cada vez más frecuencia supuestos «videos prohibidos» de celebridades e influenciadores. Las compañías detrás de la web porno no consideran que esto sea un problema" (Rodriguez, 2020).

Por el momento hay un pequeño respiro en cuanto a los peligros del *Deepfake*, y es el hecho de que no todos pueden o saben realizar un contenido de calidad que a simple vista parezca real, al respecto, hay un caso que se hizo viral como el del usuario @deeptomcruise de TikTok, esta persona, personificaba al verdadero Tom Cruise por medio de estas tecnologías, viéndose tan real para llegar al punto en el que los usuarios dudaban si este era *fake* o real, haciéndose viral, "Chris Ume es quien está detrás de estos clips virales y explicó en una entrevista en The Verge la de horas y esfuerzo detrás de estos vídeos, además de trabajar con Miles Fisher, imitador de Tom Cruise: <<No puedes hacerlo con solo presionar un botón>>". (Reino, 2021), lo cual es un posible alivio teniendo en cuenta la cantidad de personas que desean recrear situaciones de celebridades o ciudadanos comunes en escenarios de contenido ilegal, ilícito, sexual, entre otros.

De acuerdo a lo anterior, se presenta la Figura 4, el cual resume detalladamente uno de los objetivos más importantes, el cual consiste en describir las actividades realizadas con el uso incorrecto de tecnologías audiovisuales y materiales producto de inteligencia artificial que se han usado para difamar y engañar, esto, de acuerdo a los autores y a los casos encontrados que ejemplifican la repercusión ética, tanto en víctimas o sus creadores.

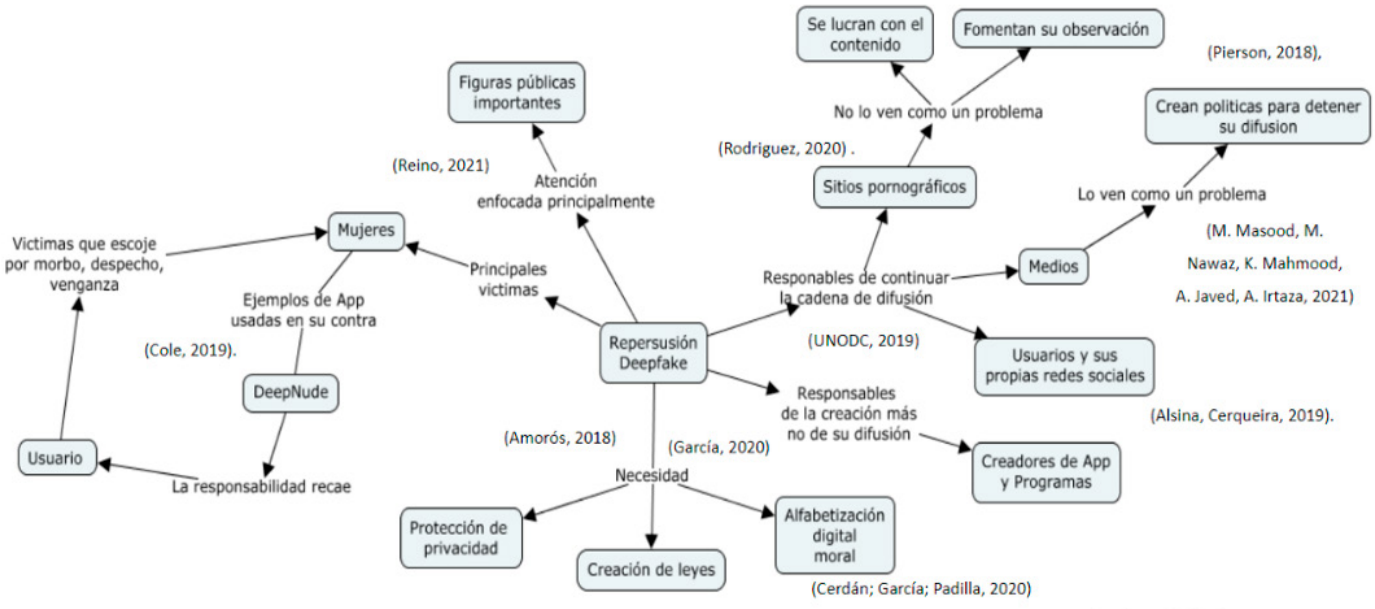

Figura 4. Descripción de las actividades realizadas con el uso incorrecto de tecnologías audiovisuales y materiales producto de inteligencia artificial que se han usado para difamar y engañar
Fuente. Elaboración propia

Una de las mayores responsabilidades que recaen sobre el uso del *Deepfake* es del usuario quien reproduce la información, si el contenido ya está creado, no tiene porqué seguir siendo difundido y viralizado en internet, cada uno puede detener su difusión si toma responsabilidad y detiene la

cadena de emisión, realizando acciones como bloqueo del contenido, dejar de compartir, evitar mirar el contenido, no brindar likes ni comentarios, no brindar aquellos aspectos que permiten que las publicaciones alcancen mayor visibilidad.

Para que el usuario logre distinguir un *deepfake* de un video real y así evitar perjudicar a una persona inocente, se tienen diversas pautas para la detección del *deepfake*, uno de los más mencionados es el parpadeo de la persona que esté hablando en el video, al respecto:

"Los humanos solemos parpadear una vez cada 2-8 segundos durando cada parpadeo entre 1 y 4 décimas de segundo. Si quieres saber si estás ante una *Deepfake*, fíjate en el número de veces que parpadea, te darás cuenta de que lo hace mucho menos que una persona real. El algoritmo de los *Deepfake* no puede parpadear a la misma velocidad que un humano" (Galiana, 2019)

Así, al estar frente a un video muy realista en el que no se tenga seguridad si lo que se ve es real o no, el mayor enfoque debe estar en los ojos del hablante; pese a que hay personas que pueden parpadear más despacio que otras, una singular manera lenta de parpadear, con poco ritmo o de forma extraña, debe alertarnos para fijarnos en demás elementos del video para hallar más inconsistencias. según Galiana (2019), los elementos a observar serían la consistencia entre el cuello y la cara, la duración del video, la fuente de la grabación, los sonidos, algunos detalles como discrepancias en el fondo o los cambios repentinos sea en texturas, las luces, colores, entre otros y, por último, el interior de la boca.

Aunque se cuenten hoy en día con diversas pautas para su detección, no todas las personas cuentan con el tiempo suficiente para fijarse en los pequeños detalles, por lo que se cuenta con inteligencia artificial que analiza cualquier video sospechoso de ser un *Deepfake*, al realizarlo de manera automática, facilita tiempo por lo que se puede detener de una forma más oportuna su difusión, por ejemplo, este se encarga de "detectar el número de fotogramas en que las personas aparecen con los ojos cerrados y dictaminar a partir de ese dato si se trata o no de un *deepfake*. Su tasa de detección es de más del 95%". (Merino, 2019)

Otra pauta para la detección del *Deepfake* recae en la creación de mejores software, que automáticamente, puedan hallar videos falsos, en el que la inteligencia artificial detecta por sí misma los aspectos mencionados anteriormente y halle otras inconsistencias que el ojo humano no pueda observar fácilmente, actualmente, se sigue trabajando y mejorando programas como MediFor, el cual puede reconocer en videos discrepancias, pero que actualmente aún siguen mejorando, paradójicamente, entre mejor sea un *deepfake*, más se puede aprender de él y mejorar la tecnología para detectarlo.

"DARPA ha tenido dos programas dedicados a la detección de falsificaciones profundas: Media Forensics (MediFor) y Semantic Forensics (SemaFor). MediFor, (...) desarrolló algoritmos para evaluar automáticamente la integridad de fotos y videos (...) exploró técnicas para identificar las inconsistencias audiovisuales presentes en las falsificaciones profundas, incluyendo inconsistencias en píxeles (integridad digital), inconsistencias con las leyes de la física (integridad física) e inconsistencias con otras fuentes de información (integridad semántica)". (Congressional Research Service, 2021)

El *Deepfake* también aprende de las mejoras constantes para pulir su funcionamiento, por ejemplo, se descubrió "que los primeros métodos de *Deepfake* no podían imitar la velocidad a la que una persona parpadea; mientras que los programas recientes han solucionado la falta de parpadeo o parpadeo no natural después de que se publicaron los hallazgos" (Gonzalez, 2020) por lo que muchos creadores de *deepfake* están al corriente de las mejoras tecnológicas para mejorar, lo cual supone un peligro para su detección y contención del problema.

A continuación, se presenta la última figura correspondiente al objetivo final de la investigación, el cual consiste en orientar sobre el uso de diversos softwares para la detección del *deepfake*, en el que se resumen algunas pautas para la orientación actualizada en cuanto a la mejora de programas que lo detecten y combatan, para evitar así, la mala praxis de acuerdo a los autores anteriormente mencionados.

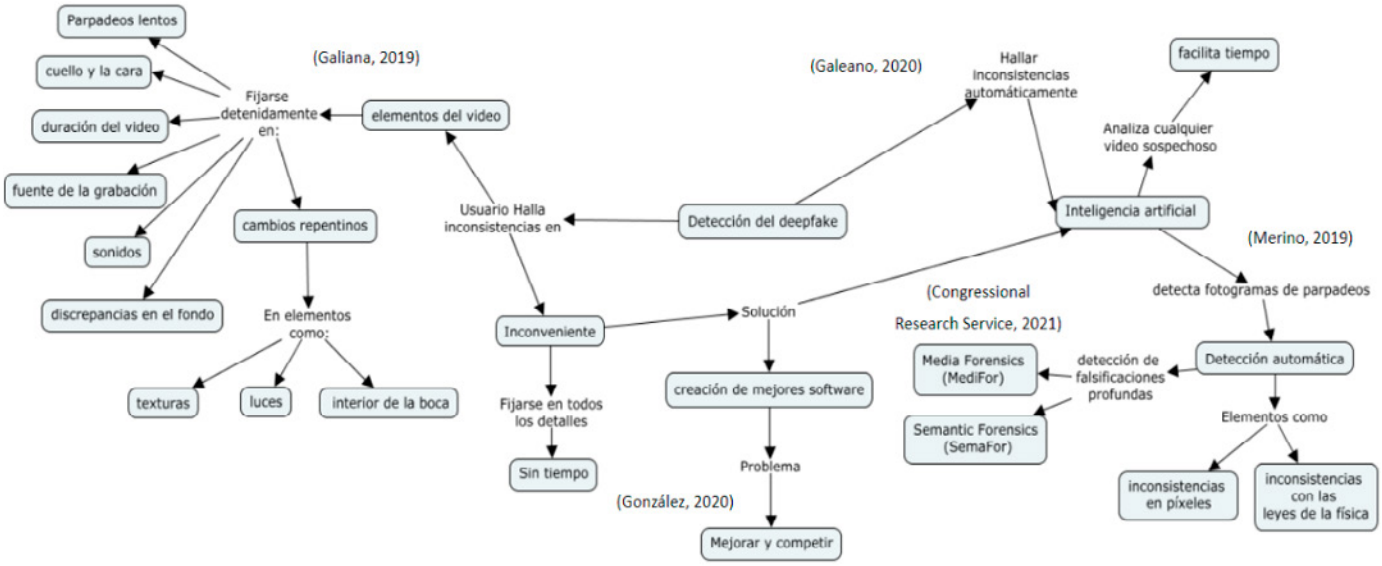

Figura 5. Orientación sobre el uso de diverso software para la detección del *deepfake*.
Fuente. Elaboración propia

Conclusiones

Gracias al estudio de los casos documentados del uso del *Deepfake* por sus principales exponentes, se puede evidenciar y recolectar información pertinente para combatir el uso indebido de este, advirtiendo en la comunidad perteneciente a las TIC las principales repercusiones éticas que afectan a individuos inocentes y sus principales consecuencias sociales y legales que influyen en el bienestar social; gracias al análisis de las consecuencias del *Deepfake*, se brindan estrategias que pueden servir de guía para crear mejores software de detección y pautas éticas en la comunidad de las TIC, por lo que se puede concienciar a la comunidad que hace parte de las TIC para advertir y evitar la continuidad de la mala praxis, creando conciencia de no distribuir ni crear material que afecte gravemente la vida de las personas, su reputación, libertad o resolución de conflictos como es el caso de evidencias falsas o alteradas en audiencias y situaciones legales.

Referencias:

AECOC, (2020). *"La capacidad de la IA para crear deepfakes: ¿Cuáles son los potenciales impactos positivos?"* AECOC, INNOVATION HUB. Barcelona, *Deepfakes*: ¿Cuáles son los potenciales impactos positivos? (aecoc.es)

Alsina, Rodrigo, Cerqueira, L. (2019). *Periodismo, ética y posverdad*. Cuadernos.info, (44), 225-239. https://doi.org/10.7764/cdi.44.1418

Amorós García, M. (2018). *Fake news, la verdad de las noticias falsas*. Plataforma Editorial. 1.a Ed. España. ISBN: 978-84-17114-72-5 https://www.tse.go.cr/revista/art/28/*fake*_news.pdf

Bañuelos Capistrán Jacob. (2020). *Deepfake: la imagen en tiempos de la posverdad*. Universidad Autónoma de Baja California, Revista Panamericana de Comunicación | Año 2 No. 1 enero - junio 2020.

Barbas Rebollo, David. (2020) *Detección de noticias falsas y caras de personas manipuladas*. Universitat Politècnica de València https://riunet.upv.es/bitstream/handle/10251/149028/Barbas%20-%20Detecci%C3%B3n%20de%20noticias%20falsas%20y%20caras%20de%20personas%20manipuladas.pdf?sequence=1

Calcagni, Laura R. (2020). *Redes Generativas Antagónicas y sus aplicaciones*. Universidad de la Plata.

Centro Belfer de Ciencias y Asuntos Internacionales. (2020). *Deepfakes*. Tech factsheets for policymakers. Spring 2020 Series. Harvard Kennedy School 79 John F. Kennedy Street, Cambridge, MA 02138.

Cole, S. (2019). *Esta terrorífica app crea un nude de cualquier mujer con un simple clic*. Vice. Consultado el 3 de febrero de 2019 en: https://bit.ly/2Ii3WbS

Congressional Research Service (2021). *Deep Fakes and National Security. Library of Congress*. IF11333. Washington, D.C.: oct. 14, 2019. https://crsreports.congress.gov/product/pdf/IF/IF11333

Federación Internacional de Periodistas (The International Federation of Journalists - IFJ). (2018). *¿Qué son las fake news? Guía para combatir la desinformación en la era de post verdad*. https://www.ifj.org/fileadmin/user_upload/Fake_News_-_FIP_AmLat.pdf

Galeano, Susana. (2020). *Qué son los deepfakes... y así se prepara Facebook para eliminarlos de sus redes sociales*. Qué son los *deepfakes*... y así se prepara Facebook para eliminarlos de sus redes sociales - Marketing 4 Ecommerce - Tu revista de marketing online para e-commerce

Galiana, Patricia. (2019). *¿Qué son los Deepfakes y cómo detectarlos?* https://www.iebschool.com/blog/*deepfakes*-como-detectarlas-business-tech/

García, Jose. (2020). *Las redes sociales frente a los deepfakes: un repaso a las políticas de Facebook, Instagram, Twitter y LinkedIn. Xataka México, Sección, inteligencia artificial*. Las redes sociales frente a los *deepfakes*: un repaso a las políticas de Facebook, Instagram, Twitter y LinkedIn (xataka.com)

GAO U.S. Government Accountability Office. (2020). *SCIENCE & TECH SPOTLIGHT: DEEPFAKES*. https://www.gao.gov/assets/gao-20-379sp.pdf

Geradts, Zeno, Koopman, Marissa, Macarulla, Andrea. (2018). *Detection of Deepfake Video Manipulation*. University of Amsterdam & Netherlands Forensic Institute.

Gonzalez, Yolanda. (2020). *Qué son los Deepfakes y cómo protegerte de ellos*. Grupo Atico 34, Madrid. ¿Qué son los *Deepfakes* y cómo protegerte de ellos? (protecciondatos-lopd.com)

Hauser, Andrea. (2018). *Deepfake - An Introduction*.

KIETZMANN, J., LEE, L., MCCARTHY, I. and KIETZMANN, T. (2019). *Deepfakes: trick or treat?* Business Horizons. ISSN 0007-6813. https://core.ac.uk/download/pdf/250590695.pdf

Masood Momina, Mariam Nawaz, Khalid Mahmood Malik, Ali Javed, Aun Irtaza (2021). *Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward*. https://arxiv.org/ftp/arxiv/papers/2103/2103.00484.pdf

Merino, Marcos, (2019) *Así es posible saber si un vídeo es un deepfake con sólo un abrir y cerrar de ojos, literalmente*. https://www.xataka.com/inteligencia-artificial/posible-saber-video-*deepfake*-solo-abrir-cerrar-ojos-literalmente-quizas-eso-no-sea-suficiente

Ocaña-Fernandez, Y., Valenzuela-Fernandez, L., & Garro-Aburto, L. (2019). *Inteligencia artificial y sus implicaciones en la educación superior. Propósitos y Representaciones*, 7(2), 536-568 p. 540. http://www.scielo.org.pe/pdf/pyr/v7n2/a21v7n2.pdf.

OSI, Oficina de Seguridad para el Internauta (2020). *"Deepfakes, ¿cómo se aprovechan de esta tecnología para engañarnos?"* Blog: *Deepfakes*, ¿cómo se aprovechan de esta tecnología para engañarnos? | Oficina de Seguridad del Internauta (osi.es)

Padilla Castillo. G; Cerdán Alenda, V. (2020). *Alfabetización moral digital para la detección de deepfakes y fakes audiovisuales en CIC*. Cuadernos de Información y Comunicación 25, 165-181.

Pierson, David. (2018), *Videos falsos: mientras el realismo aumenta, ver no siempre debe implicar creer*. Los Angeles Times. https://www.latimes.com/espanol/eeuu/la-es-videos-falsos-mientras-el-realismo-aumenta-ver-no-siempre-debe-implicar-creer-20180219-story.html

Reino, Antia. (2021). *Los 'deepfakes' y el debate sobre la suplantación de identidad con el uso de IA vuelven con fuerza a las redes sociales*. RTVE, Corporación de Radio y Televisión Española.

Rodríguez Ansorena, Tomas. (2020). *El fin de la realidad ¿Qué son los deepfakes*? Nueva sociedad. El fin de la realidad ¿Qué son los «*deepfakes*»? | Nueva Sociedad (nuso.org)

Servimedia (2018), *Influencia de las noticias falsas en la opinión pública,* estudio de comunicación en colaboración con Servimedia. https://www.servimedia.es/sites/default/files/documentos/informe_sobre_*fake*_news.pdf

Sorrivas, Nicolás. (2015). *El videoarte como herramienta pedagógica* fue publicado de la página 81 a página 94 en Cuadernos del Centro de Estudios en Diseño y Comunicación Nº 52

Thanh Thi Nguyen, Cuong M. Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, Saeid Nahavandi. (2020). *Deep Learning for Deepfakes Creation and Detection: A Survey*, Arxiv: 1909.11573v2. https://arxiv.org/pdf/1909.11573.pdf

UNODC, Oficina de las Naciones Unidas contra la droga y el delito, (2019) *Integridad y ética de los medios de comunicación*. Educación para la justicia serie de módulos universitarios. https://www.unodc.org/documents/e4j/IntegrityEthics/MODULE_10_-_Media_Integrity_and_Ethics_-_Spanish.pdf

Westerlund, Mika. (2019) *The Emergence of Deepfake Technology: A Review*. Technology Innovation Management Review, vol. 9, no. 11: pp. 39-52. https://timreview.ca/sites/default/files/article_PDF/TIMReview_November2019%20-%20D%20-%20Final.pdf