# Una nueva taxonomía del uso de la imagen en la conformación interesada del relato digital. *Deep fakes* e inteligencia artificial

## A new taxonomy for image use in the intentional shaping of the digital narrative: deep fakes and artificial intelligence

**Ángel Gómez-de-Ágreda; Claudio Feijóo; Idoia-Ana Salazar-García**

Cómo citar este artículo:

**Gómez-de-Ágreda, Ángel**; **Feijóo, Claudio**; **Salazar-García, Idoia-Ana** (2021). "Una nueva taxonomía del uso de la imagen en la conformación interesada del relato digital. *Deep fakes* e inteligencia artificial". *Profesional de la información*, v. 30, n. 2, e300216.

*https://doi.org/10.3145/epi.2021.mar.16*

Artículo recibido el 29-04-2020
Aceptación definitiva: 11-07-2020

**Ángel Gómez-de-Ágreda** ✉
*https://orcid.org/0000-0003-1036-6324*

*Universidad Politécnica de Madrid*
*Ministerio de Defensa*
Paseo de la Castellana, 109
28071 Madrid, España
*angel@angelgomezdeagreda.es*

**Claudio Feijóo**
*https://orcid.org/0000-0002-9499-7790*

*Universidad Politécnica de Madrid*
*Centro de Apoyo a la Innovación Tecnológica* (*CAIT*)
Parque Científico y Tecnológico UPM
Campus de Montegancedo, s/n.
28223 Pozuelo de Alarcón (Madrid), España
*claudio.feijoo@upm.es*

**Idoia-Ana Salazar-García**
*https://orcid.org/0000-0002-9540-8740*

*OdiseIA*
*Universidad San Pablo CEU*
*Facultad de Humanidades y Ciencias de la Comunicación*
Paseo de Juan XXIII, 3. 28040 Madrid, España
*idoiaana.salazargarcia@ceu.es*

**Resumen**

Cualquier confrontación pretende la imposición, siquiera parcial, de los criterios y la voluntad propios sobre los del adversario. En los últimos años, las tecnologías digitales y la ciencia de datos se han combinado para favorecer la aparición de nuevas formas de control del discurso y para establecer potentes campañas de desinformación que constituyen un nuevo tipo de conflicto con características digitales. Precisamente, a partir de la bibliografía disponible y del trabajo directo de los autores en diversos grupos de expertos, este artículo estudia el uso de las imágenes –reales, alteradas o generadas artificialmente por completo– estáticas o en movimiento como uno de los medios más eficientes para alterar las percepciones y, por lo tanto, los relatos dentro del denominado ámbito cognitivo. Para ello, el artículo recoge de forma ordenada y exhaustiva la más reciente doctrina del entorno militar y de inteligencia relativa a las llamadas "operaciones de influencia", desde una doble perspectiva tecnológica y sociológica. A partir de aquí, determina las diferencias entre las técnicas de desinformación tradicionales y las que incorporan tecnologías digitales y de inteligencia artificial sobre imágenes. El artículo propone una nueva taxonomía de doble entrada que refleje el grado de manipulación de las imágenes y el objetivo que se persigue con la desinformación. Este puede ser un medio para identificar y priorizar los casos más relevantes y, de aquí, adoptar las correspondientes contramedidas. Éstas se examinan en detalle para concluir que sólo una combinación de transparencia, formación del consumidor, tecnología y legislación puede contrarrestar el creciente uso de imágenes con contenido falseado.

**Palabras clave**

Operaciones de influencia; *Fake news*; *Deep fakes*; Conformación del relato; Narrativa; Desinformación; Periodismo; Manipulación; Inteligencia artificial; Tecnología; Imagen; Vídeo; Conflicto bélico; Ámbito cognitivo.

**Abstract**

Any confrontation seeks the (partial) imposition of one party's will on an adversary. In recent years, digital technologies and data science have combined to create new ways of controlling the narrative and carrying out powerful information or disinformation campaigns that are part of a new type of warfare with digital characteristics. In particular, based on a literature review and the direct work of authors in different expert groups, this paper studies the use of either static or moving images (real, altered, or wholly artificially generated) as one of the most efficient means of altering perceptions and thereby narratives in the so-called cognitive domain. To this aim, this article collects in an orderly and exhaustive way the most recent military and intelligence doctrine related to such "influence operations," from a dual technological–sociological perspective. Based on this, the paper determines the differences between traditional disinformation techniques and those incorporating digital and artificial intelligence technologies in the form of images and video. The paper proposes a new double-entry taxonomy that can display the degree of image faking and the objective of disinformation. This helps to identify and prioritize the most relevant cases and thus adopt the most appropriate countermeasures. These are also examined in detail, leading to the conclusion that only a combination of transparency, consumer education, technology, and regulation can counteract the increasing use of images and video with false content.

**Keywords**

Influence operations; Fake news; Deep fakes; Narrative shaping; Disinformation; Journalism; Manipulation; Artificial intelligence; Technology; Image; Video; Warlike conflict; Cognitive domain.

> "Mientras que la mayoría de los políticos y académicos entienden qué hacen las armas nucleares y los tanques, las posibilidades, trampas y procesos de las misiones de hackeo les resultan comparativamente opacas" (**Buchanan**, 2020).

## 1. Contexto y antecedentes

La pandemia de la Covid-19 ha resaltado la importancia de los medios digitales en la transmisión de las percepciones, la creación de narrativas y la relevancia de disponer de una información veraz. Privados durante el confinamiento de la experiencia directa de lo que no fuera el mundo más inmediato al domicilio, la verdadera ventana al exterior se abría en las pantallas de ordenadores y teléfonos móviles.

Ante la falta de evidencias científicas, rumores y bulos más o menos interesados suplían el apetito del público por datos con los que aliviar el miedo y la incertidumbre (*EU vs Disinformation*, 2020). Y lo hacían aprovechando el inmenso caudal que permiten las redes sociales, la inmediatez del medio digital y la ubicuidad de "fuentes" que supone la interactividad e hiperconectividad de los dispositivos personales de miles de millones de internautas. La posibilidad de integrar elementos textuales y audiovisuales en un mismo documento –lo que Jean Cloutier denominó *l´audioscriptovisuel*–

> "abre la posibilidad a un lenguaje polisintético e integrador. Polisintético porque aglutina diversos elementos que toman sentido sólo dentro del conjunto. Es un lenguaje sincrético y proporciona acceso a las tres dimensiones espaciales y a la temporal" (**Cloutier**, 1994).

Este flujo constante de contenidos mezcla, sin posibilidad práctica de distinción en tiempo útil, editorial y noticia. Contenidos procedentes de fuentes solventes y contrastadas, las sátiras y contenidos humorísticos, y las piezas elaboradas se mezclan con intención de confundir al receptor. Todo ello aprovechando que nuestra receptividad se agudiza en momentos de gran vulnerabilidad en los que sufrimos problemas sistémicos o cuando experimentamos un shock (**Klein**, 2012) "que destruye lo cotidiano y nos llena de incertidumbre" (**Peirano**, 2020).

Aunque las circunstancias fueran excepcionales, el uso interesado de la información y la desinformación no son fenómenos asociados únicamente a ellas. Ni tampoco son nuevos. Como no lo es la utilización de imágenes genuinas o manipuladas para la transmisión o refuerzo de dicha desinformación. Sin embargo, la rapidez de los cambios de los medios tradicionales a los digitales sin que haya habido un relevo en el marco de la veracidad ha provocado desequilibrios. La autoridad de la palabra impresa en un medio público, como el periódico, era sinónimo de veracidad incuestionable, tanto más si venía acompañada de un documento gráfico que constituía la evidencia, la prueba última de la verdad.

La novedad llega, por un lado, de las aportaciones de la tecnología –en concreto la inteligencia artificial (IA) y, en breve, la realidad virtual (RV) y aumentada (RA)– en la generación o manipulación de dichas imágenes. Y, por otro lado, de la superficialidad de la atención y el juicio de una población sobresaturada de datos, y de cambios en la doctrina militar que incorpora la influencia entre su arsenal de estrategias híbridas [las que mezclan el empleo de armamento convencional con acciones políticas, cibernéticas y de influencia, así como de guerra irregular (terrorismo, guerrillas…)].

> Las falsas noticias, especialmente en forma de imágenes, tienen repercusiones en el terreno de lo personal y en lo corporativo, pero pueden afectar también a la seguridad nacional dado su valor histórico como garantía de veracidad

Precisamente en el presente trabajo se aborda el papel de las nuevas tecnologías digitales en las estrategias de empleo de herramientas –o armas– gráficas de desinformación. En especial, se tratará el papel que pueden tener los *deep-fakes*, las manipulaciones más sofisticadas en imágenes estáticas y videos, en la alteración de las percepciones y de las narrativas. Para esto, después de esta sección de contexto, el siguiente apartado explica el procedimiento de trabajo que se ha seguido. A partir de aquí se revisa el estado del arte y los conceptos clave sobre desinformación y sus efectos en el ámbito cognitivo. Sobre estas bases, los apartados centrales del artículo examinan el uso de las imágenes para la desinformación proponiendo una nueva taxonomía, así como una revisión de los medios actuales para contrarrestarla. Un apartado de conclusiones cierra el artículo.

Todo ello se enmarca ya dentro de los incipientes desarrollos doctrinales en las Fuerzas Armadas relativos al ámbito cognitivo y a las operaciones de influencia, que afectan al conjunto de la población. La estabilidad estratégica, interna y externa, es altamente dependiente de las percepciones públicas. Éstas están cada vez más configuradas desde medios digitales susceptibles de ser manipulados por procesos sofisticados de inteligencia artificial (**Lin**, 2018) que permiten modificar imágenes de maneras antes prácticamente imposibles.

El artículo utiliza una revisión de artículos científicos, libros e informes que abordan la desinformación, su impacto en el ámbito cognitivo y el uso de imágenes en los medios digitales.

En la parte científica, se ha seguido una revisión sistematizada de la bibliografía existente con especial énfasis en Ciencias Humanas y Sociales (**Codina**, 2018). Para ello se han utilizado buscadores de artículos científicos[1] siguiendo la metodología que describen **Piasecki**, **Waligora** y **Dranseika** (1948), un procedimiento ya estándar (**De-Granda-Orive** *et al.*, 2013). Para la parte correspondiente a libros e informes, los autores, dada su posición, se encuentran en condiciones de acceder a la mayor parte de la bibliografía relevante que se produce en el tema que trata este artículo y, en cualquier caso, no existe una forma unificada de acceder a esta "literatura gris". Dentro de este segmento, se ha prestado particular atención al creciente cuerpo doctrinal sobre operaciones de influencia en las fuerzas armadas de los principales países[2].

Para completar el análisis, se han considerado asimismo las herramientas tecnológicas actualmente disponibles para los interesados en la manipulación de imágenes, así como el estado del arte y su evolución futura. Se han tenido en cuenta, por ejemplo, estudios sobre la evolución reciente de la tecnología como el de **Nguyen** *et al.* (2019). También se han considerado las últimas novedades relativas a *GPT-3*, el modelo de reconocimiento de patrones de datos y su replicación coherente más avanzado actualmente que, a pesar de estar diseñado principalmente para la manipulación de textos, también puede emplearse de forma satisfactoria con imágenes (**Chen** *et al.*, 2020) o sistemas para la prevención de *deep-fakes* como, por ejemplo, *Fawkes*:
*http://sandlab.cs.uchicago.edu/fawkes/#code*

Con todo ello, el trabajo propone una taxonomía novedosa sobre el uso de los softwares de manipulación de imágenes para la generación de desinformación o la configuración del relato que trasciende las existentes hasta ahora. **Molina** *et al.* (2019) categorizan siete tipos de contenido falso –desde las noticias falsas hasta el periodismo ciudadano– encuadrados en cuatro dominios: mensaje, fuente, estructura y red. Una aproximación similar siguen **Kapantai** *et al.* (2020), también centrada en tipos y dominios. Por su parte **Valchanov** (2018) resulta original en su tipología al incluir en la clasificación la incompetencia o la falta de exhaustividad de los periodistas.

Sin embargo, en el presente documento se pretende centrar el foco en el uso de las imágenes concretamente e incorporar a los actores que llevan a cabo estas actividades y las finalidades que persiguen, así como el nivel de desarrollo tecnológico que se precisa en cada caso. Los autores entienden que, más allá de las técnicas empleadas en cada caso concreto, lo verdaderamente relevante son los efectos que se persiguen –y eventualmente consiguen– con la manipulación.

## 2. Desinformación con imágenes en medios digitales

En un entorno mediático interrelacionado como el actual, los consumidores siguen las historias a través de muchas fuentes en un proceso *transmedia* (**Aguado**; **Feijóo**; **Martínez-Martínez**, 2013). Los propios ciudadanos, gracias a la potente combinación de redes sociales y tecnologías digitales, pueden expandir los contenidos periodísticos por los medios tradicionales o directamente crear nuevos relatos informativos. Esta potenciación de la libertad de información no está exenta de riesgos.

> "Las *fake news* o noticias falsas se han convertido en una constante dentro del periodismo y en un verdadero problema a la hora de discernir la verisimilitud de una información" (**Alonso-González**, 2019).

Los bulos no son un fenómeno nuevo dentro del periodismo (**Allport**; **Postman**, 1947), sin embargo, internet y las redes sociales facilitan exponencialmente su difusión debido a la velocidad, la amplitud y la universalidad que ofrecen esos canales. La parte visual es un componente que tradicionalmente ha añadido verosimilitud a una información, pero que ahora también debe someterse a revisión (**Hamd-Alla**, 2007).

> La desinformación se enmarca en las estrategias híbridas que no emplean necesariamente medios militares convencionales ni requieren de un estado declarado de hostilidades para ser empleadas en el denominado ámbito cognitivo

De hecho, el periodismo ha de verificar la información que comparte, especialmente cuando la probabilidad de que circulen mentiras se incrementa. Se ha llegado al punto en el que especificar el grado de confianza de las fuentes, y verificar la información y combatir la viralización de las falsas noticias es casi tan importante como el hecho de informar en sí mismo (**López-Borrull**; **Vives-Gràcia**; **Badell**, 2018). Los propios usuarios de la información son parte del problema, ya que las nuevas tecnologías favorecen un proceso comunicativo bidireccional mediante el cual los usuarios pueden convertirse en el origen de cualquier noticia sin pasar ningún tipo de filtro que indique si la información que se comparte es veraz.

Utilizando el ejemplo de la pandemia de la Covid-19, la desinformación ha tenido un crecimiento exponencial aprovechando la situación de incertidumbre y la sobreexposición a la Red que se produjo durante el confinamiento (**Marqués**, 2020). El estudio calcula en un 200% los ataques de ingeniería social y en un 300% el incremento del número de falsas noticias. Se trata de falsas noticias, no "noticias falsas", ya que la falsedad se encuentra en el hecho de que sean noticia y no necesariamente en su contenido (**Gómez-de-Ágreda**, 2018), relatos que, en un creciente número de casos, se limitan a imitar formatos periodísticos (**Bennett**; **Livingston**, 2018).

De forma simultánea proliferaron las verificaciones de hechos (*fact-checks*) con un crecimiento del 900% en el caso de las noticias en lengua inglesa tan sólo en el primer trimestre de 2020 (**Brennen** *et al.*, 2020). Las limitaciones en cuanto a medios o intenciones, o las intrínsecas del modelo actual de *fact-checking* arrojaban resultados dispares en función de la implicación y las políticas de cada red. Así, **Brennen** *et al.* (2020) estiman en un 59% el número de falsas noticias que no son retiradas de la red de *Twitter*, en un 27% en *YouTube* y en un 24% en *Facebook*. A la vista de que pueden existir dudas sobre si un contenido ha sido alterado mediante técnicas de *deep-fake*, **Gerardi**, **Walters** y **James** (2020) sugieren que aquel que resulte identificado como falso, pero no reúna los requisitos para ser eliminado sea, al menos, marcado como "alterado" para alertar a la audiencia.

Formalmente, desinformación es información equívoca que tiene la función de engañar, información diseñada para generar falsas creencias, o bien, información equívoca no accidental (**Fallis**, 2015). A pesar de existir desde siempre, el término no aparece en los diccionarios en inglés hasta los años 80 del siglo XX. El nombre parece derivar de la técnica precedente de los servicios rusos de inteligencia cuyo objetivo era la diseminación de información falsa para engañar a la opinión pública. Es una técnica subversiva, no de captación de información (**Mahariras**; **Dvilyanski**, 2018). Por su parte, el término *fake-news* –utilizado inicialmente para describir publicaciones satíricas como *The daily show* o *The onion*– ha pasado a tener connotaciones políticas tendentes a desacreditar las opiniones o la credibilidad de los adversarios (**Nielsen**; **Graves**, 2017; **Tandoc**; **Lim**; **Ling**, 2017).

La desinformación supone una amenaza para la capacidad de las audiencias para construir su propio relato en función de percepciones no sesgadas. Siguiendo a Alonso-González se puede decir que

> "el conocimiento de la realidad es lo que nos permite a los receptores formarnos opinión sobre el mundo que nos rodea, sin embargo, el flujo permanente de información y las noticias falsas erosionan esa credibilidad" (**Alonso-González**, 2019).

Sin embargo, no implica la existencia de falta de coherencia interna del discurso. De hecho, normalmente, la desinformación bien diseñada basa su fortaleza en mantener una coherencia interna (**Arendt**, 2006) –si bien normalmente forzada– igual o mayor a la percibida respecto de la realidad (**Maddock**, 2020). La audiencia demanda un marco conceptual en el que se sienta segura y, a ser posible, cómoda. Las generaciones más digitales tienden a construir su propia realidad paralela, muy basada en el audiovisual –sea real, ficticio o virtual–, antes que a enfrentar los problemas que encuentran en la que habitan (**Castells**, 2005).

El fenómeno de la desinformación va mucho más allá de la transmisión de información falsa y de los engaños puntuales que se puedan producir. Pretender combatirlo identificando y desmontando las mentiras una a una, amén de laborioso y poco práctico, deja sin acometer el verdadero problema de la desconfianza y pérdida de referencias que persigue (**Bennett**; **Livingston**, 2018).

En una suerte de neolengua orwelliana, la coherencia se mantiene en el relato y en el lenguaje como continente con independencia de la concordancia que exista respecto de la realidad del contenido. Se manipula la percepción y el mensaje sin necesidad de afectar a aquello de lo que se habla (**Austin**, 1956). Según **Howard** (2020), durante la campaña electoral estadounidense de 2016 se compartían en *Twitter* tantas noticias falsas como verídicas. Este sería el escenario más desfavorable posible para la audiencia ya que la probabilidad de encontrar una mentira sería la misma que la de una verdad. De esta circunstancia ya alertaba la empresa de consultoría *Gartner* en 2019 cuando afirmaba que, en el plazo de dos años,

> "el público occidental consumirá más noticias falsas que verdaderas" (**Flores-Vivar**, 2019).

En el caso de la información audiovisual, la proliferación de tecnologías capaces de alterar las percepciones de modo convincente degrada la confianza en las evidencias basadas en este tipo de información (**Barnes**; **Barraclough**, 2019). Siendo la definición de verdad una convención cultural, política y social, la capacidad para generar nuevas evidencias, sobre todo en el caso audiovisual, dejaría la verdad en manos de aquellos con la posibilidad de crear percepciones que sirvieran para construirla (**Paris**; **Donovan**, 2019). Es decir, las relaciones de poder se establecerían en función de la capacidad para alterar las percepciones más que las realidades.

El estado actual de la desinformación no refleja únicamente la aplicación de la tecnología digital a las tácticas clásicas. La capacidad de personalización de los mensajes basados en millones de datos minados en redes sociales sobre la víctima y su entorno, y la universalización de la audiencia suponen un importante cambio cualitativo (**Freedland**, 2020). A ello hay que añadir un objetivo de negocio, ya que el advenimiento de la web 2.0 posibilita la monetización de los afectos por parte de las redes sociales y las plataformas (**Aguado**, 2020).

Las tendencias actuales en el uso de la tecnología para facilitar la desinformación se centran así en:

- la minería de información de las redes sociales (**Manfredi-Sánchez**, 2021) para la elaboración de ataques de *spear phishing* –ataques de ingeniería social en los que se personaliza el "anzuelo" en función del conocimiento profundo que se tiene sobre el objetivo–;
- la manipulación de imágenes y videos (*deep-fakes*);
- el uso de RV y RA para intensificar el engaño (**Miller**, 2020).

Estas acciones corren a cargo de grupos de expertos al servicio de clientes con intenciones criminales. **Watts** (2019) denomina a estos grupos "manipuladores persistentes avanzados" (*advanced persistent manipulators*, *APM*, por similitud con los *advanced persistent threats*, *APT*, del mundo ciber), y los define como

> "un actor o combinación de actores que perpetran un ataque de información amplio y sofisticado, multimedia y multiplataforma sobre un objetivo específico".

Los objetivos de un *APM* serían: ejercer influencia sobre la audiencia, desacreditar a un adversario, provocar conflictos, reclutar partidarios, o distorsionar la realidad.

No obstante, la desinformación no suele generar brechas en el discurso del adversario, sino que aprovecha las existentes para llenar los huecos y ampliarlas (**Mahariras**; **Dvilyanski**, 2018). Es el vacío informativo y la falta de transparencia lo que propicia la aparición de un discurso alternativo que amortigüe el vértigo de la incertidumbre (**Calvo-Albero** *et al.*, 2020). Identificada la línea de fractura, el siguiente paso es la obtención o fabricación del contenido, su difusión y amplificación por diversos canales hasta conseguir la implicación activa de la audiencia en participaciones frecuentemente surgidas de la indignación a favor o en contra del mensaje. Cuando el relato pasa a ocupar el espacio informativo o de debate, el objetivo está conseguido; un objetivo que suele consistir más en crear discordia y confusión que en convencer al público (**Mahariras**; **Dvilyanski**, 2018). Ya en 1984, un desertor de la *KGB* de alto rango, Yuri Bezmenov, describía en este video

*https://www.youtube.com/watch?v=bX3EZCVj2XA*

las operaciones de desinformación de la agencia de inteligencia soviética de este modo:

> "Lo que básicamente significa es cambiar la percepción de la realidad (…) hasta el punto de que, a pesar de la abundancia de información, nadie pueda llegar a conclusiones sensatas en interés de la defensa de sí mismo, sus familias, sus comunidades o su país".

Es de esperar que la influencia conseguida se ejerza en las distintas fases del conflicto por ambos bandos. A través de una polarización de las posturas o una exacerbación de los sentimientos, la manipulación afectaría a la gestión de crisis, la escalada, la capacidad de disuasión e, incluso, a la toma de decisiones en materia de guerra nuclear (**Davis**, 2019). Se han publicado ya numerosas alteraciones de imágenes y videos con la finalidad de:

- fomentar un discurso de odio, como en:
  *https://www.youtube.com/watch?v=NGIE1HrBQJI*
- sembrar dudas sobre la catadura moral de personajes públicos, como el famoso video de Barak Obama en:
  *https://www.youtube.com/watch?v=cQ54GDm1eL0*
- para hacer revisionismo histórico sobre el que construir una narrativa:
  *https://www.youtube.com/watch?v=FzOVqClci_s*

La primacía de lo emocional frente a lo racional concede un papel protagonista a lo visual, a la imagen, frente al discurso elaborado a través del texto (**Hameleers** *et al.*, 2020). Al mismo tiempo, la economía de la atención en la que muchas empresas de nuevas tecnologías basan su negocio propicia la rápida obsolescencia de las noticias, rebajando así las exigencias de calidad de la manipulación (salvo cuando se desarrollan herramientas automatizadas de verificación) (**Lorenz-Spreen** *et al.*, 2019).

En esta construcción de la desinformación audiovisual se busca una alteración altamente convincente de las percepciones. La mentira debe tener un porcentaje de verdad, con una parte comprobable del mensaje, para resultar más creíble.

**Grijelmo** (2017) propone el uso de técnicas como la insinuación, la presuposición y el sobreentendido, la falta de contexto, la inversión de la relevancia y la poscensura. **Bjola** y **Pamment** (2019), por su parte, afirman que las técnicas de producción de desinformación son:

- la polarización
- la invocación de emociones
- la distribución de teorías de la conspiración
- el "troleo"
- el desvío de culpa
- la asunción de identidades falsas.

No se trata ya solamente del uso directo de la VR o la AR para la generación de imágenes falseadas en dos o tres dimensiones, sino también de su uso en la generación de realidades paralelas (**Rettberg**, 2014). Una parte de la sociedad tiende a refugiarse en estos mundos alternativos creados artificialmente a caballo entre el mundo real y el virtual. En algunos casos, se han detectado vínculos entre estas realidades construidas y las más novedosas teorías de la conspiración (como QAnon) (**Warzel**, 2020).

El umbral de entrada en el negocio de la distribución de noticias –verídicas o falsas– se ha ido haciendo cada vez más accesible, una afirmación particularmente importante para el caso de las imágenes, antes sólo al alcance de actores que dispusieran de la tecnología y los recursos apropiados (**Paris**; **Donovan**, 2019). La reputación e imagen de marca de los medios, por el tiempo que se tarda en construir, es la única barrera real que sigue existiendo en la era digital (**Hameleers** *et al.*, 2020). No obstante, la posibilidad de falsificar medios consagrados introduciendo desinformación entre sus contenidos digitales destapa un nuevo frente en el combate contra las falsas noticias. El caso reciente de medios occidentales utilizados por servicios extranjeros para publicar digitalmente contenidos sesgados contra la *OTAN* ilustra esta posibilidad (ver *BBC*, 2020).

Tan relevante es este nuevo campo que al menos 70 países han establecido equipos de desinformación en redes sociales formados bien por personas, bien por *bots* automatizados (**Howard**, 2020). En algunos casos el objetivo son terceros países mientras que en otros se trata de la población propia o ambos.

## 3. Desinformación visual en el ámbito cognitivo

"El corazón del hombre necesita creer algo, y cree mentiras
cuando no encuentra verdades que creer"
Mariano José de Larra

En los conflictos híbridos que caracterizan lo que llevamos de siglo XXI, se cree que la utilización de la fuerza convencional se limitará a ocasiones muy puntuales. Las agresiones tendrán lugar más frecuentemente en el mundo financiero o el digital. Más insidiosa todavía será la manipulación de la opinión pública a través de la desinformación, especialmente con el uso de técnicas digitales sofisticadas basadas en imágenes como las *deep-fakes* (**Wheeler**, 2018).

La opinión pública se encuentra hoy sometida a una incidencia muy alta de informaciones de dudosa fiabilidad que, indirectamente, modelan su criterio ante los distintos acontecimientos que la rodean. Las redes sociales acrecientan este fenómeno y contribuyen a la distribución de las *fake-news* con una rapidez hasta 70 veces mayor que aquellas de veracidad probada. Especialmente las relacionadas con contenidos de política, terrorismo, desastres naturales, ciencia, leyendas urbanas o información financiera (**Vosoughi**; **Roy**; **Aral**, 2018).

Dado el hecho de que el periodista dejó, hace tiempo, de ser el único en ofrecer esta influencia, el profesional de la información tiene ahora un reto importante a la hora de dar credibilidad a su trabajo y centrar el debate de la opinión pública en base a los criterios de veracidad y contraste de información que han guiado desde siempre el trabajo periodístico de calidad (**Rodríguez-Fernández**, 2019). En el mundo audiovisual, en el que la imagen transmite una mayor sensación de adecuación de la realidad, la utilización de técnicas avanzadas de inteligencia artificial, como las citadas *deep-fakes*, puede contribuir a la desinformación. Sin embargo, también pueden ayudar a mejorar la calidad del trabajo periodístico y de las consiguientes producciones informativas (**Manfredi-Sánchez**; **Ufarte-Ruiz**, 2020). Así, se reafirma el hecho de que el problema no está en la tecnología en sí misma, sino en la forma en la que se usa.

Estas nuevas técnicas, junto con la compartición continuada de datos personales, y la situación contextual histórica, están fomentando un cambio en la definición clásica de guerra, en la que se implica el uso de la violencia en un grado determinado de muerte y destrucción física. La capacidad para influir en las decisiones de los adversarios se ha fiado desde la antigüedad en la fuerza de las armas. Sin embargo, también desde los primeros tratados bélicos, se recomendaba reservar esta opción como último recurso. Tanto Sun-Tzu, el estratega chino, como Kautilya, el consejero político indio, recogen en sus tratados –*El arte de la guerra* (**Sun-Tzu**, 2013) y el *Artashastra* (**Kauti-**

> La introducción de la inteligencia artificial para la generación de *deep-fakes*, así como la generalización del uso de la realidad virtual y la aumentada, supone un salto cualitativo respecto de las formas convencionales de desinformación

**lya**, 2016)– la conveniencia de no confiar exclusivamente en la fuerza bruta y la recomendación de evitar, en la medida de lo posible, la destrucción del adversario.

El engaño es el arte supremo para el primero, para el segundo lo es la generación de desasosiego entre la población rival, la debilitación de su voluntad para que se someta a la propia. No se está hablando, por consiguiente, tanto de una medida puramente militar como de una acción política dirigida y coordinada desde las más altas instancias del Estado, un concepto que sigue presente en las estrategias y manuales militares en la actualidad (**Metz**; **Johnson**, 2001). La responsabilidad de la decisión está, por lo tanto, en el nivel político con independencia de a quién corresponda su ejecución concreta (**Calvo-Albero** *et al.*, 2020).

Por ello no es de extrañar que el concepto del ámbito cognitivo en las fuerzas armadas de todo el mundo está actualmente en pleno desarrollo doctrinal aunque los textos doctrinales de algunos países no están disponibles públicamente. Entre otros países y servicios hay publicaciones de:

- Australia (**Holbrook**, 2018; **Bienvenue**; **Rogers**; **Troath**, 2019);
- República Popular China (**Kania**, 2020) y *http://www.81.cn/jfjbmap/content/2019-10/22/content_245810.htm*
- Reino Unido (*UK Ministry of Defence*, 2017);
- Estados Unidos, tanto a nivel conjunto en el *Pentágono* (**Wright**, 2019), como del *Army* (**Schmidt**, 2020) o el *Air Force* (**Nettis**, 2020).

En España, el concepto de ámbito cognitivo está desarrollado a falta de sanción oficial. Se definirá como el espacio no físico de las operaciones que, dentro del entorno de la información, abarca las acciones, procesos y los efectos relativos a las percepciones del ser humano, considerado éste de forma individual, organizada o en sociedad. Es consustancial a la capacidad de juicio y toma de decisiones de las personas, y a su conciencia de grupo. Para Defensa, alcanza al sistema de creencias, emociones y la motivación con el fin de modificar la conducta manifiesta de las personas afectadas por el conflicto y repercute en el resto de los ámbitos.

También alcanza a los sistemas técnicos afectados por el conflicto que pudieran ser o llegar a ser susceptibles de imitar –aun parcialmente– o condicionar los procesos intelectuales humanos, es decir, a la percepción y procesado de datos de las máquinas. Atiende esencialmente a la gestión conceptual y estratégica de la información, que permita obtener el grado de influencia deseado en la audiencia objetivo (**Calvo-Albero** *et al.*, 2020).

El ámbito cognitivo se une de este modo al terrestre, naval, aéreo, espacial y cibernético entre aquellos que se consideran en las modernas operaciones bélicas (**Paul** *et al.*, 2018). Al igual que las operaciones en todos los demás, pretende afectar las decisiones humanas. Sin embargo, se diferencia de los otros en que no aspira a hacerlo alterando la realidad, sino las percepciones y las emociones. La característica peculiar del ámbito cognitivo es que no es posible actuar directamente sobre el mismo, sino que se afecta a su naturaleza desde el resto de los ámbitos operacionales. Es decir, las operaciones físicas o cibernéticas tienen un impacto sobre sus propios entornos, pero al mismo tiempo repercuten en las percepciones y emociones influyendo sobre la voluntad de la audiencia objetivo (**Peco-Yeste**, 2020).

La manipulación de las percepciones aprovecha los sesgos individuales y grupales del ser humano, tanto aquellos que están presentes en sus procesos de decisión como personas, como los que afecten a sistemas algorítmicos de toma de decisión automatizados o autónomos en los cuales se puedan haber volcado estos mismos prejuicios a través del entrenamiento o de su programación. Como seres fundamentalmente visuales que somos, son las percepciones que provienen de las imágenes las que tienen un potencial de manipulación más inmediato (**Moran**, 2005).

Para Ortega y Gasset, nuestro comportamiento viene condicionado más por la interpretación de las percepciones que tenemos que por los sucesos reales (**Santos-Porras**, 2020). La manipulación de las percepciones permitiría, por lo tanto, escapar a la tiranía que la realidad impone sobre el yo y la circunstancia que lo rodea. Se trata de una búsqueda de la felicidad basada en la experimentación emocional (de ahí la "mejora de la experiencia de cliente") más que en una reflexión racional (**Ruiz**, citado en **Santos-Porras**, 2020). El internauta entra en el juego de la manipulación para ser capaz de construir su propia realidad virtual, más si contiene imágenes, para evitar enfrentarse a la física en una suerte de drogodependencia de percepciones que no desafíen su forma de entender el mundo o su proyecto vital.

La forma práctica de llevar a cabo la manipulación es operar en el espacio informativo, sea desde el establecimiento activo de las narrativas o desde el desafío a éstas mediante la presentación de discursos alternativos. En esta labor, la desinformación es una herramienta básica empleada tanto por los agentes que controlan el relato desde una posición dominante en los medios como por aquellos otros que aspiran a rebatirlo desde plataformas distribuidas.

En resumen, las operaciones en el ámbito cognitivo están orientadas a obtener ventajas en el terreno de lo emocional, al ser a este último al que apela la desinformación. Se puede afirmar que el centro de gravedad de las operaciones es –y ha sido siempre– la voluntad del combatiente y que, por lo tanto, es sobre ella sobre la

> "Las medidas de protección contra imágenes de contenidos falseado comienzan con la discreción respecto de las propias imágenes online"

que tienen que incidir los efectos de las acciones (**Gómez-de-Ágreda**, 2019). Esta voluntad, y la de los dirigentes o la opinión pública, es particularmente sensible a los mensajes audiovisuales con alta carga emotiva (**Arsenault**, 2020).

Las imágenes con contenido falseado se distribuyen, principalmente, a través de redes sociales y aplicaciones de mensajería, y webs de medios marginales o de propaganda política. La diseminación de falsas noticias facilita la captación de la atención propia de las redes sociales en base a su capacidad para resultar sorprendente y novedoso (**Fard**; **Lingeswaran**, 2020). La contaminación de un número suficientemente significativo de fuentes genera desconfianza en el conjunto del ecosistema informativo. A ello contribuye la falta de criterio en la selección de dichas fuentes, esa confusión que existe entre editorial y noticia. Dos tercios de los estadounidenses se informan total o parcialmente a través de unas redes sociales que han hecho bandera de su irresponsabilidad respecto a los contenidos que publican (**Mahariras**; **Dvilyanski**, 2018).

Un estudio de la *Harvard University* sobre imágenes en los grupos públicos de la red *WhatsApp* en India identificó como falsas alrededor de un 13% de las imágenes del servicio de mensajería frente a un 10% de prevalencia general. Más de la tercera parte (34%) de las mismas eran meras descontextualizaciones de la imagen (relativamente fáciles de identificar automáticamente mediante *image hashing*, que consiste en convertir cada imagen en una secuencia numérica que la identifica unívocamente facilitando su comparación posterior), un 30% resultaron ser memes satíricos y un 10% imágenes retocadas mediante lo que se podrían denominar técnicas clásicas (**Garimella**; **Eckles**, 2020).

Su mayor difusión se produce cuando se combinan las distintas plataformas de forma que unas aporten la credibilidad, otras la viralización, y la combinación de todas ellas la ubicuidad necesaria para camuflar sus orígenes. La existencia de cámaras fotográficas digitales en cada bolsillo y de las redes sociales proporcionan los medios para hacerlo (**Paris**; **Donovan**, 2019). La relación que se establece entre los medios tradicionales y las redes sociales, que muchas veces se percibe como una competición por la audiencia, termina por ser necesariamente simbiótica. Alberto Artero, director de *El confidencial*, afirma que el "cuarto poder" y lo que Mark Zukerberg denominó "quinto poder" en su comparecencia pública ante el Senado de los Estados Unidos, se complementan en las respectivas funciones de elaboración y distribución de contenidos, por mucho que sus puntos de partida estén muy diferenciados (**Fard**; **Lingeswaran**, 2020). Artero afirmó en su participación en el Seminario "Huella digital" que

> "el 'quinto poder' no reconoce los valores fundamentales de los profesionales de la comunicación".
> *https://www.fpablovi.org/index.php/sintesis-huella-digital/998-cuestiones-eticas-en-el-tratamiento-y-gestion-de-datos-en-los-medios-de-comunicacion*

## 4. Taxonomía

Algunas taxonomías generales relativas a falsas noticias, como la que se recoge en (**Molina** *et al.*, 2019), las divide en:

- noticias falseadas
- contenido polarizado
- sátira
- informes falsos
- comentarios
- información persuasiva;
- periodismo ciudadano simulado.

Esta tipología estaría insertada en cuatro dominios: el mensaje, la fuente, la estructura y la red.

En cuanto a imágenes con contenido falseado, la clasificación más habitual considera tres grandes grupos: descontextualización, memes y manipulación (**Garimella**; **Eckles**, 2020). Sin embargo, deja fuera las categorías más sutiles de manipulación, que son también las más conflictivas y de mayor interés. Por este motivo, el artículo propone una nueva clasificación más completa y que incluye actores e intenciones, como se discute a continuación.

A partir de las ideas recogidas por **Garimella** y **Eckles** (2020) y **Paris** y **Donovan** (2019), que sugieren una graduación desde *deep-fakes* hasta descontextualización, se propone en primer lugar una ampliación de las clasificaciones existentes que recoja los siguientes tipos:

- imágenes descontextualizadas, es decir utilización de dobles o atribución de la acción a un momento o lugar diferente del original; como denuncia *https://www.elespanol.com/reportajes/20171006/252225238_0.html*
- imágenes parciales;
- imágenes/videos retocados, incluyendo alteraciones en la velocidad;
- imágenes/videos alterados digitalmente;
- *deep-fakes*;
- imágenes/videos generados digitalmente *ex novo*, que se pueden utilizar en la generación de perfiles falsos en redes sociales (**Povolny**; **Chick**, 2020; **Hartman**; **Satter**, 2020).

La clasificación propuesta tiene como elementos destacados los siguientes. En primer lugar, se mantiene la descontextualización como una categoría per sé, pero se distingue de las imágenes parciales que, manteniendo el contexto, presentan una visión interesadamente parcial de la realidad que induce una interpretación errónea de lo mostrado .

Un caso famoso de distorsión de la realidad se realizó sobre una foto de *The Associated Press*, disponible en los siguientes urls:
*https://networkedthought.substack.com/p/reusable-media-fact-checking*
*https://www.demilked.com/media-manipulating-truth*
*https://steemitimages.com/640x0/http://consciousreporter.com/wp-content/uploads/2015/05/media-manupilate.jpg*
*https://digmedia.lucdh.nl/wp-content/uploads/2019/11/73258819_2399417883517500_4203145005495222272_n-2.jpg*

Sobre la fotografía original (centro) se pueden ofrecer dos relatos antagónicos mostrando solamente una parte de la realidad: a la izquierda parece que un soldado va a disprar su arma sobre otro y a la derecha otro soldado da a beber con una cantimplora.

Este tipo de imágenes se ha usado profusamente, por ejemplo, para mostrar supuestas aglomeraciones de gente durante los primeros momentos de la pandemia de Covid-19 o en actos de campaña electoral, como se describe en
*https://www.demilked.com/media-manipulating-truth*

En segundo lugar, se distingue también entre retoques y alteraciones digitales del contenido de los videos o fotografías
*https://english.elpais.com/elpais/2017/10/06/inenglish/1507278297_702753.html*

Con ello se pretende diferenciar entre técnicas como la modificación de la velocidad de reproducción de un video o la utilización de programas informáticos para alterar la composición de las imágenes. Es muy conocido el video que muestra a la presidenta del Congreso de los Estados Unidos, Nancy Pelosi, en aparente estado de embriaguez. La reducción de la velocidad de reproducción consigue el efecto sin afectar a los píxeles concretos del mismo
*https://www.theguardian.com/us-news/video/2019/may/24/real-v-fake-debunking-the-drunk-nancy-pelosi-footage-video*

La forma de identificar la manipulación e, incluso, el tratamiento legal que podría recibir es diferente en ambos casos. Es similar el caso del juicio a los policías en el caso de Rodney King. Los agentes fueron absueltos tras la exhibición del video del apaleamiento a una fracción de su velocidad normal en el que las reacciones de King a los golpes se pueden interpretar como intentos de incorporarse para seguir resistiendo (**Paris**; **Donovan**, 2019).

Cuando estos programas suponen la utilización de redes neuronales convolucionales (CNN) (**Krizhevsky**; **Sutskever**; **Hinton**, 2017) o generativas [lo que se denomina *synthetic media* (**Barnes**; **Barraclough**, 2019)] para generar alteraciones profundas (*deep-fakes*), entramos en una categoría que merece una diferenciación en función de su sofisticación y potencial disruptivo. Finalmente, estos mismos sistemas son capaces de generar, *ex novo*, imágenes realistas partiendo de bases de datos de entrenamiento como es el caso de
*https://thispersondoesnotexist.com*

Algunas de estas categorías pueden parecer relativamente inocentes si se aplican a objetivos concretos. Sin embargo, la prevalencia de alteraciones de videos o imágenes que contienen rostros humanos no debe hacer olvidar que estas mismas técnicas se pueden aplicar sobre multitud de soportes y contenidos: personas, documentos de texto o gráficas, paisajes u obras de arte (por la capacidad de la inteligencia artificial para replicar pinturas de autores clásicos siguiendo exactamente las pautas del original), y mapas y planos.

Otro tipo de clasificación es el que resulta de la distinción entre los diferentes propósitos o intenciones del manipulador. Un resumen de los casos presentes en la bibliografía hecha por los autores arroja los siguientes tipos:

- Sátira: se incluye en la clasificación el contenido satírico por su potencial para provocar equívocos siguiendo a **Klein** y **Wueller** (2017) a pesar de opiniones en contra como **Allcott** y **Gentzkow** (2017) que argumentan que no existe intencionalidad de engañar;
- Publicidad/propaganda;
- Desinformación;
- Manipulación/construcción del relato.

Estas dos clasificaciones se pueden combinar para dar lugar a un mapa de cuadrantes que identifica los objetivos de las imágenes falseadas, los actores principales involucrados y el nivel tecnológico necesario en la actualidad para llevar a cabo el proceso de creación de imágenes con contenido falseado. Se presenta en la figura 2 y se discute a continuación.

Los cuadrantes generados por esta clasificación crean cuatro tipos principales de escenarios de uso. Los dos superiores se caracterizan por el uso necesario de tecnologías avanzadas de IA –no disponibles para cualquiera–, mientras que los dos inferiores solo requieren un conocimiento experto y técnicas ampliamente disponibles:

- en el cuadrante superior derecho es donde suceden las *operaciones de influencia* propiamente dichas, en las que están interesados típicamente agentes estatales. Se trata de imágenes *ex novo* o *deep fakes* que buscan la desinformación y la manipulación;
- en el cuadrante superior izquierdo, que podemos denominar como de *marketing*, tienen interés típicamente grandes corporaciones o partidos políticos que buscan propaganda o desprestigiar al rival;

- en el cuadrante inferior derecho, donde se busca prioritariamente la creación de un *conflicto*, se sitúan grupos de interés que no tienen necesariamente acceso a la última tecnología y que utilizan imágenes retocadas, parciales o descontextualizadas para influenciar la percepción de la audiencia;
- en el cuadrante inferior izquierdo se encuentra el *prosumer*, productor y agente de la *economía de la atención* que genera y contribuye a viralizar imágenes con contenido falseado como parte del modelo de negocio de las redes sociales.

Los dos cuadrantes situados a la derecha son especialmente problemáticos. La combinación de técnicas digitales avanzadas sobre imágenes con propósitos criminales o bélicos resulta particularmente inquietante. Los tempos políticos, por ejemplo, son muy distintos de los tecnológicos o los jurídicos por lo que la atribución de la autoría de un montaje o la constatación misma de la manipulación de un recurso puede resultar demasiado lenta para evitar efectos indeseables sobre las percepciones y las decisiones que provoquen. Esta misma circunstancia se da en los ataques cibernéticos.

Es fácil imaginar un video alterado a través de medios "sintéticos" en el que se modifique o se construya directamente una realidad que afecte a las relaciones diplomáticas de dos o más países. Podría pensarse en un caso similar al del video que mostraba al Emir de Qatar expresando afirmaciones contrarias a sus socios del *Consejo de Cooperación del Golfo* y que resultó ser el detonante de la ruptura de relaciones entre varios de ellos. La presión de la opinión pública podría llevar a la adopción de medidas antes de que pudiese demostrarse fehacientemente la autenticidad del video.

Hay que notar que la descripción en cuadrantes es una aproximación a la complejidad de la cuestión y que idénticas técnicas pueden dar lugar a cualquiera de los tipos de imágenes con contenido falseado expuestos. La transición entre un tipo y otro resulta, en ocasiones, difícil de establecer y puede no ser lineal en el sentido de seguir el orden propuesto. Así, un *face swap*, una sustitución del rostro de un personaje por otro, puede formar parte de un meme satírico perfectamente lícito y legal o –aplicado, por ejemplo, a un video de contenido pornográfico– puede suponer una manipulación delictiva que atenta contra el honor de las personas y su reputación social o política. Del mismo modo, puede aplicarse a una campaña de marketing o a una de desinformación en la que se pretenda presentar una realidad distinta de la verdad.

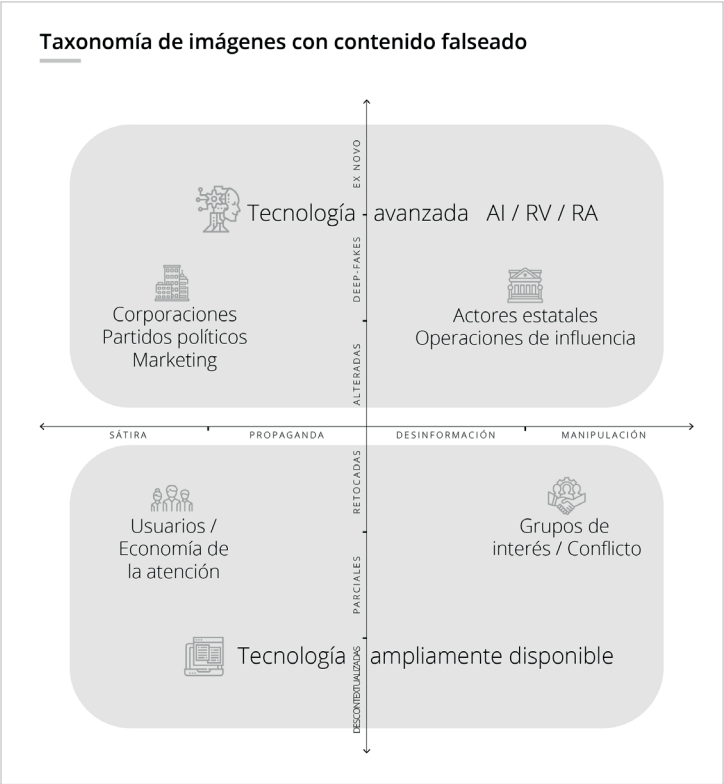


Figura 2. Taxonomía de las imágenes con contenido falseado y escenarios de uso. Propuesta de los autores.

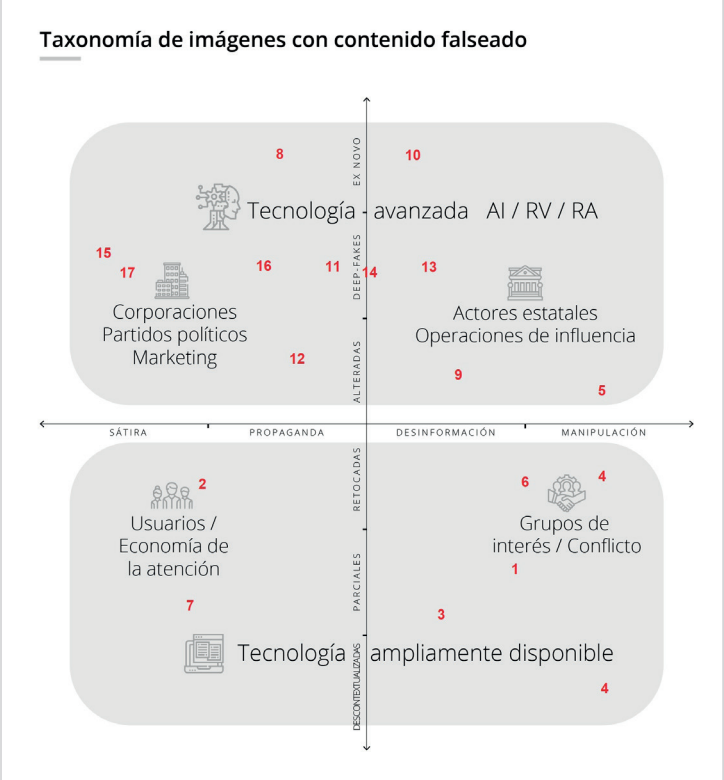


Figura 3. Ejemplos de contenidos manipulados según la taxonomía propuesta. La numeración corresponde a las imágenes de la tabla 1.

Todos los actores, si disponen de los recursos necesarios, pueden utilizar en alguna medida todas las técnicas. Por ejemplo, los que buscan atención también emplean *deep-fakes*, o los actores estatales imágenes descontextualizadas. Lo que el diagrama presentado intenta explicar es que hay herramientas que están preferentemente al alcance de unos actores y no de otros, o que tienen más aplicación en unos objetivos que en otros.

Tabla 1. Ejemplos de contenidos manipulados recogidos en la figura 3.

| | | |
|---|---|---|
| 1 | *https://networkedthought.substack.com/p/reusable-media-fact-checking*<br>*https://www.demilked.com/media-manipulating-truth*<br>*https://steemitimages.com/640x0/http://consciousreporter.com/wp-content/uploads/2015/05/media-manupilate.jpg*<br>*https://digmedia.lucdh.nl/wp-content/uploads/2019/11/73258819_2399417883517500_4203145005495222272_n-2.jpg* | Sobre la fotografía original (centro) se pueden ofrecer dos relatos antagónicos mostrando solamente una parte de la realidad:<br>*https://cheezburger.com/7036933/11-fascinating-examples-showing-how-the-media-can-manipulate-our-point-of-view* |
| 2 |  | En principio, una utilización satírica de la misma imagen del perfil de la persona afectada. Su utilización para suplantar identidades, tanto en las redes sociales como en correos electrónicos es muy frecuente, incluyendo el nombre de usuario altamente similar al original:<br>*https://www.newtral.es/bulos-fakes-falso-protestas-cataluna/20191021* |
| 3 |  <br>  <br>Abby D. Phillip<br>@abbydphillip<br>Big, boisterous crowd here in Omaha for Hillary Clinton and Warren Buffett<br>9:52 PM - Aug 1, 2016<br>132 583 people are talking about this<br>  <br>#TPPIsTreason<br>@CNNPolitics Are you reporting on the 50 people who showed up to Hillary rally in Omaha today?<br>521 172 people are talking about this | Las imágenes utilizan el zoom para ofrecer una visión parcial de la realidad que no se corresponde con el conjunto:<br>*https://www.demilked.com/media-manipulating-truth* |
| 4 |  | A la imagen se le añadió la bandera posteriormente y se atribuyó a un contexto distinto del real:<br>*https://english.elpais.com/elpais/2017/10/02/inenglish/1506943013_999238.html?rel=mas* |

| | | |
|---|---|---|
| 5 |  <br>  | Imágenes atribuidas a los disturbios en Cataluña: la primera se corresponde con los disturbios en Chile, la segunda a manifestaciones de mineros, no de independentistas:<br>*https://www.iprofesional.com/actualidad/302134-chile-pinera-Las-imagenes-de-las-protestas-masivas-en-Santiago-por-suba-del-precio-del-metro*<br>*https://www.elespanol.com/reportajes/20171006/252225238_0.html* |
| 6 |  | Imagen retocada para insertar una esvástica tatuada en el brazo del policía:<br>*https://www.newtral.es/cataluna-policia-tatuaje-esvastica-manipulada/20191021* |
| 7 | *https://twitter.com/stuart_viner/status/990549397300764673/photo/1*<br>*https://pbs.twimg.com/media/DcmuA-9WkAAZsoL?format=jpg&name=900x900*<br>*https://pbs.twimg.com/media/Fbwz3maXgAE15zO?format=jpg&name=small* | Imagen descontextualizada con un punto de vista más sugerente a mitad de camino entre la sátira y la manipulación política:<br>*https://parksrecreation.wordpress.com/2019/01/23/a-one-sided-story* |
| 8 | *Deep-fake* de audio para elaborar un "fraude del CEO" simulando la voz de un directivo:<br>*https://www.vice.com/en_us/article/pkyqvb/deepfake-audio-impersonating-ceo-fraud-attempt* | |
| 9 |  | Manipulación sobre imagen de un cementerio judío para mostrar un vehículo militar alemán como presunto culpable de la vandalización de las lápidas:<br>*https://www.wired.com/story/hackers-broke-into-real-news-sites-to-plant-fake-stories-anti-nato* |

| | | |
|---|---|---|
| 10 | *[Human prompt /* GTP-3 answer]<br><br>*The sentence: "A giraffe in a forest with trees in the background" is correct but boring. Make it sound more interesting.*<br><br>*A better way to write it is:* "A giraffe, with his long neck stretched high, in a forest filled with tall trees."<br><br>*A poetic way to write it is:* "The lonely giraffe makes me think about the loneliest person in my life. I know I am not that person because I have a lot of family and friends who care about me, but I wonder what my life would have been like if I hadn't had any friends or a family to love."<br><br>*A funny way to write it is:* "A giraffe walks slowly and peacefully in a forest full of tall trees. He probably hopes the other animals in the forest don't notice him."<br><br>*A five-year old would say:* "A giraffe looks funny with his long neck in a forest filled with tall trees."<br><br>  | Aplicación de inteligencia artificial (*GPT-3*) a la generación *ex novo* de un texto:<br>*https://towardsdatascience.com/gpt-3-101-a-brief-introduction-5c9d773a2354*<br>*https://maraoz.com/2020/07/18/openai-gpt3*<br><br>Los mismos algoritmos que se emplean en la generación de textos –hasta el momento, *GPT-2*– pueden emplearse con resultados satisfactorios para la de imágenes (**Chen** *et al.*, 2020).<br><br>Generación de caras *ex novo* mediante algoritmos de inteligencia artificial. Se pueden utilizar en la generación de perfiles falsos en redes sociales, por ejemplo:<br>*https://www.mcafee.com/blogs/other-blogs/mcafee-labs/dopple-ganging-up-on-facial-recognition-systems*<br><br>También en:<br>*https://graphics.reuters.com/CYBER-DEEPFAKE/ACTIVIST/nmovajgnxpa/index.html* |
| 11 | *Vocodes*, generador de *fakes* de audio:<br>*https://vo.codes/#speak* | |
| 12 | Manipulación de la velocidad de un video para simular embriaguez:<br>https://www.theguardian.com/us-news/video/2019/may/24/real-v-fake-debunking-the-drunk-nancy-pelosi-footage-video<br>*https://edition.cnn.com/2020/08/02/politics/fake-nancy-pelosi-video-facebook/index.html* | |
| 13 | Alteración digital de imágenes para afectar a la lectura que hace de las mismas un sistema de reconocimiento sin cambiar la apariencia para el ojo humano:<br>*https://www.labsix.org/physical-objects-that-fool-neural-nets*<br>Su contraparte sería el programa *Fawkes*, que evita que funcione el reconocimiento facial:<br>*http://sandlab.cs.uchicago.edu/fawkes/#code* | |
| 14 | *Avatarify*, programa para hacer *deep-fakes* en videoconferencias en tiempo real:<br>*https://github.com/alievk/avatarify#configure-video-meeting-app* | |
| 15 | *Deep-fakes* de videos con ánimo de sátira y entretenimiento:<br>*https://www.youtube.com/cannel/UCUix6Sk2MzkVOr5PWQrtH1g*<br><br>También en 3D:<br>*https://www.youtube.com/watch?v=nJd_2mJ4u-I* | |
| 16 | *Deep-fakes* como denuncia o generadores de confrontación social:<br>*https://www.youtube.com/watch?v=NGlE1HrBQJI*<br>*https://www.youtube.com/watch?v=cQ54GDm1eL0* (el famoso video de Barak Obama) | |
| 17 | Alteración *deep-fake* de recursos audiovisuales históricos:<br>*https://www.youtube.com/watch?v=FzOVqClci_s* | |

Aunque el primer software que puede clasificarse como *deep-fake* no se empleó sobre una imagen sino sobre un audio, la aplicación del *deep learning* al tratamiento gráfico supone un salto cualitativo importante. En aquella primera aplicación, el programa utilizó inteligencia artificial para detectar fonemas en una base de datos de grabaciones para conseguir posteriormente reproducirlos con otro discurso. El proyecto se describe en **Bregler**, **Covell** y **Slaney** (1997).

Técnicamente, para alterar un video, el algoritmo de IA genera un juego de imágenes sintéticas en un proceso de extracción de parámetros –aprendizaje– y generación. El proceso requiere una elevada capacidad de computación y un tiempo proporcional a lo convincente que sea el resultado final (**Thompson** *et al.*, 2020). El uso de algoritmos entrenados previamente sobre bases de datos existentes reduce significativamente la duración del proceso (**Vougioukas**; **Petridis**; **Pantic**, 2019). Existen numerosos algoritmos capaces de generar estas imágenes, si bien es cierto que la mayor parte no están todavía disponibles para el gran público debido al elevado consumo de recursos computacionales que requieren (**Gerardi**; **Walters**; **James**, 2020).

Existen numerosos ejemplos positivos de alteraciones sintéticas de imágenes médicas para incrementar la base de datos sobre la que entrenar algoritmos[3] para la detección de enfermedades (**Iqbal**; **Ali**, 2018). Sin embargo, su principal uso actualmente está en la industria cinematográfica y en la criminal. El término mismo se hizo popular por su aplicación a

la generación de videos pornográficos, y el primer fraude documentado data de 2019. En él se sintetizó digitalmente un audio con el fin de engañar al CEO de una compañía energética británica para que pagase 240.000 dólares a un contratista (**Stupp**, 2019).
*https://www.vice.com/en_us/article/pkyqvb/deepfake-audio-impersonating-ceo-fraud-attempt*

Existen recursos disponibles online para demostrar estas tecnologías:
*https://vo.codes/#speak*

Diversos estudios demuestran que el nivel de sofisticación que se emplea habitualmente es el menor requerido para alcanzar el propósito inmediato que se buscaba. De hecho, el uso de *deep-fakes* es relativamente marginal todavía o, al menos, no ha sido detectado en cantidades significativas. Tanto la tecnología que se precisa como el conocimiento técnico que requiere para su elaboración cuidada (no amateur) siguen sin estar al alcance de la mayoría (**Paris**; **Donovan**, 2019). Es probable no obstante que su número crezca con soluciones algorítmicas que permitan detectar formas simples de manipulación –obligando, por tanto, a un incremento en la sofisticación– y el abaratamiento de los programas capaces de generar *deep-fakes* y la creciente disponibilidad de potencia de cálculo. Hay una amplia oferta de programas en el mercado que permiten la manipulación razonablemente convincente de imágenes y videos. *AI DeepFake* o *DeepNude* habilitan al usuario para insertar imágenes o sintetizar contenidos (**Greengard**, 2019).

También hay aplicaciones de código abierto disponibles online como:

- *FakeApp*
  *https://www.malavida.com/en/soft/fakeapp*
- *FaceSwap*
  *https://github.com/deepfakes/faceswap*
- *DeepFaceLab*
  *https://github.com/iperov/DeepFaceLab*

Matt Turek, director de programa en la *Information Innovation Office (I2O)* de la *Defense Advanced Research Projects Agency* (*Darpa*) del *U.S. Department of Defense*, reconoce el diferencial de peligrosidad de los *deep-fakes* cuando dice que

> "puede afectar al proceso político, al mantenimiento de la ley y el orden, al comercio y a otros más. El hecho de que la gente pueda manipular imágenes y video fácilmente produce un amplio impacto… y reduce significativamente la confianza de la sociedad en los medios audiovisuales" (citado en **Greengard**, 2019).

Es interesante señalar que las manipulaciones proporcionan un mayor alcance mediático que las generaciones de contenidos nuevos. Así, casi el 60% de los ejemplares analizados en el estudio que llevan a cabo **Brennen** *et al.* (2020) implican alteraciones en el contenido que parten de una base real. Estas manipulaciones representan un 87% de las interacciones totales en las redes sociales.

Del mismo modo, la difusión está íntimamente vinculada a intereses concretos políticos o económicos. A pesar de que la desinformación de este tipo generada desde grupos de interés apenas supone un 20% de los contenidos totales, su viralización alcanza a casi el 70% de las interacciones totales. Resulta asimismo llamativo cómo casi el 40% de la desinformación pretende apoyarse en la credibilidad de instituciones como la *Organización Mundial de la Salud* o *Naciones Unidas*.

La falta de certeza sobre la autenticidad de un documento gráfico puede incluso permitir a un sujeto negar lo que habría sido una evidencia irrefutable en caso de no existir la posibilidad técnica de su manipulación indetectable. Es lo que se denomina "dividendo del mentiroso" y que desvía la responsabilidad sobre conductas o declaraciones que se recogen en el video a la edición del mismo (**Chesney**; **Citron**, 2018). En el ejemplo del video sobre el Emir de Qatar, efectivamente, éste declaró que el video estaba manipulado y que él jamás había hecho dichas declaraciones. Igualmente, la oposición al presidente de Gabón, Ali Bongo, tachó de *deep-fake* un video publicado por aquel en el que se pretendía demostrar su buen estado de salud. El video está disponible en
*https://www.facebook.com/tvgabon24/videos/324528215059254*

Establecida la sospecha, se dio el salto argumental hasta la certeza de lo contrario a lo que se mostraba en el video para justificar el golpe de Estado que siguió.

La mayor parte de las aplicaciones para las que puede utilizarse una manipulación profunda de contenidos gráficos están relacionadas con la justificación de actitudes o la generación de estados de ánimo puntuales. Por lo tanto, la persistencia del engaño no es necesaria más allá del momento en que se haya conseguido el efecto deseado. Una ventaja tecnológica puntual puede permitir elaborar un video convincente que no se pueda desenmascarar hasta después de que haya tenido su efecto en la opinión pública (en lo que sería una versión tecnológicamente avanzada de las supuestas pruebas presentadas por Estados Unidos para justificar su invasión de Irak en 2003).

La sofisticación de la tecnología está alcanzando un estadio en el cual la detección "a ojo desnudo" de la manipulación se va a convertir en imposible en breve (**Sayler**; **Harris**, 2019). No se trata solamente de la sustitución de la cara de una

persona por la de otra (*face-swapping*), sino que se puede sincronizar el movimiento de los labios con el discurso que se quiere insertar (*lip-syncing*) o reproducir los movimientos de un actor sobre la base de los rasgos de la persona objetivo (*puppet-master*). Aun así, debemos hablar todavía de un estado experimental de estas tecnologías (**Gerardi**; **Walters**; **James**, 2020).

Independientemente de lo sofisticada que sea la técnica empleada para la elaboración de un *deep-fake*, a partir de un cierto umbral de credibilidad lo que se explota son los procesos psicológicos de procesado de la información (**Woolley**; **Joseff**, 2020) sobre los cuales no podemos ejercer mejoras tecnológicas.

Además de los algoritmos de IA, el otro gran grupo de tecnologías que pueden influir en las imágenes con contenido falseado son la realidad virtual (RV) y la realidad aumentada (RA). Podemos definir la RV como un medio generado por simulaciones interactivas, que responden a la posición y acciones del usuario, reemplazando las percepciones de uno o más de sus sentidos, dándole la sensación de estar inmerso mentalmente o presente en la simulación (**Sherman**; **Craig**, 2018). Por su parte, la RA lo que hace es añadir nuevas capas de información sobre la realidad percibida, modificándola y enriqueciéndola (**Yndurain** *et al.*, 2010). Ambos tipos de medios comparten la idea de inmersividad.

El concepto de inmersión es común a casi todos los tipos de medios de comunicación, tanto una película, como un videojuego y hasta un libro, en todos estos casos se habla de inmersión mental o presencial, donde el usuario se aleja de alguna manera de la realidad física, para trasladarse a la realidad ofrecida por el medio en cuestión (**Otero-Franco**; **Flores-González**, 2011). Con la RV y la RA se busca dar un paso más allá y alcanzar una inmersión física sensorial, es decir:

> "la desconexión de los sentidos del mundo real y la conexión al mundo virtual. Como consecuencia, el usuario deja de percibir el entorno que le rodea y pasa a estar inmerso dentro del mundo virtual" (*Universidad Andina de Cuzco*, 2019).

Los experimentos con televisión 3D inmersiva ya muestran que el usuario puede valorar positivamente la experiencia (**Galloso** *et al.*, 2016).

La utilización de RV y RA en el contexto de las imágenes con contenido falseado no está todavía generalizada. Su uso en numerosas aplicaciones publicitarias, de medicina, de entrenamiento industrial, deportivo o militar, y de virtualización de encuentros y reuniones tiene lugar, por el momento, en entornos controlados (**Mir**; **Rodríguez**, 2020). Sin embargo, algunos estudios ya apuntan a una aceptación por parte del cerebro de los estímulos recibidos de este modo que generarían una sensación de "haber estado allí" y una interiorización en la memoria de la experiencia inmersiva (**Makowski** *et al.*, 2017).

El potencial manipulador de la RV y la RA puede provocar "contagios emocionales masivos" (**Kramer**; **Guillory**; **Hancock**, 2014). Sus aplicaciones médicas permitirían el tratamiento de la ansiedad (**Modina** *et al.*, 2015), pero también estarían en condiciones de provocarla (**Riva** *et al.*, 2007).

En resumen, el potencial de estas tecnologías para inducir conductas es mayor que el de los *chatbots* (**O'Brolchain** *et al.*, 2015). *Siri*, *Alexa* o *Xiaoice* operan con un creciente grado de empatía al que una presencia antropomórfica virtual añadiría una potente vía de acceso a nuestras emociones (**Zhang**, 2020).

## 5. Contramedidas para las imágenes con contenido falseado. *Fact-checking*

Las contramedidas para luchar contra las imágenes con contenidos falseado no tienen por qué comenzar por la tecnología. De hecho, la iniciativa en la elaboración del discurso es clave para el éxito contra la desinformación. Una estrategia comunicativa proactiva que mantenga

> "una política de comunicación clara, honesta, empática y que llegue al ciudadano" (**Calvo-Albero** *et al.*, 2020)

hace muy difícil la identificación de oportunidades para introducir elementos de desinformación. El establecimiento de una buena reputación es la mejor defensa posible (**Castro-Martínez**, 2019). En este sentido, la tecnología explota

> "la desconfianza que se ha instalado en grandes sectores de la población hacia el poder, las instituciones y las elites" (**Calvo-Albero** *et al.*, 2020).

Ningún país democrático está a salvo de estos ataques desde dentro y desde fuera de su propia estructura, y la pérdida de confianza en las comunicaciones institucionales abre una vía de entrada a fuentes alternativas (**Bennett**; **Livingston**, 2018).

Igualmente, la lucha contra la desinformación, tanto en el campo de batalla como fuera de él, es un asunto de la máxima prioridad para las autoridades militares. La *Alianza Atlántica* ha establecido en Riga (Letonia) un *Centro de Excelencia para la Comunicación Estratégica* (*Stratcom Centre of Excellence*), como centro de estudio y desarrollo de conceptos sobre algunos aspectos relacionados con la misma.
*https://www.stratcomcoe.org*

Hay que notar que, en general, este combate contra la desinformación no puede llevarse a cabo aportando datos contrapuestos y contrastados ya que el sujeto se halla inmerso en una experiencia emocional que rechaza aquellos argumentos que la pongan en crisis (**Arendt**, 2006). Es, por lo tanto, el marco emocional el que hay que poner en evidencia para

permitir que sea la audiencia la que lo rechace. A este respecto, se ha llegado a decir que el foco de la protección debería ponerse en la información contenida en los metadatos que dejan al descubierto nuestra realidad inconsciente más que en los datos que compartimos conscientemente (**Howard**, 2020).

Sin embargo, mientras que en los medios tradicionales existen tres niveles –representados por la junta de accionistas, la redacción y la dirección que intermedia entre ambas–, las redes sociales están únicamente volcadas en el negocio al haber renunciado formalmente a su rol editorial. En cualquier caso, esta labor editorial se lleva a cabo a través de los algoritmos de optimización que priorizan unos contenidos sobre otros (lo que ha dado lugar a acusaciones en Estados Unidos por parte del *Partido Republicano* de que las redes privilegian la visibilidad de los mensajes de sus rivales políticos). La solución resulta atractiva por su eficacia, escalabilidad y economía. No obstante, resulta poco fiable en la identificación y supresión de desinformación si se utiliza como único sistema (**Fard**; **Lingeswaran**, 2020).

> "La tecnología puede jugar un papel para contrarrestar la información falsa, pero solo es una parte de la solución y su mera utilización siempre estará sujeta a una carrera armamentística entre desinformación y fact-checking"

El papel central de las redes en la recolección de datos en beneficio de actores públicos y privados ha propiciado que se haya permitido, mayoritariamente, su autorregulación con una mínima supervisión del legislador. Este modelo ha permitido una mayor cercanía a la realidad del producto, pero carece de criterios neutrales que equilibren las decisiones que se adoptan. **Fard** y **Lingeswaran** (2020) concluyen que los gobiernos están, por alguna razón, menos implicados que las organizaciones en la lucha contra la desinformación.

Desde el punto de vista tecnológico, que es sólo una parte de la solución como se ha explicado más arriba, se han llevado a cabo sistemas capaces de evaluar la veracidad de una noticia de forma automatizada. Aunque sigue siendo un desafío dar con los datos precisos para preparar estos bots,

> "algunos son tan sofisticados que superan a los profesionales verificadores en el análisis de los atributos cuantificables de la noticia, como la estructura gramatical, la elección de palabras, la puntuación y la complejidad del texto" (como el de la *University of Michigan*, que "ha conseguido identificar *fake news* con un 76% de éxito, frente al 70% de los cazadores humanos") (**Flores-Vivar**, 2019).

La identificación de imágenes *deep-fake* manipuladas mediante algoritmos de inteligencia artificial no es sencilla[4]. La precisión de sus resultados, por satisfactorios que puedan ser en un momento dado [*ResNext* llega a alcanzar un 95% de precisión (**Garimella**; **Eckles**, 2020)], no está garantizada con el paso del tiempo. El rápido avance de las técnicas empleadas lleva a una competición constante entre la generación y la identificación que incrementa el grado de incertidumbre (**Gregory**, 2020). Esto ha llevado a crear una autentica competición entre equipos sobre la capacidad de cada cual para descubrir alteraciones de forma automática: el *DeepFake Detection Challenge*:
*https://www.kaggle.com/c/deepfake-detection-challenge*

La citada agencia *Darpa* ha creado un programa que permite garantizar la integridad de fotografías y videos, y describir el modo en que se llevó a cabo la manipulación en su caso: *MediFor* (*Media Forensics*). Es un programa pensado para mandos operativos y agencias de inteligencia (**Greengard**, 2019; **Sayler**; **Harris**, 2019). *MediFor* explora tanto la integridad digital (de los píxeles), como la física y semántica de la imagen. *Darpa* dispone también de otro programa relacionado con las *deep-fakes*, *SemaFor* (*Semantic Forensics*), que permite el chequeo semántico.

Igual que en el caso de la ciberseguridad –y de la seguridad en su conjunto–, las técnicas defensivas y las ofensivas avanzan en paralelo alternándose en la delantera. Como se ha dicho, no parece que sea posible una solución estrictamente tecnológica al problema de la desinformación. Más aún cuando, en el caso de las imágenes, los aspectos psicológicos y sociológicos son tan críticos como los técnicos (**Donoso-Rodríguez**, 2020). Igualmente, las técnicas de manipulación más sencillas suelen estar disponibles en versiones comerciales al alcance de casi cualquier usuario, lo que amplía también el número de potenciales agresores.

Ante la imposibilidad de actuar eficazmente sobre las amenazas, la solución que se ha sugerido es la mitigación del perímetro sobre el que pueden actuar. Será, por lo tanto, preciso operar sobre las audiencias objetivo para minimizar la probabilidad de éxito de la desinformación. Como demuestra el *Media Literacy Index* 2018 de la *Open Society Institution*, existe una correlación directa entre el nivel educativo de una sociedad, su capacidad de juicio crítico, y la libertad de prensa (**Lessenski**, 2018). Estos dos factores, educación y acceso a una información plural y veraz, son los que identifica **Horowitz** (2019) como fundamentales para blindar a las personas frente a las falsas noticias.

Los criterios que emplea un algoritmo para identificar una imagen son diferentes a los del cerebro humano. Una imagen que se corresponde de forma inequívoca con una tortuga a ojos de una persona puede ser identificada como un fusil por una máquina:
*https://www.labsix.org/physical-objects-that-fool-neural-nets*

se puede ver el resultado en
*https://www.youtube.com/watch?time_continue=3&v=qPxlhGSG0tc&feature=emb_logo*

O también *Inception*, una red neuronal de *Google*, puede confundir una fotografía de un gato con una de salsa guacamole con una probabilidad de un 99% (**Zittrain**, 2019).

Esta característica es particularmente preocupante ya que se pueden generar imágenes que resulten completamente inocentes para el observador humano, pero que sean capaces de engañar a los algoritmos. Algunos experimentos llevados a cabo con modificaciones sutiles de señales de tráfico (**Sitawarin** *et al.*, 2018) ilustran a la perfección el riesgo oculto que suponen para la conducción autónoma, especialmente cuando dichas imágenes pueden alterarse mediante proyecciones momentáneas sin dejar rastro alguno
*https://www.youtube.com/watch?time_continue=3&v=qPxlhGSG0tc&feature=emb_logo*

Se está experimentando con la introducción de marcas de agua digitales en los videos que dificultasen su falsificación (**Alipour**; **Gerardo**; **Medina**, 2019), pero no deja de ser una capa adicional que el falsificador conseguirá manipular también.

Más allá de la transparencia y la tecnología, desde 2018 han empezado a aparecer soluciones legislativas al fenómeno de las *deep-fakes* como la *Malicious deep fake prohibition act* (**Sasse**, 2018) en Estados Unidos. Las penas que se proponen varían en función de los efectos causados y de la naturaleza de la víctima. En cualquier caso, muchas voces se muestran contrarias a una legislación específica para una tecnología tan concreta y evolutiva (**Greengard**, 2019), o que es sólo un instrumento para llevar a cabo acciones criminales que ya están tipificadas (**Greene**, 2018). Es la línea que separa la libertad de expresión (**Barnes**; **Barraclough**, 2019) y el abuso de la misma lo que será preciso delimitar con cuidado (**Paris**; **Donovan**, 2019).

El carácter multifacético de la desinformación supone también el riesgo de una dispersión excesiva de la legislación aplicable [privacidad, penal, constitucional (electoral), civil…] a un mismo fenómeno en el que, además, la falta de veracidad no es necesariamente el elemento punible. Por ello, algunos países lo que regulan es su contenido ofensivo o la forma en que se disemina. Para **Barnes** y **Barraclough** (2019) cualquier legislación se encontrará con problemas a la hora de encontrar suficientes juristas expertos en estas materias, la ineficiencia de las investigaciones de faltas y delitos menores que se generan mucho más rápidamente de lo que pueden ser investigados, y –al igual que en la ciberseguridad– las dificultades de atribución de la autoría del ataque.

## 6. Conclusiones

El medio digital se presenta como el vector ideal para la generación y la transmisión de relatos interesados. En particular, la "democratización" de las tecnologías que permiten su manipulación y la viralización que se consigue a través de las redes sociales han hecho de la desinformación un fenómeno ubicuo. Si nos atenemos al uso de imágenes estáticas o en movimiento, en su apelación al componente más emocional del público permite una más rápida y profunda identificación con el mensaje.

Las falsas noticias, especialmente en forma de imágenes, tienen repercusiones en el terreno de lo personal y en lo corporativo, pero pueden afectar también a la seguridad nacional (**Bonfanti**, 2020), dado su valor histórico como garantía de veracidad. Por ello, las fuerzas armadas de todo el mundo están generando doctrinas y estrategias para la defensa frente a las técnicas digitales de desinformación y las operaciones de influencia. A este respecto, la desinformación se enmarca en las estrategias híbridas que no emplean necesariamente medios militares convencionales ni requieren de un estado declarado de hostilidades para ser empleadas en el denominado ámbito cognitivo. Al igual que el terrorismo y otras estrategias con objetivos de creación de conflicto, no solamente se emplean por parte del actor más débil para intentar contrarrestar el discurso dominante, sino que su tentadora efectividad hace que típicamente ambos bandos las utilicen como parte de su arsenal.

La desinformación no es un fenómeno nuevo, pero la utilización de la moderna tecnología digital supone un cambio cualitativo en sus capacidades. La introducción de la inteligencia artificial para la generación de *deep-fakes*, así como la generalización del uso de la realidad virtual y la aumentada, supone un salto cualitativo respecto de las formas convencionales de desinformación. En concreto, los llamados *deep-fakes*, y el uso de la realidad virtual y la aumentada generan contenidos que son aceptados inconscientemente por el receptor del mensaje como vivencias reales hasta el punto que *Facebook* ha prohibido aquellos que no resulten evidentes al público (**Bickert**, 2020). Su generalización podría derivar en un agnosticismo que termine por desvincular verdad y realidad, percepciones y realidades, de forma permanente, negando paradójicamente a lo visual su carácter de evidencia (**Barnes**; **Barraclough**, 2019). Igualmente, en una nueva paradoja, en un mundo en el que cualquier persona puede publicar, el papel del periodista con reputación ha cobrado cada vez mayor importancia, en un intento de mantener la calidad de la información.

En el caso de las imágenes con contenido falseado, utilizando tecnología ya disponible y teniendo en cuenta el casi ilimitado caudal de material gráfico que se puede encontrar en las redes sociales, cualquier atacante sólo necesitaría capacidad de computación, determinación y paciencia para llegar a producir un *deep-fake* convincente. Para ello es más importante la coherencia interna del relato falseado que su adecuación a la realidad exterior. En todo caso, por ahora, la tecnología más sofisticada está en poder de agentes estatales y grandes corporaciones que la utilizan para operaciones de influencia y marketing respectivamente.

Una taxonomía de doble entrada con un eje que describe el grado de manipulación de la información (desde descontextualización hasta imágenes generadas *ex novo* pasando por extracciones parciales, retocadas, alteradas y *deep-fakes*), y con otro eje que describe los objetivos perseguidos (desde la sátira hasta la manipulación pasando por la propaganda y la desinformación) puede ser, bajo el punto de vista de los autores, una forma de identificar los casos más relevantes y tomar las medidas más adecuadas a cada uno de ellos. Como cualquier taxonomía es imperfecta y requiere una revisión continua para adaptarse a los cambios en los dos ejes que se han propuesto. Sin embargo, desde el punto de vista de los autores, la recopilación de los casos de interés, muchos presentados a lo largo del texto, y una visión prospectiva de la evolución tanto de la tecnología como del concepto de confrontación se espera que le confieran alguna estabilidad. Tiene además la ventaja de permitir priorizar y utilizar los recursos más adecuados para contrarrestar las situaciones que se consideren más problemáticas en un entorno en el que no es posible atender todas las situaciones de interés.

Estas contramedidas o medidas de protección contra imágenes de contenido falseado comienzan con la discreción respecto de las propias imágenes online, la primera y principal estrategia defensiva a adoptar (**Gerardi**; **Walters**; **James**, 2020). A partir de ahí, el entorno digital requiere de ciudadanos conscientes del valor de sus datos, formados específicamente para reconocer sus técnicas y resiliente ante sus efectos. La resiliencia depende especialmente de una sólida identificación con los valores y principios que constituyen la base de la sociedad y de la credibilidad de sus líderes (**Reynolds**; **Parker**, 2018). La tecnología puede jugar un papel para contrarrestar la información falsa, pero sólo es una parte de la solución y su mera utilización siempre estará sujeta a una carrera armamentística entre desinformación y *fact-checking*. La credibilidad y la reputación de los medios y el uso de la legislación, nueva o existente, que aborde los distintos aspectos del ciberespacio también serán clave. Aquí, la velocidad de respuesta es el elemento crítico para evitar el posible daño.

## 7. Notas

1. Para la búsqueda se ha utilizado la expresión

> ("disinformation" or "misinformation" or "deception" or "desinformación" or "fake news" or "deep-fakes") and ("video" or "image")

restringida al período 2018-2020, dado lo novedoso de algunas de las técnicas utilizadas. En *Google Scholar* esta búsqueda arrojó más de 15.000 resultados de interés a la fecha de elaboración del artículo, de los que 426 estaban relacionados con el ámbito hispánico. Después de una revisión manual, 330 artículos realmente trataban este tema.

2. El concepto de ámbito cognitivo en las Fuerzas Armadas en España está en su fase de definición por un equipo de más de treinta expertos de varias disciplinas académicas liderado por el primer autor de este documento.

3. Un algoritmo es un conjunto de reglas que, aplicadas sistemáticamente a unos datos de entrada apropiados, resuelven un problema en un número finito de pasos elementales (**Berlanga-de-Jesús**, 2016).

4. Al tratarse de técnicas distintas y de contenidos diferentes, la casuística es muy amplia y evoluciona muy rápidamente por lo que cualquier referencia pronto queda obsoleta. **Korshunov** y **Marcel** (2018) y **Yu**, **Chang** y **Ti** (2019) recogen algunos casos recientes y su efectividad.

## 8. Referencias

**Aguado, Juan-Miguel** (2020). *Mediaciones ubicuas*. Gedisa. ISBN: 978 84 18193 58 3

**Aguado, Juan-Miguel**; **Feijóo, Claudio**; **Martínez-Martínez, Inmaculada-José** (2013). *La comunicación móvil hacia un nuevo ecosistema digital*. Gedisa. ISBN: 978 84 9784 782 7

**Alipour, Manocher C.**; **Gerardo, Bobby D.**; **Medina, Ruji P.** (2019). "A secure image watermarking architecture based on DWT-DCT domain and pseudo-random number". *International journal of recent technology and engineering*, v. 8, n. 4, pp. 4096-4102.
*https://doi.org/10.35940/ijrte.d8724.118419*

**Allcott, Hunt**; **Gentzkow, Matthew** (2017). "Social media and fake news in the 2016 election". *Journal of economic perspectives*, v. 31, n. 2, pp. 211-236.
*https://doi.org/10.1257/jep.31.2.211*

**Allport, Gordon W.**; **Postman, Leo** (1947). *The psychology of rumor*. Henry Holt and Company.

**Alonso-González, Marián** (2019). "Fake news: desinformación en la era de la sociedad de la información". *Ámbitos. Revista internacional de comunicación*, n. 45, pp. 29-52.
*https://doi.org/10.12795/ambitos.2019.i45.03*

**Arendt, Hannah** (2006). *Los orígenes del totalitarismo*. Alianza Editorial. ISBN: 978 84 20647715

**Arsenault, Amelia** (2020). *Microtargeting, automation, and forgery: Disinformation in the age of artificial intelligence*. Major research paper, University of Ottawa.
*https://ruor.uottawa.ca/handle/10393/40495*

**Austin, John-Langshaw** (1956). "A plea for excuses". In: *Proceedings of the Aristotelian Society*, v. 57, pp. 1-30. *https://www.jstor.org/stable/4544570*

**Barnes, Curtis**; **Barraclough, Tom** (2019). *Perception inception: Preparing for deepfakes and the synthetic media of tomorrow*. New Zealand: The Law Foundation. ISBN: 978 0 473 48214 5

*BBC* (2020). "Hackers post fake stories on real news sites 'to discredit NATO'". *BBC news*, 30 July. *https://www.bbc.com/news/technology-53594440*

**Bennett, W. Lance**; **Livingston, Steven** (2018). "The disinformation order: Disruptive communication and the decline of democratic institutions". *European journal of communication*, v. 33, n. 2, pp. 122-139. *https://doi.org/10.1177/0267323118760317*

**Bickert, Monika** (2020). "Enforcing against manipulated media". *Facebook*, January 6. *https://about.fb.com/news/2020/01/enforcing-against-manipulated-media*

**Bienvenue, Emily**; **Rogers, Zac**; **Troath, Sian** (2019). "Cognitive warfare". *The cove*, May 14. *https://cove.army.gov.au/article/cognitive-warfare*

**Bjola, Corneliu**; **Pamment, James** (2019). *Countering online propaganda and extremism: The dark side of digital diplomacy*. Routledge. ISBN: 978 1 138578630

**Bonfanti, Matteo E.** (2020). "The weaponisation of synthetic media: what threat does this pose to national security?". *Ciber Elcano*, n. 57. *https://cutt.ly/yxwkBG4*

**Bregler, Christoph**; **Covell, Michelle**; **Slaney, Malcolm** (1997). "Video rewrite: Driving visual speech with audio". In: *Siggraph'97: Proceedings of the 24th Annual conference on computer graphics and interactive techniques*. *https://doi.org/10.1145/258734.258880*

**Brennen, J. Scott**; **Simon, Felix**; **Howard, Philip N.**; **Nielsen, Rasmus-Kleis** (2020). "Types, sources, and claims of Covid-19 misinformation". *Reuters Institute for the Study of Journalism*, 7 April. *https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation*

**Buchanan, Ben** (2020). *The hacker and the state*. Harvard University Press. ISBN: 978 0 67498755

**Calvo-Albero, José-Luis**; **Andrés-Menárguez, David-Fernando**; **Peirano, Marta**; **Moret-Millás, Vicente**; **Peco-Yeste, Miguel**; **Donoso-Rodríguez, Daniel** (2020). *Implicaciones del ámbito cognitivo en las operaciones militares*. Documento de trabajo 01/2020. Instituto Español de Estudios Estratégicos; Ceseden. *http://www.ieee.es/Galerias/fichero/docs_trabajo/2020/DIEEET01_2020CCDC_cognitivo.pdf*

**Castells, Manuel** (2005). *La era de la información*. Economía, sociedad y cultura. 3ª edición. Alianza Editorial. ISBN: 978 84 20677002

**Castro-Martínez, Andrea** (2019). "Ciberdiplomacia y comunicación institucional: La presencia de la diplomacia digital española en redes sociales". *Revista estudios institucionales*, v. 6, n. 10, pp. 45-72. *https://doi.org/10.5944/eeii.vol.6.n.10.2019.23243*

**Chen, Mark**; **Radford, Alec**; **Child, Rewon**; **Wu, Jeffrey**; **Jun, Heewoo**; **Luan, David**; **Sutskever, Ilya** (2020). "Generative pretraining from pixels". In: *Proceedings of the 37th International conference on machine learning*, n. 119, pp. 1691-1703. *http://proceedings.mlr.press/v119/chen20s.html*

**Chesney, Robert**; **Citron, Danielle** (2018). "Deepfakes and the new disinformation war. The coming age of post-truth geopolitics". *Foreign affairs*, January/February. *https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war*

**Cloutier, Jean** (1994). "L'audioscriptovisuel et le multimédia". *Communication et langages*, n. 99, pp. 42-53.

**Codina, Lluís** (2018). "Revisiones bibliográficas sistematizadas: Procedimientos generales y framework para Ciencias Humanas y Sociales". En: Lopezosa, Carlos; Díaz-Noci, Javier; Codina, Lluís. *Methodos. Anuario de métodos de investigación en comunicación social*. Barcelona: Universitat Pompeu Fabra, pp. 50-60. *https://doi.org/10.31009/methodos.2020.i01.05*

**Davis, Zachary S.** (2019). "Artificial intelligence on the battlefield: An initial survey of potential implications for deterrence, stability, and strategic surprise". *Prism*, v. 8, n. 2, pp. 141-131. *https://www.jstor.org/stable/pdf/26803234.pdf*

**De-Granda-Orive, José-Ignacio**; **Alonso-Arroyo, Adolfo**; **García-Río, Francisco**; **Solano-Reina, Segismundo**; **Jiménez-Ruiz, Carlos-Andrés**; **Aleixandre-Benavent, Rafael** (2013). "Ciertas ventajas de *Scopus* sobre *Web of Science* en un análisis bibliométrico sobre tabaquismo". *Revista española de documentación científica*, v. 36, n. 2. *https://doi.org/10.3989/redc.2013.2.941*

**Donoso-Rodríguez, Daniel** (2020). "Aspectos psicológicos en el ámbito cognitivo de las operaciones militares". En: Calvo-Albero, José-Luis; Andrés-Menárguez, David-Fernando; Peirano, Marta; Moret-Millás, Vicente; Peco-Yeste, Miguel; Donoso-Rodríguez, Daniel. *Implicaciones del ámbito cognitivo en las operaciones militares*. Documento de trabajo 01/2020. Instituto Español de Estudios Estratégicos; Ceseden.
*http://www.ieee.es/Galerias/fichero/docs_trabajo/2020/DIEEET01_2020CCDC_cognitivo.pdf*

*EU vs Disinformation* (2020). *Actualización del informe especial del SEAE: breve evaluación de las narrativas y la desinformación en torno a la pandemia de covid-19*.
*https://cutt.ly/1xwQVZC*

**Fallis, Don** (2015). "What is disinformation?". *Library trends*, v. 63, n. 3, pp. 401-426.
*https://doi.org/10.1353/lib.2015.0014*

**Fard, Amir-Ebrahimi**; **Lingeswaran, Shajeeshan** (2020). "Misinformation battle revisited: Counter strategies from clinics to artificial intelligence". In: *Proceedings WWW'20. Misinformation battle revisited: Counter strategies from clinics to artificial intelligence*, pp. 510-519.
*https://doi.org/10.1145/3366424.3384373*

**Flores-Vivar, Jesús-Miguel** (2019). "Inteligencia artificial y periodismo: diluyendo el impacto de la desinformación y las noticias falsas a través de los bots". *Doxa comunicación*, n. 29, pp. 197-212.
*https://doi.org/10.31921/doxacom.n29a10*

**Freedland, Jonathan** (2020). "Disinformed to death". *The New York review*, August 20.
*https://www.nybooks.com/articles/2020/08/20/fake-news-disinformed-to-death*

**Galloso, Iris**; **Palacios, Juan F.**; **Feijóo, Claudio**; **Santamaría, Asunción** (2016). "On the influence of individual characteristics and personality traits on the user experience with multi-sensorial media: an experimental insight". *Multimedia tools and applications*, v. 75, n. 20.
*https://doi.org/10.1007/s11042-016-3360-z*

**Garimella, Kiran**; **Eckles, Dean** (2020). "Images and misinformation in political groups: Evidence from WhatsApp in India". *Misinformation review*, v. 1, n. 5.
*https://doi.org/10.37016/mr-2020-030*

**Gerardi, Francesca**; **Walters, Nikolay**; **James, Tomas** (2020). *Cyber-security implications of deepfakes*. University College London. NCC Group.
*https://cutt.ly/LxrKLEk*

**Gómez-de-Ágreda, Ángel** (2018). "Falsas noticias, no noticias falsas". *Telos*, n. 109.
*https://telos.fundaciontelefonica.com/telos-109-asuntos-de-comunicacion-falsas-noticias-no-noticias-falsas*

**Gómez-de-Ágreda, Ángel** (2019). *Mundo Orwell. Manual de supervivencia para un mundo hiperconectado*. Barcelona: Editorial Ariel. ISBN: 978 84 33429789

**Greene, David** (2018). "We don't need new laws for faked videos, we already have them". *Electronic Frontier Foundation*, February 13.
*https://www.eff.org/es/deeplinks/2018/02/we-dont-need-new-laws-faked-videos-we-already-have-them*

**Greengard, Samuel** (2019). "Will deepfakes do deep damage?". *Communications of the ACM*, v. 63, n. 1, pp. 17-19.
*https://doi.org/10.1145/3371409*

**Gregory, Sam** (2020). "Deepfakes and synthetic media: What should we fear? What can we do?". *Witness*.
*https://blog.witness.org/2018/07/deepfakes*

**Grijelmo, Alex** (2017). "El arte de la manipulación masiva". *El país*, 22 agosto.
*https://elpais.com/elpais/2017/08/22/opinion/1503395946_889112.html*

**Hamd-Alla, Tarek-Bahaa-El-Deen** (2007). "Credibility and connotations of image in the world of post-digital photography". In: *Philadelphia 12th Conference (Image culture)*, pp. 220-237.
*https://cutt.ly/2xrXpw1*

**Hameleers, Michael**; **Powell, Thomas E.**; **Van-der-Meer, Toni G. L. A.**; **Bos, Lieke** (2020). "A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media". *Political communication*, v. 37, n. 2, pp. 281-301.
*https://doi.org/10.1080/10584609.2019.1674979*

**Hartman, Travis**; **Satter, Raphael** (2020). "These faces are not real". *Reuters graphics*, 15 July.
*https://graphics.reuters.com/CYBER-DEEPFAKE/ACTIVIST/nmovajgnxpa/index.html*

**Holbrook, Deric J.** (2018). "Information-age warfare and defence of the cognitive domain". *The strategist*, 13 December.
*https://www.aspistrategist.org.au/information-age-warfare-and-defence-of-the-cognitive-domain*

**Horowitz, Minna-Aslama** (2019). "Disinformation as warfare in the digital age : dimensions, dilemmas, and solutions". *Journal of Vicentian social action*, v. 4, n. 2, pp. 5-21.
*https://scholar.stjohns.edu/cgi/viewcontent.cgi?article=1104&context=jovsa*

**Howard, Phillip N.** (2020). *Lie machines. How to save democracy from troll armies, deceitful robots, junk news operations, and political operatives*. Yale University Press. ISBN: 978 0 300250206

**Iqbal, Talha**; **Ali, Hazrat** (2018). "Generative adversarial network for medical images (MI-GAN)". *Journal of medical systems*, v. 42, n. 11.
*https://doi.org/10.1007/s10916-018-1072-9*

**Kania, Elsa B.** (2020). "Minds at war. China's pursuit of military advantage through cognitive science and biotechnology". *Prism*, v. 8, n. 3, pp. 83-101.
*https://ndupress.ndu.edu/Portals/68/Documents/prism/prism_8-3/prism_8-3_Kania_82-101.pdf*

**Kapantai, Eleni**; **Christopoulou, Androniki**; **Berberidis, Christos**; **Peristeras, Vassilios** (2020). "A systematic literature review on disinformation: Toward a unified taxonomical framework". *New media & society*, first online.
*https://doi.org/10.1177/1461444820959296*

**Kautilya** (2016). *Arthashastra*. CreateSpace Independent Publishing Platform. ISBN: 978 1 987699364

**Klein, David O.**; **Wueller, Joshua R.** (2017). "Fake news: A legal perspective". *Journal of internet law*, v. 20, n. 10, pp. 5-13.
*http://governance40.com/wp-content/uploads/2018/12/Fake-News-A-Legal-Perspective.pdf*

**Klein, Naomi** (2012). *La doctrina del shock: El auge del capitalismo del desastre*. Booket. ISBN: 978 84 08006732

**Korshunov, Pavel**; **Marcel, Sébastien** (2018). "DeepFakes: A new threat to face recognition? Assessment and detection". *arXiv*, 5 pp.
*http://arxiv.org/abs/1812.08685*

**Kramer, Adam D. I.**; **Guillory, Jamie E.**; **Hancock, Jeffrey T.** (2014). "Experimental evidence of massive-scale emotional contagion through social networks". In: *Proceedings of the National Academy of Sciences of the United States of America*, v. 111, n. 24, pp. 8788-8790.
*https://doi.org/10.1073/pnas.1320040111*

**Krizhevsky, Alex**; **Sutskever, Ilya**; **Hinton, Geoffrey E.** (2017). "ImageNet classification with deep convolutional neural networks". *Communications of the ACM*, v. 60, n. 6.
*https://doi.org/10.1145/3065386*

**Lessenski, Marin** (2018). *Sense wanted resilience to 'post-truth' and its predictors in the new media literacy index 2018*. Open Society Foundation.
*https://osis.bg/wp-content/uploads/2018/04/MediaLiteracyIndex2018_publishENG.pdf*

**Lin, Herb** (2018). "Developing responses to cyber-enabled information warfare and influence operations". *Lawfare*, September 6.
*https://www.lawfareblog.com/developing-responses-cyber-enabled-information-warfare-and-influence-operations*

**López-Borrull, Alexandre**; **Vives-Gràcia, Josep**; **Badell, Joan-Isidre** (2018). "Fake news, ¿Amenaza u oportunidad para los profesionales de la información y la documentación?". *El profesional de la información*, v. 27, n. 6, pp. 1346-1356.
*https://doi.org/10.3145/epi.2018.nov.17*

**Lorenz-Spreen, Philipp**; **Mørch-Mønsted, Bjarke**; **Hövel, Philipp**; **Lehmann, Sune** (2019). "Accelerating dynamics of collective attention". *Nature communications*, v. 10, 1759.
*https://doi.org/10.1038/s41467-019-09311-w*

**Maddock, Jay** (2020). "Your brain's built-in biases insulate your beliefs from contradictory facts". *The conversation*, 1 diciembre.
*https://theconversation.com/your-brains-built-in-biases-insulate-your-beliefs-from-contradictory-facts-150509*

**Mahariras, Aristedes**; **Dvilyanski, Mikhail** (2018). "Dezinformatsiya". *The cyber defense review*, v. 3, n. 3, pp. 21-28.
*https://cyberdefensereview.army.mil/Portals/6/Documents/CDR%20Journal%20Articles/CDR_V3N3_Full.pdf*

**Makowski, Dominique**; **Sperduti, Marco**; **Nicolas, Serge**; **Piolino, Pascale** (2017). "'Being there' and remembering it: Presence improves memory encoding". *Consciousness and cognition*, v. 53, pp. 194-202.
*https://doi.org/10.1016/j.concog.2017.06.015*

**Manfredi-Sánchez, Juan-Luis** (2021). *El impacto de Covid-19 en la narrativa estratégica internacional*. Instituto Español de Estudios Estratégicos.
*http://www.ieee.es/contenido/noticias/2021/01*

**Manfredi-Sánchez, Juan-Luis**; **Ufarte-Ruiz, María-José** (2020). "Inteligencia artificial y periodismo: una herramienta contra la desinformación". *Revista Cidob d'afers internacionals*, n. 124, pp. 49-72.
*https://doi.org/10.24241/rcai.2020.124.1.49*

**Marqués, David** (2020). "Se calcula que las 'fake news' han crecido un 300% con la pandemia en España". *Seguritecnia*, 1 junio.
*https://www.seguritecnia.es/entrevistas/se-calcula-que-las-fake-news-han-crecido-un-300-con-la-pandemia-en-espana_20200601.html*

**Metz, Steven**; **Johnson, Duglas V.**; (2001). *Asymmetry and U.S. military strategy*. Strategic studies institute. ISBN: 1584870419

**Miller, M. Nina** (2020). *Digital threats to democracy : A double-edged sentence*. Technology for Global Security; CNAS.
*https://www.cnas.org/publications/commentary/digital-threats-to-democracy-a-double-edged-sentence*

**Mir, Rory**; **Rodriguez, Katitza** (2020). "If privacy dies in VR, it dies in real life". *Electronic frontier foundation*, August 25.
*https://www.eff.org/deeplinks/2020/08/if-privacy-dies-vr-it-dies-real-life*

**Modina, Nexhmedin**; **Ijntema, Hiske**; **Meyerbröker, Katharina**; **Emmelkamp, Paul M. G.** (2015). "Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments". *Behaviour research and therapy*, n. 74, pp. 18-24.
*https://doi.org/10.1016/j.brat.2015.08.010*

**Molina, María D.**; **Sundar, S. Shyam**; **Le, Thai**; **Lee, Dongwon** (2019). "'Fake news' is not simply false information: A concept explication and taxonomy of online content" *American behavioral scientist*, v. 65, n. 2, pp. 180-212.
*https://doi.org/10.1177/0002764219878224*

**Moran, Richard** (2005). "Getting told and being believed". *Philosopher´s imprint*, v. 5, n. 5.
*http://hdl.handle.net/2027/spo.3521354.0005.005*

**Nettis, Maj-Kimber** (2020). "Multi-domain operations: Bridging the gaps for dominance". *Air forces cyber*, 16 March.
*https://www.16af.af.mil/News/Article/2112873/multi-domain-operations-bridging-the-gaps-for-dominance*

**Nguyen, Thanh-Thi**; **Nguyen, Cuong M.**; **Nguyen, Dung-Tien**; **Nguyen, Duc-Thanh**; **Nahavandi, Saeid** (2019). "Deep learning for deepfakes creation and detection: A survey". *Arxiv*.
*http://arxiv.org/abs/1909.11573*

**Nielsen, Rasmus-Kleis**; **Graves, Lucas** (2017). *News you don't believe: Audience perspectives on fake news*. Reuters Institute for the Study of Journalism.
*https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Nielsen%26Graves_factsheet_1710v3_FINAL_download.pdf*

**O'Brolchain, Fiachra**; **Jacquemard, Tim**; **Monaghan, David**; **O'Connor, Noel**; **Novitzky, Peter**; **Gordijn, Bert** (2016). "The convergence of virtual reality and social networks: Threats to privacy and autonomy". *Science and engineering ethics*, n. 22, pp. 1-29.
*https://doi.org/10.1007/s11948-014-9621-1*

**Otero-Franco, Antonio**; **Flores-González, Julián** (2011). "Realidad virtual: Un medio de comunicación de contenidos. Aplicación como herramienta educativa y factores de diseño e implantación en museos y espacios públicos". *Icono14*, v. 9, n. 2, pp. 185-211.
*https://doi.org/10.7195/ri14.v9i2.28*

**Paris, Britt**; **Donovan, Joan** (2019). *Deepfakes and cheap fakes*. *The manipulation of audio and visual evidence*. Data & society.
*https://datasociety.net/library/deepfakes-and-cheap-fakes*

**Paul, Christopher**; **Clarke, Colin P.**; **Triezenberg, Bonnie L.**; **Manheim, David**; **Wilson, Bradley** (2018). *Improving C2 and situational awareness for operations in and through the information environment*. RAND corporation. ISBN: 978 1 9774 0131 1
*https://doi.org/10.7249/rr2489*

**Peco-Yeste, Miguel** (2020). "Los aspectos militares". En: Calvo-Albero, José-Luis; Andrés-Menárguez, David-Fernando; Peirano, Marta; Moret-Millás, Vicente; Peco-Yeste, Miguel; Donoso-Rodríguez, Daniel. *Implicaciones del ámbito cognitivo en las operaciones militares*. Documento de trabajo 01/2020. Instituto Español de Estudios Estratégicos; Ceseden.
*http://www.ieee.es/Galerias/fichero/docs_trabajo/2020/DIEEET01_2020CCDC_cognitivo.pdf*

**Peirano, Marta** (2020). "Medios de comunicación". Calvo-Albero, José-Luis; Andrés-Menárguez, David-Fernando; Peirano, Marta; Moret-Millás, Vicente; Peco-Yeste, Miguel; Donoso-Rodríguez, Daniel. *Implicaciones del ámbito cognitivo en las operaciones militares*. Documento de trabajo 01/2020. Instituto Español de Estudios Estratégicos; Ceseden.
*http://www.ieee.es/Galerias/fichero/docs_trabajo/2020/DIEEET01_2020CCDC_cognitivo.pdf*

**Piasecki, Jan**; **Waligora, Marcin**; **Dranseika, Vilius** (2017). "Google Search as an additional source in systematic reviews". *Science and engineering ethics*, v. 24, pp. 809-810.
*https://doi.org/10.1007/s11948-017-0010-4*

**Povolny, Steve**; **Chick, Jesse** (2020). "Dopple-ganging up on facial recognition systems". *McAfee*, August 5.
*https://www.mcafee.com/blogs/other-blogs/mcafee-labs/dopple-ganging-up-on-facial-recognition-systems*

**Rettberg, Jill W.** (2014). *Seeing ourselves through technology. How we use selfies, blogs and wearable devices to see and shape ourselves*. Palgrave MacMillan. ISBN: 978 1 137 47664 7
*https://doi.org/10.1057/9781137476661*

**Reynolds, Louis**; **Parker, Lucie** (2018). *Digital resilience : Stronger citizens online*. Institute for Strategic Dialogue.
*https://www.isdglobal.org/isd-publications/digital-resilience-stronger-citizens-online*

**Riva, Giuseppe**; **Mantovani, Fabrizia**; **Capideville, Claret-Samantha**; **Preziosa, Alessandra**; **Morganti, Francesca**; **Villani, Daniela**; **Gaggioli, Andrea**; **Botella, Cristina**; **Alcañiz, Mariano** (2007). "Affective interactions using virtual reality: The link between presence and emotions". *CyberPsychology and behavior*, v. 10, n. 1.
*https://doi.org/10.1089/cpb.2006.9993*

**Rodríguez-Fernández, Leticia** (2019). "Desinformación: retos profesionales para el sector de la comunicación". *El profesional de la información*, v. 28, n. 3.
*https://doi.org/10.3145/epi.2019.may.06*

**Santos-Porras, Borja** (2020). "Cultivar el pensamiento crítico es más necesario que nunca". *The conversation*, 1 mayo.
*https://theconversation.com/covid-19-cultivar-el-pensamiento-critico-es-mas-necesario-que-nunca-137448*

**Sasse, Ben** (2018). "Malicious deep fake prohibition act of 2018". *US Congress*.
*https://www.congress.gov/bill/115th-congress/senate-bill/3805/text*

**Sayler, Kelley M.**; **Harris, Laurie A.** (2019). "Deep fakes and national security". *Congressional Research Service*.
*https://crsreports.congress.gov*

**Schmidt, Todd** (2020). "The missing domain of war: Achieving cognitive overmatch on tomorrow's battlefield". *Modern War Institute*, 4 July.
*https://mwi.usma.edu/missing-domain-war-achieving-cognitive-overmatch-tomorrows-battlefield*

**Sherman, William R.**; **Craig, Alan B.** (2018). *Understanding virtual reality. Interface, application, and design*. M. Kaufmann. ISBN: 978 0 12 800965 9
*https://doi.org/10.1016/C2013-0-18583-2*

**Sitawarin, Chawin**; **Bhagoji, Arjun-Nitin**; **Mosenia, Arsalan**; **Chiang, Mung**; **Mittal, Prateek** (2018). *DARTS: Deceiving autonomous cars with toxic signs*. Association for Computing Machinery.
*https://arxiv.org/pdf/1802.06430.pdf*

**Stupp, Catherine** (2019). "Fraudsters used AI to mimic CEO's voice in unusual cybercrime case". *The Wall Street Journal*, August 30.
*https://cutt.ly/3xt0wuo*

**Tandoc, Edson C.**; **Lim, Zheng-Wei**; **Ling, Richard** (2017). "Defining 'fake news'. A typology of scholarly definitions". *Digital journalism*, v. 6, n. 2, pp. 137-153.
*https://doi.org/10.1080/21670811.2017.1360143*

**Thompson, Neil C.**; **Greenewald, Kristjan**; **Lee, Keeheon**; **Manso, Gabriel F.** (2020). *The computational limits of deep learning*. Cornell University.
*https://arxiv.org/abs/2007.05558*

**Tzu, Sun** (2013). *El arte de la guerra*. CreateSpace Independent Publishing Platform. ISBN: 978 1 484072912

*UK Ministry of Defence* (2017). *JCN 1/17, Future force concept*. Ministry of Defence.
*https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/643061/concepts_uk_future_force_concept_jcn_1_17.pdf*

*Universidad Andina de Cuzco* (2019). *Realidad virtual introducción e historia*. Universidad Andina de Cusco.
*https://www.docsity.com/es/realidad-virtual-introduccion-e-historia/5147526*

**Valchanov, Ivan** (2018). "A taxonomy approach to fake news". *21st Century media and communications*, v. 4, n. 1, pp. 24-30.
*https://journals.uni-vt.bg/mc/eng/vol2/iss1/4*

**Vosoughi, Soroush**; **Roy, Deb**; **Aral, Sinan** (2018). "The spread of true and false news online". *Science*, v. 359, n. 6380, pp. 1146-1151.
*https://doi.org/10.1126/science.aap9559*

**Vougioukas, Konstantinos**; **Petridis, Stavros**; **Pantic, Maja** (2019). "Realistic speech-driven animation with GANs". *Facial animation*.
*https://sites.google.com/view/facial-animation*

**Warzel, Charlie** (2020). "How QAnon creates a dangerous alternate reality". *The New York Times*, August 4.
*https://www.nytimes.com/2020/08/04/opinion/qanon-conspiracy-theory-arg.html*

**Watts, Clint** (2019). *Advanced persistent manipulators, part one: The threat to the social media industry*. Alliance For Securing Democracy.
*https://securingdemocracy.gmfus.org/advanced-persistent-manipulators-part-one-the-threat-to-the-social-media-industry*

**Wheeler, Tarah** (2018). "In cyberwar, There are no rules". *Foreign policy*, September 12.
*https://foreignpolicy.com/2018/09/12/in-cyberwar-there-are-no-rules-cybersecurity-war-defense*

**Woolley, Samuel**; **Joseff, Katie** (2020). *Demand for deceit: How the way we think drives disinformation*. National Endowment for Democracy.
*https://www.ned.org/wp-content/uploads/2020/01/Demand-for-Deceit.pdf*

**Wright, Nicholas D.** (2019). *Mind space: Cognition in space operations.* Report for the Pentagon Joint Staff Strategic Multilayer Assessment Group, Intelligent Biology.
*https://www.intelligentbiology.co.uk*

**Yndurain, Elena**; **Feijóo, Claudio**; **Ramos, Sergio**; **Campos, Celeste** (2010). "Context-aware mobile applications design: implications and challenges for a new industry". *The journal of the Institute of Telecommunications Professionals*, v. 4, n. 4, pp. 16-28.
*https://www.cedint.upm.es/en/publicacion/context-aware-mobile-applications-design-implications-and-challenges-new-industry*

**Yu, Chia-Mu**; **Chang, Ching-Tang**; **Ti, Yen-Wu** (2019). *Detecting deepfake-forged contents with separable convolutional neural network and image segmentation*. Cornell University.
*http://arxiv.org/abs/1912.12184*

**Zhang, Wanqing** (2020). "The AI girlfriend seducing China's lonely men". *Sith tone*, December 7.
*https://www.sixthtone.com/news/1006531/the-ai-girlfriend-seducing-chinas-lonely-men*

**Zittrain, Jonathan** (2019). "The hidden costs of automated thinking". *The New Yorker*, July 23.
*https://www.newyorker.com/tech/annals-of-technology/the-hidden-costs-of-automated-thinking*