CC BY

ARTÍCULOS

CIC. Cuadernos de **Información y Comunicación**
ISSN: 1135-7991

EDICIONES COMPLUTENSE

http://dx.doi.org/10.5209/ciyc.68762

Alfabetización moral digital para la detección de *deepfakes* y *fakes* audiovisuales

Víctor Cerdán Martínez[1]; María Luisa García Guardia[2]; Graciela Padilla Castillo[3]

Evaluado: 17/04/2020 / Aceptado: 25/05/2020

Resumen. Los *deepfakes* son vídeos manipulados donde se suplanta la cara de una persona por la de otra a través de Inteligencia Artificial. Estos contenidos fotorrealistas pueden convertirse en armas de acoso, propaganda o conflicto social. Esta investigación de innovación teórica ahonda en la incipiente literatura científica sobre el *deepfake* en Ciencias Sociales, poniendo sus bases en los trabajos de Wenceslao Castañares. Sus objetivos son analizar el fenómeno, buscar herramientas para detectarlo y combatirlo, e inaugurar una nueva vía de alfabetización moral digital de especial interés en el campo de la comunicación, por su importancia social y política. Se emplea una metodología diacrónica en dos pasos. En primer lugar, se realiza una investigación bibliográfica longitudinal y se ofrece una descripción histórica, sistémica y descriptiva, con los rasgos más importantes del fenómeno, naturaleza, funcionamiento y posibles usos. En segundo lugar, se realiza una observación científica consciente de más de medio centenar de vídeos *deepfake*, para percibir el fenómeno, conseguir una mayor generalización y garantizar la validez de los resultados. Los resultados muestran que aunque se han creado softwares de detección de este tipo de *fakes*, aún no son de acceso libre y gratuito. La única manera de reconocer el *deepfake* es a través de una educación moral digital, que pueda detectar una serie de parámetros visuales hiperrealistas: frecuencia del parpadeo, efecto intermitente de las caras o transiciones entre cabeza y cuello.
Palabras clave: Alfabetización digital; *Deepfake*; *Fake* audiovisual; Inteligencia artificial; Software anti-*fake*; Parámetros visuales.

[en] Digital moral literacy for the detection of *deepfakes* and audiovisual *fakes*

Abstract. Deepfakes are manipulated videos where a person's face is supplanted by that of another through Artificial Intelligence. These photorealistic contents can become weapons of harassment, propaganda or social conflict. This theoretical innovation research delves into the emerging scientific literature on deepfake in Social Sciences, laying its foundations on the research of Wenceslao Castañares. Its objectives are to analyse the phenomenon, find tools to detect and combat it, and inaugurate a new way of moral digital literacy of special interest in the field of communication, due to its social and political importance. A two-step diachronic methodology is used. First, a longitudinal bibliographic investigation is carried out and a historical, systemic and descriptive description is offered, with the most important features of the phenomenon, nature, operation and possible uses. Second, there is a conscious scientific observation of more than fifty deepfake videos, to perceive the phenomenon,

[1] Universidad Complutense de Madrid
vicerdan@ucm.es
[2] Universidad Complutense de Madrid
mluisagarcia@ccinf.ucm.es
[3] Universidad Complutense de Madrid
gracielp@ucm.es

achieve greater generalization and guarantee the validity of the results. The results show that although detection software for this type of fakes has been created, it is not yet freely available. The only way to recognize the deepfake is through a digital moral education, which can detect a series of hyper-realistic visual parameters: the frequency of blinking, the subtle intermittent effect of the faces or the transitions between the head and neck.
Keywords: Digital literacy; Deepfake; Audiovisual fake; Artificial intelligence; Anti-fake software; Visual parameters.

[pt] Alfabetiçaçâo moral digital pra detectar os deepfakes e os fakes audiovisuais

Resumo. Deepfakes são vídeos manipulados onde o rosto de uma pessoa é suplantado pelo de outra através da Inteligência Artificial. Esses conteúdos fotorrealistas podem se tornar armas de assédio, propaganda ou conflito social. Esta pesquisa de inovação teórica investiga a literatura científica emergente sobre deepfake em Ciências Sociais, lançando suas bases nas investigações de Wenceslao Castañares. Seus objetivos são analisar o fenômeno, encontrar ferramentas para detectá-lo e combatê-lo e inaugurar uma nova forma de alfabetização moral digital de especial interesse no campo da comunicação, devido à sua importância social e política. Uma metodologia diacrônica de duas etapas é usada. Primeiramente, é realizada uma investigação bibliográfica longitudinal e é oferecida uma descrição histórica, sistêmica e descritiva, com as características mais importantes do fenômeno, natureza, operação e usos possíveis. Em segundo lugar, há uma observação científica consciente de mais de cinquenta vídeos profundos, para perceber o fenômeno, obter maior generalização e garantir a validade dos resultados. Os resultados mostram que, embora o software de detecção para este tipo de falsificações tenha sido criado, eles ainda não estão disponíveis gratuitamente. A única maneira de reconhecer o deepfake é através de uma série de parâmetros visuais: freqüência de piscar, efeito sutil intermitente das faces ou as transições entre a cabeça e o pescoço.
Palavras chave: Alfabetização digital; Deepfake; Falsa audiovisual; Inteligência artificial; Software anti-falso; Parâmetros visuais.

Sumario: 1. Introducción. 2. Método. 3. Resultados. 3.1. *Softwares* de detección de *deepfakes*. 3.2. Técnicas de observación visual. 4. Discusión. Fuentes de financiación. Referencias Bibliográficas.

Cómo citar: 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.

1. Introducción

La detección de vídeos *fake*, llamados *deepfakes*, es uno de los retos de la ciudadanía del siglo XXI. En la actualidad, existen varias técnicas para detectar si el vídeo o la imagen que se está percibiendo son reales o han sido intencionalmente manipulados. A pesar de que la Inteligencia Artificial (IA) avanza en la detección de *deepfakes*, la misma IA que crea esos contenidos falsificados también lo hace en el otro extremo. Así, el *deepfake* se erige como gran reto moral y cívico dentro de los nuevos lenguajes digitales o gramáticas generativas que han investigado algunos autores (Castañares, 1994; Aladro, 2007; Abril-Curto, 2010; Abril-Hernández, 2014; Aladro, 2017; Jivkova, Requeijo y Padilla, 2017; Márquez, 2017; Bernárdez y Padilla, 2018; Cerdán y Villa, 2019; Padilla y Presol, 2020).

Esta investigación original ahonda en la poca literatura científica de Ciencias Sociales sobre el *deepfake*, para analizar el fenómeno, buscar y comentar las herramientas para detectarlo y combatirlo, e inaugurar una nueva vía de alfabetización moral digital de especial interés en el campo de la comunicación, por su importancia social y política. Lo hace inspirándose en la obra del querido profesor Wenceslao Castañares, que estudió ampliamente la cultura visual (Castañares, 2007b), la realidad virtual (Castañares, 2011) y los valores que transmiten las imágenes mediáticas (Castañares, 1995; Castañares, Núñez y González, 1996; Castañares, 1997; Castañares, 2007b). Esta investigación aplicará las teorías del profesor Castañares sobre los valores y los sentimientos de la televisión moralista a un nuevo tipo de televisión, la desarrollada a través de Inteligencia Artificial y emitida a través de plataformas online. Esta línea de investigación la han seguido recientemente otros autores (Márquez y Tosca, 2017; Aladro, Jivkova y Bailey, 2018; Márquez y Ardèvol 2018; Fouce, Pecourt y Pedro-Carañana, 2018) desde enfoques multidisciplinares.

Los *deepfakes* (Koopman, Macarulla y Geradts, 2018; Cerdán y Padilla, 2019) son vídeos donde un usuario reemplaza la cara de alguien por la de otra persona. Estos audiovisuales ganaron penosa popularidad por su aplicación en productos de contenido sexual, donde los rostros de actrices pornográficas fueron cambiados por estrellas de Hollywood, de la televisión o de la música, y distribuidas en webs como Reddit y PornHub (Matsakis, 2018). Son posibles gracias a algoritmos gratuitos y muy fáciles de usar. El primero en hacerlo fue el usuario de la plataforma Reddit autodenominado *Deepfakes* (Cole, 2017). Seleccionó los rostros de las celebridades Gal Gadot, Maisie Williams o Taylor Swift, para incluirlas en el cuerpo de varias actrices de cine pornográfico (Cole, 2017; Cerdán y Padilla, 2019). El resultado fueron varios vídeos impúdicos, que parecen interpretados por las citadas famosas. Si, como bien asegura Castañares (2007a y 2007b), la expresión de lo que en un momento se consideró privado o íntimo, encuentran en la televisión un espacio privilegiado, en el *Deepfake* esto se lleva al extremo, ya que se falsifica esa intimidad. Y se hace a través de un ordenador casero y un algoritmo de *machine learning*, que cualquier persona puede descargarse de Internet.

Los vídeos se subieron a la plataforma Reddit, un sitio web de marcadores sociales, donde los usuarios pueden añadir texto, imágenes, vídeos o enlaces. Otros usuarios pueden votar a favor o en contra del contenido, haciendo que aparezcan más o menos destacados. Su público es mayoritariamente anglosajón y la mayoría de la actividad se realiza en inglés (Maeve, 2013; Cerdán y Padilla, 2019). Fue en dicha plataforma donde los primeros vídeos *deepfake* se popularizaron y propagaron. Según Beamonte (2018), la nueva moda ganó tantos adeptos que *Deepfakes* creó un subreddit, con su nombre de usuario, sólo para este tipo de vídeos y en sólo dos meses, contaba con 15.000 suscriptores. Esta comunidad extendió el uso de la palabra *deepfake* para referirse a todos los vídeos creados con IA. Castañares, Núñez y González (1996) aseguran que el gran problema de la televisión tradicional es la falta de regularización, sin embargo, esta idea se lleva al extremo en los *deepfakes* emitidos en Internet, donde a veces sobrepasan la barrera de lo legal con impunidad.

Otro usuario de Reddit, *Deepfakeapp*, decidió crear una aplicación llamada *FakeApp* para que cualquier persona, sin experiencia informática, pudiera hacer sus propios vídeos mediante IA (Beamonte, 2018). Esta app de escritorio está basada en el algoritmo del usuario *Deepfakes*, pero fue creada por el nuevo usuario, sin la ayuda de las *deepfakes* originales (Cole, 2017). Todas las herramientas que se necesitan

para hacer uno de estos vídeos son gratuitas y están disponibles junto con un tutorial para principiantes (Beamonte, 2018).

Las técnicas de entrenamiento de IA, como el aprendizaje automático, permiten incorporar decenas de fotografías a un algoritmo que crea máscaras humanas convincentes (Beamonte, 2018). Éstas reemplazan las caras de cualquier persona en un vídeo, mediante el uso de datos parecidos y permite que el software se capacite para mejorar con el tiempo. El problema directo es que este tipo de vídeos *fake* o *deepfakes* no sólo afecta a las celebridades (Pardo, 2018; Cerdán y Padilla, 2019). Igualmente se han recogido denuncias de mujeres anónimas, cuyos ex novios o amantes han usado la tecnología para vengarse de ellas y humillarlas en la red (Askham, 2018). Según la teoría de la televisión moralista de Castañares (2007a), estas acciones mediáticas constituyen un tipo de educación incidental e irreflexiva en la sociedad, es decir, una educación superficial basada en estímulos y emociones. En el caso de los *deepfake*s, estas emociones son creadas por individuos a través de Inteligencia Artificial y no por productoras o canales de televisión.

Las investigaciones del acoso digital son numerosas y realmente necesarias, sobre todo para acabar pronto con su existencia (Adorjan y Ricciardelli, 2019; Ashton, McDonald y Kirkman, 2019; Rice y Moffett, 2019). Sin embargo, la vía más practicada y perniciosa, el *sexting*, se centra en la difusión de vídeos originales, no manipulados (Linvill, 2019; Lucić, Baćak y Štulhofer, 2019; Setty, 2019). No serían *deepfakes* porque el o la protagonista fue grabado en la intimidad de la pareja que tiempo después, se disuelve. Una de las partes, como venganza, envía, comparte, publica y comenta fotografías y vídeos, que debieron quedar para la intimidad de cada uno de sus ex integrantes. Castañares (2007a) explica este tipo de acciones como un reflejo de los valores sociales, es decir, para el autor la televisión (y por extensión, el contenido *deepfake* publicado en Internet) representa un espejo de la sociedad en la que nos encontramos. "De esta manera se renuncia al ideal de otras épocas: la existencia de unos principios cuyo conocimiento permitía orientarse en el mundo" (Castañares, 2007a).

El *deepfake* tiene otra peligrosa variante social y política, que está muy cerca del estudio tradicional y actual de las *fake news*. Sería el referente al marketing político, mensajes electorales, o cualquier discurso, declaración o aparición de un mandatario o político. Farkas y Schou (2018) inciden en el poder de las *fake news* para desacreditar, atacar y deslegitimar a los opositores políticos. Atribuyen al término un significado flotante o líquido, que establece cómo es y debería ser la sociedad del momento en que se lanza esa noticia falsa (Bakir y McStay, 2018).

Establecen las *fake news* como una nueva categoría política, muy novedosa, y una nueva forma de campaña desacreditadora (Farkas y Schou, 2018). Definen lo *fake* como un discurso recurrente, que apoya agendas políticas particulares, y lo ejemplifican en tres momentos de Donald Trump (Farkas y Schou, 2018). Su objeto de estudio abarca publicaciones de redes sociales, artículos periodísticos y comentarios académicos, publicados entre noviembre de 2016 y marzo de 2017, en *The Washington Post*, *The Huffington Post*, The Guardian, *The Conversation*, CNN, *Monday Note*, *Business Insider*, *The New York Times*, *The Wall Street Journal*, *Buzzfeed News*, *Mashable*, *Slate*, Gizmodo y *Time Magazine*.

El problema añadido es que el algoritmo del *deepfake*, como se ha comentado, permite que cualquier usuario no profesional realice un vídeo donde un político diga algo que nunca dijo en la vida real. BuzzFeed, en 2018, alertó de los peligros de esta

técnica difundiendo un *deepfake* del expresidente norteamericano Barack Obama. Tras 56 horas de edición de vídeo, incluyeron a la perfección un nuevo audio, pronunciado por el actor Jordan Peele, que llamaba "dipshit" ("idiota" en castellano) a Donald Trump. El resultado fue perfecto y parecía que Obama había pronunciado todo lo que se escuchaba en el vídeo *fake*. Este experimento alertaba de los potenciales peligros que puede tener el *deepfake*.

El anterior ejemplo invita a pensar en el *deepfake* como la suma de los tres tiempos de las *fake news* (Farkas y Schou, 2018). El primer momento serían las noticias falsas como una crítica del capitalismo digital (Farkas y Schou, 2018), donde la estructura económica de Internet es la principal razón para la difusión de noticias falsas (Filloux, 2016; Silverman y Alexander, 2016). El segundo momento cubriría las noticias falsas como crítica de la política y los medios de comunicación conservadores. En las elecciones presidenciales norteamericanas de 2016, el término pasó de ser residual a uno de los términos más buscados en Google, justo antes del día de las elecciones, hasta alcanza un punto máximo histórico, en la toma de posesión de Donald Trump (Farkas y Schou, 2018). En este espacio de tiempo, las noticias falsas se asocian al ala derecha del espectro político estadounidense (Farkas y Schou, 2018). El tercer y último momento contendría las noticias falsas como crítica de los medios liberales y *mainstream*. Donald Trump no llevaba ni un día completo en su cargo, cuando declaró la guerra a los medios de comunicación, por atacar y deslegitimar su presidencia (Farkas y Schou, 2018). Después lo usaría para atacar a todos los medios tradicionales que parecieran oponentes directos, como la CNN (Farkas y Schou, 2018).

Aparte del potencial del *deepfake* como arma política de desinformación, propaganda o ataque al contrario, existe el problema de que es tan incipiente que no está legislado explícitamente en Europa, ni en España. Por sus fines, se podría castigar como un delito contra el derecho a la propia imagen, una injuria o un delito de odio. El primero, en España, es un derecho de la personalidad derivado de la dignidad humana, garantizado en el artículo 18 de la Constitución Española y tipificado en la Ley Orgánica 1/1982, de 5 de mayo, sobre protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen, también. En el *deepfake* no se difunden imágenes íntimas reales, pero sí creadas o figuradas, para que parezcan verosímiles, de la intimidad de sus protagonistas. En segundo lugar, la injuria está tipificada en la Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal. Por último, y en tercer lugar, los delitos de odio están contemplados en el marco jurídico europeo: Pacto Internacional de Derechos Civiles y Políticos de Naciones Unidas de 1966; Directiva 2012/29/UE del Parlamento Europeo y del Consejo, de 25 de octubre de 2012; Resolución del Parlamento Europeo, de 14 de marzo de 2013, sobre el refuerzo de la lucha contra el racismo, la xenofobia y los delitos motivados por el odio; y la Oficina para las Instituciones Democráticas y Derechos Humanos (ODHIR) de la Organización para la Seguridad y Cooperación en Europa (OSCE). También están recogidos en el Código Penal Español, en su artículo 510, y el Ministerio del Interior ha promovido distintas campañas mediáticas y digitales, para que la ciudadanía los conozca, los detecte y los denuncie.

Otros autores, en cambio, ponen la solución a todo lo *fake* en la deontología y en la propia responsabilidad de los medios digitales (Aparicio, 2015; Berlanga y Sánchez, 2018; Bernárdez y Serrano, 2018; Túñez y Martínez, 2018). Recomiendan que se preste mayor atención al papel de la publicidad digital en la causa y el combate,

tanto del fenómeno de las noticias falsas contemporáneas, como de la variante de las noticias falsas automatizadas (Catalina-García, García y Montes, 2015; Cáceres, Brändle y Ruiz, 2017; Foladori y García, 2017; Bakir y McStay, 2018). Sólo así los medios digitales podrán reencontrar un nuevo espacio de interacción riguroso con la ciudadanía (Said-Hung, Arcila-Calderón y Méndez-Barraza, 2011; Said-Hung et al., 2013; Govaert, Lagerwerf y Klemm, 2019; Soon y How Tan, 2016; Arroyo y Calle, 2018; Ireri et al., 2019; Plaut y Klein, 2019; Van Heekeren, 2019).

Haigh, Haigh y Kozak (2018) creen que el fin de lo *fake* está en la deontología, en la responsabilidad que tienen y deben ejercer los periodistas. Explican cómo un grupo de activistas voluntarios de Ucrania convirtió los hechos en un arma de contrapropaganda y citan StopFake y sus prácticas de trabajo para verificar las noticias publicadas en Internet (Haigh, Haigh y Kozak, 2018). Distinguen que StopFake evalúa las noticias para detectar signos de evidencia falsificada, como imágenes y citas manipuladas o tergiversadas, mientras que los sitios de verificación de hechos tradicionales evalúan afirmaciones políticas matizadas, pero asumen la exactitud de los informes (Haigh, Haigh y Kozak, 2018).

Stover (2018) entrevistó a Garlin Gilchrist II, Director Ejecutivo del nuevo Centro para la Responsabilidad de los Medios Sociales de la Universidad de Michigan, acerca de la responsabilidad de las *fake news*. El entrevistado definió las *deepfakes* como grabaciones de audio y vídeo, que han sido manipuladas digitalmente para convencer a la gente de que un político o una celebridad, por ejemplo, dijo algo que él o ella realmente no dijo o hizo algo que realmente no sucedió (Stover, 2018). Calificaba el *deepfake* como el apocalipsis de la información, porque impide a la gente distinguir entre real y falso.

La entrevista de Stover (2018) a Garlin Gilchrist II, Director del Center for Social Media Responsibility de la Universidad de Michigan, también observa las herramientas potenciales para acabar con la propagación de noticias falsas y reconstruir la confianza del público en fuentes confiables de noticias e información. Como solución, contempla un enfoque multidisciplinar, que reúna a expertos de Ciencias Sociales, Humanidades, Informática e Ingeniería, desde sectores públicos y privados (Stover, 2018). Según Gilchrist, el *fake* anula la capacidad de confiar en lo que vemos, leemos y escuchamos; es la pérdida de la fe en los sentidos humanos (Stover, 2018). Este trabajo esencialmente teórico, siguiendo ese espíritu de enfoque multidisciplinar, quiere inaugurar una parcela científica de alfabetización moral digital de la ciudadanía en *deepfakes*, para detectarlos y combatirlos; y que se queden solamente en la parcela del humor, para un mejor ejercicio ciudadano y cívico en el mundo digital.

2. Método

Este trabajo, que analiza teóricamente el fenómeno del *deepfake*, busca y comenta las herramientas para detectarlo y combatirlo, e inaugura una nueva vía de alfabetización moral digital de especial interés en el campo de la comunicación, por su importancia social y política. Como la bibliografía científica sobre el tema es dispar y aún incipiente en las Ciencias Sociales, la investigación se erige como una innovación teórica o prospectiva de investigación que aúna Inteligencia Artificial, *fake*

audiovisual o *deepfake*, alfabetización digital y predicción del fenómeno a corto y medio plazo. Para ello se emplea una metodología diacrónica en dos pasos.

En primer lugar, se realiza una investigación bibliográfica longitudinal, sobre el *deepfake*, poco explorado y reconocido en las publicaciones académicas del campo de las Ciencias Sociales y concretamente, de la Comunicación. Por su novedad, se ofrece una descripción histórica, sistémica y descriptiva. El fin fundamental de este método es compartir con la comunidad científica los rasgos más importantes del fenómeno para obtener las notas que resumen su naturaleza, funcionamiento y posibles usos, tal como se ha hecho en la Introducción.

En segundo lugar, se realiza una observación científica consciente de más de medio centenar de vídeos *deepfake*, para percibir directamente el fenómeno, conseguir una mayor generalización y garantizar la validez de los resultados, que se adentran en las técnicas computacionales u observacionales para detectar el *fake* audiovisual. A través de un método sintético, se relacionan esos vídeos, aparentemente aislados, con las propuestas de académicos y profesionales de varios campos de las Ciencias Sociales. El fin fundamental de este método es descifrar las posibles formas para detectar un *deepfake* o *fake* audiovisual, y conseguir esa alfabetización moral mediática de la ciudadanía, contra una nueva y perversa forma de falsedad.

3. Resultados

Tal como se indicaba en el Método, tras la revisión bibliográfica longitudinal, se ha realizado una observación científica consciente de más de medio centenar de vídeos *deepfake*, para percibir directamente el fenómeno, conseguir una mayor generalización y garantizar la validez de los resultados, que se adentran en las técnicas computacionales u observacionales para detectar el *fake* audiovisual.

A modo de resumen, se enumeran los 25 vídeos *deepfake* con más visualizaciones en YouTube, entre los 50 observados para este trabajo. Se ha considerado imprescindible reflejar 5 ítems: Título del vídeo; Protagonista del *deepfake*; Número de visualizaciones; Canal de YouTube; y Suscriptores del canal de YouTube. Los datos corresponden al momento del cierre de la investigación, tras una observación de 6 meses, en el segundo cuatrimestre de 2019:

Título del vídeo	**Protagonista del *deepfake***	**Número de visualizaciones**	**Canal de YouTube**	**Suscriptores del canal de YouTube**
Keanu Reeves Stops A ROBBERY!	Keanu Reeves	7.593.985	Corridor	6.357.165
Bill Hader impersonates Arnold Schwarzenegger [DeepFake]	Arnold Schwarzenegger	7.194.845	Ctrl Shift Face	141.448
Terminator 2 starring Sylvester Stallone [DeepFake]	Sylvester Stallone	2.706.783	Ctrl Shift Face	141.448
We Made The Best Deepfake on The Internet	Tom Cruise	1.811.165	Sam and Niko	2.187.517

The Dark Knight's Tale [Deep-Fake]	Heath Ledger	627.777	Ctrl Shift Face	141.448
Bruce Lee's Matrix [DeepFake]	Bruce Lee	534.950	Ctrl Shift Face	141.448
Tom Cruise Impersonator Olympics	Tom Cruise	495.683	San and Niko	2.187.652
Trump | Deepfakes Replacement	Donald Trump	484.729	derpfakes	18.945
The Shining starring Jim Carrey : Episode 3 - Here's Jimmy! [DeepFake]	Jim Carrey	422.455	Ctrl Shift Face	141.448
Mike Tyson and Snoop Dogg as Oprah and Gayle [Deepfake]	Mike Tyson y Snoop Dogg	386.225	DrFakenstein	7.148
Jennifer Lawrence-Buscemi on her favorite housewives [Deepfake]	Steve Bus-cemi	363.916	birbfakes	472
Superman and lois lane deep fake nicolas cage meme	Nicholas Cage	356.647	Life2Coding	5.840
Deepfakes | Christopher Reeve in Justice League	Christopher Reeve	238.278	derpfakes	18.945
Princess Leia Remastered... Again | Derpfakes	Carrie Fisher	180.307	derpfakes	18.945
Donald Trump | Mr. Bean Deepfake	Rowan Atkin-son	119.088	PLYGNfakes	95
Keanu Reeves in Forest Gump Deep Fake	Keanu Reeves	102.428	TheFakening	8.407
Keanu Reeves Sesame Street [Deepfake]	Keanu Reeves	85.275	DrFakenstein	7.148
Ross Marquand Is Everyone - 11-Deepfakes-In-One! [Deep-fake]	Ross Mar-quand	66.920	birbfakes	472
Ariana Grande Pete Davidson Deepfake	Pete Davidson	59.750	TheFakening	8.407
Keanu Reeves in Scarface | Deepfakes	Keanu Reeves	42.872	derpfakes	18.945
Deep Fake - Jennifer Lawrence - Steve Buscemi	Steve Bus-cemi	42.777	Truth Syrup	225
Marilyn Monroe Deepfake	Mira Sorvino	29.924	Deep Homage	626
Deep Fake Rami Freddie Ellen	Freddie Mer-cury	20.391	Wayne Davis	313
Nick Cage DeepFakes Movie Compilation	derpfakes	18.945	Usersub	653
Deepfakes | Nic Cage & Video Games	derpfakes	18.945	derpfakes	18.945

Tabla 1: Los 25 vídeos *deepfake* con más visualizaciones en YouTube al finalizar la investigación.

Esta investigación pretende servir de innovación teórica y por ello, la tabla anterior es solamente una prueba de la observación científica consciente que se ha llevado a cabo. Aunque en investigaciones posteriores se podrá ahondar en análisis cuantitativos, sí que se pueden deducir tres ideas principales, a partir de los datos de la tabla: los protagonistas de *deepfakes* son mayoritariamente varones; los creadores de *deepfakes*, en casi todos los casos, crean un perfil cuyo nombre se refiere directamente de la técnica; y un alto número de seguidores no asegura directamente la viralidad de un vídeo.

Tras ello, y de acuerdo a la metodología sintética, se relacionan esos cincuenta vídeos, aparentemente aislados (de distintos protagonistas y canales), con las propuestas de académicos y profesionales de las Ciencias Sociales. El fin fundamental es descifrar las formas para detectar un *deepfake* o *fake* audiovisual y conseguir esa alfabetización moral mediática de la ciudadanía contra una nueva y perversa forma de falsedad.

3.1. *Softwares* de detección de *deepfakes*

Lo que distingue los *deepfakes* de otras manipulaciones audiovisuales son unos resultados fotorrealistas. Con suficientes imágenes y tiempo de capacitación en computación, los vídeos resultantes pueden ser extremadamente convincentes. El canal *Dan it All* (YouTube, 2019) tiene publicado un ejemplo, en tiempo real, del procesamiento necesario para generar un *deepfake*: solamente 11 horas. En segundo lugar, hay que destacar la disponibilidad del software para realizarlos. Como se indicaba en la introducción, FakeApp es una interfaz creada por un usuario de Reedit que permite a casi cualquier personaje generar este tipo de vídeos desde sus hogares (Koopman, Macarulla y Geradts, 2018).

La combinación de resultados fotorrealistas y la facilidad de uso plantean un desafío para la detección de estos vídeos. Es un problema especialmente sensible y urgente en la era actual de las *fake news*, que va más allá de la aplicación de la ley y se vuelve relevante también para comunicadores, periodistas, sitios web de alojamiento de vídeos y usuarios de redes sociales (Koopman, Macarulla y Geradts, 2018). Debido a esto, cualquier método de detección o autentificación que sea accesible para la ciudadanía se convierte en algo fundamental y necesario.

La Universidad de Albany, en Estados Unidos, parece haber encontrado una grieta en los *deepfakes* para la ejecución de un software que permita identificar este tipo de vídeos manipulados (Ossorio, 2018). Un equipo de investigadores ha recurrido a la Inteligencia Artificial para analizar automáticamente vídeos, 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 superior al 95%. A veces, en los vídeos manipulados, existe un parpadeo con movimientos oculares que no imitan con exactitud los naturales, realizados por cualquier ser humano. También detectan extraños movimientos de cabeza, color de ojos incompatible con el de los seres humanos y en general, la explotación de todo tipo de señales fisiológicas extrañas. Los investigadores asumen que mientras que ellos tratan de configurar programas de detección de *deepfakes*, los creadores de este tipo de contenidos pueden mejorar la Inteligencia Artificial de los softwares que sirven para manipular vídeos (Li, Chang y Lyu, 2018; Li y Lyu, 2019).

Los investigadores del Information Sciences Institute (USC) de la Universidad del Sur de California aseguran haber descubierto un software con una precisión de hasta el 90%. El proceso de esta IA es simple. La máquina organiza todos los fotogramas uno encima del otro y luego, determina las diferencias entre estos para averiguar si el vídeo es falso. Cualquier imperfección biométrica determinará si el vídeo es una *deepfake* (Saeed, 2019). El objetivo es poder aplicar esta tecnología a una posible propagación de vídeos falsos durante la campaña electoral de 2020, en Estados Unidos. Y es que un vídeo producido por IA podría mostrar a Donald Trump diciendo o haciendo algo extremadamente indignante y calumnioso. Sería demasiado creíble y, en el peor de los casos, podría influir en una elección, desencadenar violencia en las calles o provocar un conflicto armado internacional (Knight, 2019). La investigación que se presentó en una conferencia de visión por computadora en California en 2019 fue financiada por Google y DARPA, un ala de investigación del Pentágono.

Aunque estas investigaciones y otras (Guera y Delp, 2018; Koopman, Macarulla y Geradts, 2018) han creado sus propios softwares para la detección de *deepfakes*, con una tasa de acierto que ronda el 90%, aún no son de dominio público y por lo tanto, los ciudadanos no tienen acceso a su libre y gratuito uso.

Si la detección de *deepfakes* audiovisuales está en sus orígenes, los softwares de localización de fotografías manipuladas tampoco son infalibles. Por ejemplo, *Fact-Check* de Facebook emplea el equipo humano de la red social y otros sistemas automatizados para verificar de datos y contenidos que puedan ser falsos, y así poder eliminarlos. Al principio, estas labores se limitaban a los textos y enlaces de medios de comunicación. Pero desde 2018, *Fact-Check* trabaja en el reconocimiento de vídeos y fotografías (Ticbeat, 2018). Por otro lado, *ImgOps* es un buscador que permite comparar una fotografía con otras similares en la red. Durante algunas manifestaciones, es normal que se hagan virales fotografías de conflictos sociales anteriores, como gente con el rostro ensangrentado. A través de esta plataforma, se puede reconocer si esa fotografía ha sido publicada en Internet con anterioridad y por lo tanto, no corresponde con el texto.

Google Chrome posee un complemento para navegadores, el *Fake News Detector*, que permite a los usuarios marcar una noticia, vídeo o fotografía com: *Legitimate, Fake news, Extremely bised, Satire* o *Not news*. *NewsCracker* es una tecnología de aprendizaje automático y análisis estadístico, que permite clasificar la calidad de cualquier artículo, contenido o publicación en una escala del 0 a 10 en función de su veracidad (Ticbeat, 2018). Sin embargo, a excepción de *ImgOps* (y otras que emplean un modelo de reconocimiento similar como *Google*), las demás aplicaciones o no están completamente desarrolladas o no son científicamente fiables.

A pesar de la cantidad de investigaciones centradas en la actualidad sobre la detección vídeos *deepfakes*, todavía no hay ninguna herramienta de Inteligencia Artificial que esté abierta a la ciudadanía para detectar estos contenidos manipulados. Esto obliga a buscar otras formas para averiguar la veracidad de un vídeo.

3.2. Técnicas de observación visual

Existen varias técnicas de observación visual para la detección de un vídeo falsificado con Inteligencia Artificial. El parpadeo de los ojos de un personaje *deepfake* es diferente al de una persona real. En condiciones normales, un ser humano tiende a

parpadear una vez cada 2-10 segundos y cada uno de sus parpadeos se prolonga durante 1-4 décimas de segundo (Ehrenkranz, 2018). Cualquier persona que no fuerce a propósito su parpadeo deberá encajar dentro de esos rangos durante cualquier grabación. Sin embargo, esto no ocurre en los *deepfake*, donde los rostros falsificados parpadean muchísimo menos que las personas reales (Merino, 2019). Y la razón es muy simple. En las fotografías que se publican en Internet, las personas no suelen aparecer con los ojos cerrados. Dado que los programas que generan *deepfakes* emplean este tipo de fotografías públicas para la realización de los vídeos, no hay, por ahora, forma de la Inteligencia Artificial pueda rellenar esos huecos al mismo ritmo de parpadeo que en un vídeo real (Lyu, 2018).

Además del parpadeo, que es uno de los indicios de observación visual más infalibles para la detección de *deepfakes*, hay otras vislumbres. Uno de los más obvios es el efecto de luces intermitentes en las caras manipuladas (Weber, 2018). En muchos *deepfakes*, los rostros se ven extraños y las transiciones entre la cabeza y el cuello o el cabello no siempre encajan bien. Algo parecido al parpadeo de los ojos ocurre con el interior de la boca. Según Galiana (2019), el software para crear *deepfakes* es capaz de transferir caras bastante bien, aunque carece del dominio de los detalles. Por ejemplo, un cierto desenfoque en el interior de la boca es otra señal de que podría ser una imagen falsa. Asimismo, los clips de un vídeo falsificado no suelen ser muy largos: entre 30 segundos y 1 minuto, por la complejidad de su realización (Vicent, 2019).

Aunque ya hay varios softwares que detectan fotografías manipuladas, en la actualidad, no hay softwares de acceso abierto para la detección de *deepfakes*. Sin embargo, es algo en lo que están trabajando varias compañías privadas e universidades. La plataforma Truepic ya creó una aplicación para detectar fotografías falsas que fue empleada en el conflicto sirio por la cadena Al Jazeera (*El País*, 2019). El objetivo de Truepic es hacer lo mismo ahora con los *deepfake* audiovisuales.

4. Discusión

Como se indicaba en las primeras líneas, la detección de vídeos *fake* o *deepfakes* es uno de los retos de la ciudadanía del siglo XXI. Por el momento, los usuarios que crean estos contenidos lo han hecho principalmente con tres fines: humor (Beamonte, 2018), venganza (Askham, 2018) o pornografía (Cole, 2017). Sin embargo, su potencialidad negativa como culmen del *fake* político, sobre todo en periodo electoral, hacía necesario reflexionar sobre el fenómeno, su naturaleza y características. Gracias a que la bibliografía es incipiente y casi testimonial en Ciencias Sociales, este trabajo, además de cumplir con ese fin social y cívico, ha propuesto la inauguración de posibles vías teóricas que ahonden en técnicas e ideas para que los ciudadanos puedan detectar los *deepfakes*, sobre todo, antes de la próxima vez que acudan a las urnas.

Castañares asegura (2007a y 2007b) que las motivaciones de aquellos que ven televisión no pueden descartarse en el hecho de buscar una satisfacción de ver su imagen reflejada en televisión. Siguiendo esta idea llegamos a la conclusión de que este tipo de vídeos *fake* representan, según la teoría de la televisión moralista de Castañares (2007a), un espejo en el que se miran los ciudadanos del siglo XXI. Por un lado, ansían ver a las celebridades más mediáticas en situaciones íntimas, como

son los vídeos pornográficos, a pesar de que esto constituya un delito contra los derechos de imagen y la intimidad de estas mujeres. Asimismo, quieren ver a los políticos y los actores famosos en situaciones extravagantes y/o ridículas. Ambos casos comparten el mismo trasfondo educativo: "Si eres famoso vamos a hacer lo que queramos con tu imagen y no vas a poder defenderte". Acorde a los estudios de Castañares, Núñez y González (1996) y Castañares (2007a), este sería el trasfondo educativo del contenido *deepfake* en la ciudadanía contemporánea y, al mismo tiempo, un espejo de sí misma.

Tras seguir escrupulosamente la metodología propuesta, los resultados teóricos se constituyen en dos posibles soluciones a la detección del *fake* audiovisual: 1) la creación y publicación de un software gratuito, que todavía no existe con carácter libre, para que cualquier persona, desde sus dispositivos electrónicos, pueda dilucidar si un vídeo que ha recibido está manipulado; y 2) la alfabetización moral ciudadana y también profesional, pensando en los comunicadores, para difundir y divulgar los detalles que el ojo humano puede ver en un vídeo *deepfake* para detectarlo. Al final, el cuerpo humano, como máquina compleja e inteligente que es, dentro de los umbrales de su normalidad y consciencia, 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. Las pruebas del delito del *fake* audiovisual son entonces tres: parpadeo irregular o casi inexistente de los ojos, desenfoque en el cuello o en el interior de la boca del protagonista, y efecto de luces intermitentes en las caras manipuladas.

Tras conocer estos detalles, se abren prospectivas de investigación teórica y práctica. En la parte de teoría, futuras investigaciones podrían relacionar el *deepfake* político con la tradición investigadora anterior, sobre manipulación fotográfica y marketing electoral, roles políticos, técnicas de telegenia, o censura de colaboradores políticos cuyos rastro histórico se quiere borrar. El *deepfake* se erige como la evolución necesaria y computacional de la manipulación fotográfica, sobre todo aquella que tiene el objetivo de desprestigiar al contrario.

En la parte de práctica, la tabla de los 25 vídeos con más visualizaciones en YouTube, dentro de los 50 vídeos observados para hacer este trabajo, invita también a futuras reflexiones anhelando más comprensión de fenómenos sociales en los contextos digitales actuales. Ha dejado tres ideas importantes: los protagonistas de *deepfakes* son mayoritariamente varones; los creadores de *deepfakes*, en casi todos los casos, crean un perfil cuyo nombre se refiere directamente de la técnica; y un alto número de seguidores no asegura directamente la viralidad de un vídeo. Con ello, surgen interrogantes novedosos: ¿por qué los protagonistas son varones?, ¿hay *deepfakes* de mujeres políticas?, ¿la representación es diferente o desigual?, ¿los usuarios que mantienen esas cuentas son profesionales o aficionados?, ¿ofrecen otros vídeos?, ¿cómo son las interacciones y el *engagement* que generan?, ¿qué características comparten los vídeos con más visualizaciones?, ¿por qué algunos vídeos con pocas visualizaciones han aparecido en medios de comunicación masivos?, ¿contempla YouTube el *deepfake* en sus normas comunitarias?, ¿necesitará ponerle límites a corto o medio plazo? Afortunadamente, el número de interrogantes avala el carácter novedoso de este trabajo, que pretendía una innovación teórica original y necesaria a partir de la inspiración que sugiere la obra de Wenceslao Castañares y otros teóricos. Como el Profesor manifestó siempre: "Queda la esperanza de que otros puedan ir más allá de lo que yo he ido".

Fuentes de financiación

Este trabajo es parte de la investigación llevada a cabo bajo el proyecto titulado *Produsage cultural en las redes sociales: industria, consumo popular y alfabetización audiovisual de la juventud española*. Referencia FEM2017-83302-C3-3-P. Proyecto I+D del Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia 2018-2022, Ministerio de Economía, Industria y Competitividad.

Referencias Bibliográficas

Abril Curto, G. (2010). Cultura visual y espacio público-político. *CIC. Cuadernos De Información Y Comunicación*, *15*, 21-36. Recuperado de https://revistas.ucm.es/index.php/CIYC/article/view/CIYC1010110021A

Abril Hernández, A. (2014). Narrativas transmediáticas en entornos digitales: la novela hipermedia Inanimate Alice y sus aplicaciones docentes. *CIC. Cuadernos De Información Y Comunicación*, *19*, 287-301. https://doi.org/10.5209/rev_CIYC.2014.v19.43916

Adorjan, M. y Ricciardelli, R. (2019). Student perspectives towards school responses to cyber-risk and safety: the presumption of the prudent digital citizen. *Learning, Media and Technology*, *44*(4), 430-442. https://doi.org/10.1080/17439884.2019.1583671

Aladro Vico, E. (2007). Metáforas e iconos para transmitir información. *CIC. Cuadernos De Información Y Comunicación*, *12*, 49-57. Recuperado de https://revistas.ucm.es/index.php/CIYC/article/view/CIYC0707110049A

Aladro Vico, E. (2017). El lenguaje digital, una gramática generativa. *CIC. Cuadernos De Información Y Comunicación*, *22*, 79-94. https://doi.org/10.5209/CIYC.55968

Aladro Vico, E. Jivkova-Semova, D y Bailey, O. (2018). Artivismo: Un nuevo lenguaje educativo para la acción social transformadora. *Comunicar: Revista científica iberoamericana de comunicación y educación*. 57, 9-18. https://doi.org/10.3916/C57-2018-01

Aparicio, E. (2015). Los medios de comunicación en la violencia contra las mujeres: el paradigma de la delgadez. *Historia y comunicación social*, *20*(1), 107-119. https://doi.org/10.5209/rev_HICS.2015.v20.n1.49550

Arroyo, I. y Calle, S. (2018). Los community managers de las ONGD. Estudio de percepciones y usos de las redes sociales. *Icono 14*, *16*(2), 121-142. https://doi.org/10.7195/ri14.v16i2.1189

Ashton, S., McDonald, K. y Kirkman, M. (2019). What does 'pornography' mean in the digital age? Revisiting a definition for social science researchers. *Porn Studies*, *6*(2), 144-168. https://doi.org/10.1080/23268743.2018.1544096

Askham, G. (2018, 3 de mayo). Qué son los deepfakes y por qué se están convirtiendo en el nuevo porno de la venganza. *BBC News*. Recuperado de https://www.bbc.com/mundo/noticias-43975322

Bakir, V. y McStay, A. (2018). Fake News and The Economy of Emotions. Problems, causes, solutions. *Digital Journalism*, *6*(2), 154-175. https://doi.org/10.1080/21670811.2017.1345645

Beamonte, P. (2018, 25 de enero). FakeApp, el programa de moda para crear vídeos porno falsos con IA. *Hipertextual*. Recuperado de https://hipertextual.com/2018/01/fakeapp-videos-porno-falsos-ia

Berlanga, I. y Sánchez, M. (2018). Ética y tratamiento de la información en los relatos periodísticos de corrupción. *Historia y comunicación social*, *23*(2), 477-488. https://doi.org/10.5209/HICS.62269

Bernárdez, A. y Padilla, G. (2018). Mujeres cineastas y mujeres representadas en el cine comercial español (2001-2016). *Revista Latina de Comunicación Social*, *73*, 1247-1266. DOI: 10.4185/RLCS-2018-1305

Bernárdez, A. y Serrano, M. (2018). Lo personal es político: un bebé en la sesión de constitución de las Cortes Generales. El tratamiento televisivo del caso de Carolina Bescansa y su hijo. Vivat Academia. Revista de Comunicación, 142, 79-96. http://doi.org/10.15178/va.2018.142.79-96

Cáceres, M. D., Brändle, G. y Ruiz San Román, J. A. (2017). Sociabilidad virtual: la interacción social en el ecosistema digital. *Historia y comunicación social*, *22*(1), 233-247. https://doi.org/10.5209/HICS.55910

Castañares, W. (1994). *De la interpretación a la lectura*. Madrid. Iberediciones.

Castañares, W. (1995). Géneros realistas en televisión: los reality shows. *CIC. Cuadernos De Información Y Comunicación*, *1*, 79-91. Recuperado de https://revistas.ucm.es/index.php/CIYC/article/view/CIYC9595110079A

Castañares, W., Núñez, F. y González, J.L. (1996). *La sensibilidad moral: impresiones y testimonios de nuestro tiempo*. Madrid. Nóesis.

Castañares, W. (1997). La Televisión y sus géneros: ¿una teoría imposible? *CIC. Cuadernos De Información Y Comunicación*, *3*, 167-180. Recuperado de https://revistas.ucm.es/index.php/CIYC/article/view/CIYC9797110167A

Castañares, W. (2007a). *La televisión moralista: valores y sentimientos en el discurso televisivo*. Madrid. Fragua.

Castañares, W. (2007b). Cultura visual y crisis de la experiencia. *CIC. Cuadernos De Información Y Comunicación*, *12*, 29-48. Recuperado de https://revistas.ucm.es/index.php/CIYC/article/view/CIYC0707110029A

Castañares, W. (2011). Realidad virtual, mímesis y simulación. *CIC. Cuadernos De Información Y Comunicación*, *16*, 59-81. https://doi.org/10.5209/rev_CIYC.2011.v16.3

Catalina-García, B., García, A. y Montes, M. (2015). Jóvenes y consumo de noticias a través de Internet y los medios sociales. *Historia y comunicación social*, *20*(2), 603-621. https://doi.org/10.5209/rev_HICS.2015.v20.n2.51402

Cerdán Martínez, V. y Villa, D. (2019). Creación de un formato de concienciación social para la televisión pública española «Héroes Invisibles». *Revista Latina de Comunicación Social*, *74*(13), 1565-1579. DOI: 10.4185/RLCS-2019-1399-82

Cerdán Martínez, V. y Padilla Castillo, G. (2019). Historia del «fake» audiovisual: «deepfake» y la mujer en un imaginario falsificado y perverso. *Historia Y Comunicación Social*, *24*(2), 505-520. https://doi.org/10.5209/hics.66293

Cole, S. (2017, 11 de diciembre). AI-Assisted Fake Porn Is Here and We're All Fucked. *Vice*. Recuperado de https://motherboard.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn

Ehrenkranz, M. (2018, 16 de junio). Hay un truco infalible para detectar si un vídeo ha sido manipulado por una IA Deep Fake: fíjate en los ojos. *Gizmodo*. Recuperado de https://bit.ly/2Y7iuRd

El País (2018, 15 de noviembre). Estas aplicaciones te ayudan a detectar imágenes y vídeos falsos desde tu móvil. Recuperado de https://bit.ly/2zl1CwU

Fake News Detector (2018). Recuperado de https://bit.ly/2Y86oY5

Farkas, J. y Schou, J. (2018). Fake News as a Floating Signifier: Hegemony, Antagonism

and the Politics of Falsehood. *Javnost - The Public. Journal of the European Institute for Communication and Culture*, *25*(3), 298-314. https://doi.org/10.1080/13183222.2018.1463047

Filloux, F. (2016, 5 de diciembre). Facebook's Walled Wonderland Is Inherently Incompatible with News. *Monday Note*. Recuperado de https://mondaynote.com/facebooks-walled-wonderland-is-inherently-incompatible-with-news-media-b145e2d0078c#.v0txzx82e

Foladori, G. y García, M. (2017). El papel de la experiencia histórica y la confianza en la comunicación de tecnologías emergentes: el caso de las nanotecnologías. *Historia y comunicación social*, *22*(1), 221-230. https://doi.org/10.5209/HICS.55909

Fouce, H., Pecourt, J., & Pedro-Carañana, J. (2018). El conocimiento en la universidad como problema público. Condiciones de producción en la comunicación y las ciencias sociales. *CIC. Cuadernos De Información Y Comunicación*, 23, 9-23. https://doi.org/10.5209/CIYC.60909

Galiana, P. (2019, 10 de mayo). ¿Qué son los Deepfakes y cómo detectarlos? IEBS. Recuperado de https://bit.ly/2wcpXmA

Govaert, C., Lagerwerf, L. y Klemm, C. (2019). Deceptive Journalism: Characteristics of Untrustworthy News Items. *Journalism Practice*. https://doi.org/10.1080/17512786.2019.1637768

Guera, D. y Delp, E. (2018). Deepfake Video Detection Using Recurrent Neural Networks. *Conference: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. DOI: 1-6. 10.1109/AVSS.2018.8639163.

Haigh, M., Haigh, T. y Kozak, N. I. (2018). Stopping Fake News. The work practices of peer-to-peer counter propaganda. *Journalism Studies*, *19*(4), 2062-2087. https://doi.org/10.1080/1461670X.2017.1316681

Ireri, K., Chege, N., Kibarabara, J. y Onyalla, D. B. (2019). Frame Analysis: Newspaper Coverage of Kenya's Oil Exploration in the Post-2012 Discovery Era. *African Journalism Studies*, *40*(2), 34-50. https://doi.org/10.1080/23743670.2019.1635035

Jivkova, D.; Requeijo, P. y Padilla, G. (2017). Usos y tendencias de Twitter en la campaña a elecciones generales españolas del 20D de 2015: hashtags que fueron *trending topic*". *El profesional de la información*, *26*(5), 824-837. https://doi.org/10.3145/epi.2017.sep.05

Knight, W. (2019, 21 de junio). A new deepfake detection tool should keep world leaders safe—for now. *Technology review*. Recuperado de https://bit.ly/2M3kqYA

Koopman, M., Macarulla Rodriguez, A. M. y Geradts, Z. (2018). Detection of Deepfake Video Manipulation. *Proceedings of the 20th Irish Machine Vision and Image Processing conference IMVIP 2018*, 133-136. Recuperado de https://www.researchgate.net/publication/329814168_Detection_of_Deepfake_Video_Manipulation.

Li, Y., Chang, M. C. y Lyu, S (2018). In ictu oculi: Exposing ai generated fake face videos by detecting eye blinking. *IEEE International Workshop on Information Forensics and Security (WIFS)*. Recuperado de https://bit.ly/2S0oppv

Li, Y. y Siwei, L. (2018). *Exposing DeepFake Videos By Detecting Face Warping Artifacts*. New York: Computer Vision Foundation.

Linvill, D. L. (2019). Addressing social media dangers within and beyond the college campus. *Communication Education*, *68*(3), 371-380. https://doi.org/10.1080/03634523.2019.1607885

Lucić, M., Baćak, V. y Štulhofer, A. (2019). The role of peer networks in adolescent pornography use and sexting in Croatia. *Journal of Children and Media*, *14*(1), 110-127. https://doi.org/10.1080/17482798.2019.1637356

Lyu, S. (2018, 31 de agosto). The best defense against deepfake AI might be… blinking. *Fast Company*. Recuperado de https://bit.ly/2N78iqv

Maeve, D. (2013, 3 de julio). 6% of Online Adults are reddit Users. *Pew Research Center*. Recuperado de https://www.pewinternet.org/2013/07/03/6-of-online-adults-are-reddit-users

Márquez, I. (2017). Del muro a la pantalla: el graffiti en la cibercultura. *CIC. Cuadernos De Información Y Comunicación*, 22, 95-106. https://doi.org/10.5209/CIYC.55969

Márquez, I. y Tosca, S. (2017). Playing with the city: street art and videogames. *Arte, Individuo y Sociedad*, 29(1), 105-120. http://dx.doi.org/10.5209/ARIS.51892

Márquez, I. y Ardèvol, E. (2018). Hegemonía y contrahegemonía en el fenómeno *youtuber*. *Desacatos: Revista de Ciencias Sociales*, 56, 34-49.

Matsakis, L. (2018, 14 de febrero). Artificial intelligence is now fighting fake porn. *Wired*. Recuperado de https://www.wired.com/story/gfycat-artificial-intelligence-deepfakes/

Merino, M. (2019, 3 de febrero). Así es posible saber si un vídeo es un *deepfake* con sólo un abrir y cerrar de ojos, literalmente. *Xataka*. Recuperado de https://bit.ly/2BlE4J3

Ossorio, M. A. (2018, 30 de agosto). DARPA ha encontrado la forma de detectar Deepfakes: la clave está en los ojos. *Media Tics*. Recuperado de https://bit.ly/2Lzfrj2

Padilla Castillo, G. y Presol Herrero, África. (2020). Ética y deontología en publicidad. Nike 'Dream Crazier' 2019 como campaña feminista en Instagram. *Comunicación y Género*, *3*(1), 3-15. https://doi.org/10.5209/cgen.63975

Pardo, L. (2018, 26 de enero). FakeApp: 200 fotos tuyas son suficientes para incluir tu rostro en cualquier vídeo. *NeoTeo.* Recuperado de https://www.neoteo.com/fakeapp-200-fotos-tuyas-suficientes-incluir-rostro-cualquier-video/

Plaut, S. y Klein, P. (2019). "Fixing" the Journalist-Fixer Relationship: A Critical Look Towards Developing Best Practices in Global Reporting. *Journalism Studies*, *20*(12), 1696-1713. https://doi.org/10.1080/1461670X.2019.1638292

Rice, L. L. y Moffett, K. W. (2019). Snapchat and civic engagement among college students. *Journal of Information Technology & Politics*, *16*(2), 87-104. https://doi.org/10.1080/19331681.2019.1574249

Saeed, A. (2019, 25 de junio). Deep Fake Detection Software Can Work as Game Changer for Authenticity. *Digital information world*. Recuperado de https://bit.ly/2Y875k9

Said-Hung, E., Arcila-Calderón, C. y Méndez-Barraza, J. (2011). Desarrollo de los cibermedios en Colombia. *El profesional de la información*, 2011, *20*(1), 47-53. DOI: 10.3145/epi.2011.ene.06

Said-Hung, E., Serrano-Tellería, A., García de Torres, E., Yezers'ka, L. y Calderín, M. S. (2013). La gestión de los Social Media en los medios informativos iberoamericanos. *Communication & Society*, *26*(1), 67-92. Recuperado de https://dadun.unav.edu/bitstream/10171/35436/1/20130425132044.pdf

Setty, E. (2019). 'Confident' and 'hot' or 'desperate' and 'cowardly'? Meanings of young men's sexting practices in youth sexting culture. *Journal of Youth Studies*. https://doi.org/10.1080/13676261.2019.1635681

Silverman, C. y Alexander, L. (2016, 4 de noviembre). How Teens in the Balkans are Duping Trump Supporters with Fake News. *Buzzfeed News*. Recuperado de https://www.buzzfeed.com/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo?utm_term=.ebrpdWyymW#.ijbg9Q44YQ

Soon, C. y How Tan, T. (2017). The media freedom-credibility paradox. *Media Asia*, 43(3-4), 176-190. https://doi.org/10.1080/01296612.2016.1276315

Stover, D. (2018). Garlin Gilchrist: Fighting fake news and the information apocalypse. *Bulletin of the Atomic Scientists*, *74*(4), 283-288. https://doi.org/10.1080/00963402.2018.1486618

Tic Beat (2018, 14 de octubre). 5 herramientas para ayudarnos a detectar 'fake news'. *Tic Beat*. Recuperado de https://bit.ly/2THd6Go

Túñez, M. y Martínez, Y. (2018). Impacto de las editoriales y las revistas "depredadoras" en el área de Comunicación. *Historia y comunicación social*, *23*(2), 439-458. https://doi.org/10.5209/HICS.62267

Van Heekeren, M. (2019). The Curative Effect of Social Media on Fake News: A Historical Re-evaluation. *Journalism Studies*, *21*(3), 306-318. https://doi.org/10.1080/1461670X.2019.1642136

Vincent, J. (2019, 27 de junio). Deepfake detection algorithms will never be enough. Spotting fakes is just the start of a much bigger battle. *The Verge*. Recuperado de https://bit.ly/2Xidj0o

Weber, K. (2018, 8 de noviembre). 8 steps to verify Deep Fake videos. *Medium Corporation*. Recuperado de https://bit.ly/32HxFEd

YouTube (2019). Creating deepfakes live. Enter to win your own deepfake video! *Dan It All*. Recuperado de https://bit.ly/2y1ppkj