REVISTA
DE INVESTIGACIÓN
EN ARTES VISUALES

| Año 201X | septiembre | Nº. 11
https://doi.org/10.4995/aniav.2022.17309

La imagen como forma de (des)conocimiento en la era del deepfake

Este artículo investiga las repercusiones de la irrupción del *deepfake* en nuestra confianza epistémica en las imágenes

The image as a form of (un)knowing in the era of deepfake

León-Mendoza, Raúl
Universitat Politècnica de València, raulemen@esc.upv.es

Recibido: 10-03-2022
Aceptado: 07-09-2022

Citar como: León-Mendoza, Raúl. (2022). La imagen como forma de (des)conocimiento en la era del deepfake. ANIAV - Revista de Investigación en Artes Visuales, n. 11, p. 53-70, septiembre. 2022. ISSN 2530-9986.
Doi: https://doi.org/10.4995/aniav.2022.17309

PALABRAS CLAVE
Imagen, representación, conocimiento, *deepfake*, verdad, inteligencia artificial.

RESUMEN
Las máquinas que producen *deepfakes* no necesitan del mundo real para producir representaciones verosímiles. Esto debería producir la crisis del viejo *contrato de veridicción* que mantenemos con las imágenes. La irrupción del *deepfake* convierte la totalidad de las imágenes en materiales para una agencia artística. Sin embargo, instituciones como la justicia mantienen la confianza en la imagen como dispositivo de conocimiento. En un contexto social de profunda sospecha sobre cualquier tipo de representación, la imagen aún se utiliza para determinar lo verdadero y lo falso. Desde la aparición del *deepfake* se suceden medidas y contramedidas, cada vez más dependientes de la inteligencia artificial, destinadas a seguir manteniendo a las imágenes como parte del aparato epistemológico que empleamos para conocer el mundo.

En este artículo, a través de diferentes estudios de caso, bibliografía y documentos relevantes sobre el tema, se traza una línea sobre el esfuerzo por comprobar la autenticidad de las imágenes como representaciones fieles de la realidad. Esta línea une las viejas *fake pictures*, con el nuevo fenómeno del *deepfake*. Concluimos cuestionándonos si las imágenes no son ya sólo un simple reflejo de nuestros deseos de ver, al tiempo que

hemos abandonado y entregado totalmente a las máquinas nuestro aparato crítico de determinación de la verdad. En este contexto, lo que es real se dirime en una conversación en entre máquinas (GAN) por su apariencia verosímil, pero no por su contenido.

KEY WORDS
Image, representation, knowledge, deepfake, truth, artificial intelligence.

ABSTRACT
The machines that produce deepfakes do not need the real world to produce plausible representations. This should produce a veracity crisis of the image. However, institutions such as justice maintain confidence in the image as an epistemological tool. In a social context of deep suspicion about any type of representation, the image is still used to determine what is true and what is false.

Since the appearance of deepfake, measures and countermeasures have been taking place, mostly dependent on artificial intelligence, aimed to maintain the images as part of the methods we use to interpret the world.

In this article, through different case studies, bibliography and relevant documents on the subject, the different methods of authentication of images as faithful representations of reality are described. These unite the old fake pictures, with the new phenomenon of deepfake.

We conclude by questioning whether at present images are a simple reflection of our desire to see what we wish. So all images are an artistic agency. At the same time that we have abandoned completely and handed over to machines our critical capacity for determining truth. In this context, reality is decided by its plausible appearance, but not by its meaning.

INTRODUCCIÓN

La aceleración, automatización y nivel de verosimilitud que han alcanzado los medios de creación algorítmica de imágenes con el empleo del *deeplearning (IA)*, producen un estado de indeterminación de la *verdad* a través de la imagen, que debilita el paradigma científico en lo referente la construcción de la verdad por correspondencia, abriendo una brecha en nuestras formas tradicionales de acceso al conocimiento.

A pesar de que "la difusión de imágenes científicas falsas (...) tiene una larga tradición en la historia del conocimiento"(López-Cantos y Maestre Gasteazi 2019), la irrupción y desarrollo exponencial de los *deepfakes,* complican hasta el límite nuestra vieja y problemática relación con las imágenes. Con su salida de su ámbito primigenio de la pornografía y su incursión en otros ámbitos del conocimiento, los *DF* finiquitan el *contrato de veridicción* mantenido religiosamente entre nosotros, la imagen y el conocimiento. Sólo la verosimilitud de las imágenes producidas por *Redes generativas*

antagónicas de computadoras *(GAN)*[1], que se manifiesta en los *DF*, está consiguiendo erosionar el papel de la imagen como "objeto epistémico con naturaleza probatoria" (López-Cantos, 2010).

Nuestra relación hacia los *dispositivos* (documentos, instrumentos, instituciones) que tradicionalmente contenían la *verdad* está cambiando por la falta de *confianza epistémica*, lo que está debilitando es el acceso a la *realidad* y disipando la formación de un mínimo consenso intersubjetivo sobre el mundo, que está ahora orientado hacia la producción-satisfacción de un deseo.

Los *DF* abren la imagen a la potencia pura que supone no vincularse con ningún referente. En un contexto donde la imagen no garantiza ningún régimen de conexión con la *realidad*, todo es representable con verosimilitud. Y sin embargo nos resistimos a *flotar* con *autonomía* en esta nueva distribución de lo plausible. Seguimos necesitando mantener el estatus de conocimiento objetivo de las imágenes, que sostiene la reproducción de ciertas relaciones de poder (autoridad), para lo que se hace necesario también la emergencia de dispositivos de juicio sobre las condiciones de *verdad* de las imágenes, que está transitando desde el juicio experto forense (humano) al juicio algorítmico forense (máquinas). Con este último giro hacia las *máquinas*, las fronteras de la epistemología se sitúan donde marca el resultado de una conversación entre dos *máquinas* de computo.

1. Parecer-ser. Estatuto de las imágenes como herramientas de conocimiento

Las características de la imagen como material de acceso al conocimiento de la verdad (prueba) corren en paralelo en los ámbitos de **conocimiento científico** y **judicial**[2].

> "Una imagen digital que se quiera usar como prueba en un juicio o ser presentada en una publicación científica debe ser una evidencia que, en derecho procesal, se refiere la certeza clara, manifiesta y tan perceptible que nadie pueda dudar de ella."(López-Cantos y Maestre Gasteazi 2019)

El tradicional papel de la imagen en el campo de la identificación de personas, sucesos y cosas sigue vigente en todos los entornos administrativos de nuestras vidas. En especial, es aún un *dispositivo duro* en la lógica de la administración de justicia[3]. Para el juez Joaquim Bosch:

> "Si muge como una vaca, tiene cuernos como una vaca y da leche como una vaca, está claro que estamos ante una vaca, aunque queramos cambiarle el nombre.(Viúdez, 2015)"[4]

[1] En adelante, *las máquinas* o *la máquina.*

[2] "Es esta condición de competencia pública radical la que hace de todo proceso jurídico el principal foco de atención a la hora de analizar las transformaciones intensa y profundas de la condición de verdad en nuestra vida en común actual, (...)."(Marzo 2019)

[3] Genealogía de la relación entre la imagen y la administración de justicia en: (Tagg 2005)

[4] Coronando esta reflexión podría estar: Si es capaz de juzgar, sabiendo que la realidad es inescrutable, estamos ante un juez.

Esta confusión entre una **correlación** de hechos que se entienden como **causas** y su institución como verdad-**efecto**, tiene en derecho el nombre de ***verdad judicial***[5]. Esta correlación de hechos se nutre con lo que conocemos bajo el nombre de ***pruebas***. Entre otros *medios de prueba*, las leyes españolas (Código Penal[6], LEC[7], LECrim) conceden a las imágenes un estatus de *prueba documental* con *eficacia* probatoria.

A pesar de todo, la relación de ciertas categorías de la imagen con la **representación de la *verdad*** ha mantenido un **consenso intersubjetivo** desde el inicio de la fotografía hasta nuestros días, que sigue operativo a pesar de falsificaciones, manipulaciones conocidas e intentos de todo tipo por evidenciar su inconsistencia. Es inevitable, citar a Joan Fontcuberta, la persona que en el Estado Español ha ejercido un papel relevante en la divulgación de la inconsistencia de la relación imagen-verdad:

> "Toda fotografía es una ficción que se presenta como verdadera. Contra lo que nos han inculcado, contra lo que solemos pensar, la fotografía miente siempre, miente por instinto, miente porque su naturaleza no le permite hacer otra cosa." (Fontcuberta, 1997)

Figura 1. Imagen de caso "Cura de Churra" distribuidas por el Huffingtonpost, en las que presuntamente se identifica al párroco en una zona de *cruising*. El párroco denunció las imágenes, asegurando que no había estado en ese lugar. Tras un proceso judicial que confirmaba que las imágenes "no habían sido manipuladas" el párroco se enfrenta a un juicio por denuncia falsa.

[5] Aunque se asume de partida que esta *verdad judicial* pueda coincidir sólo en parte con la llamada *verdad material* (nosotros la hemos llamado *realidad),* es *verdad judicial* la que determina la *pena* se impone.

[6] (*Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal.* s. f.). Artículo 8(*Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal.* s. f.): A los efectos de este Código se considera documento todo soporte material que exprese o incorpore datos, hechos o narraciones con eficacia probatoria o cualquier otro tipo de relevancia jurídica.

[7] (*Ley 1/2000, de 7 de enero, de Enjuiciamiento Civil.* s. f.) Sección 8. *Artículo 382.1. Instrumentos de filmación, grabación y semejantes. Valor probatorio:* Las partes podrán proponer como medio de prueba la reproducción ante el tribunal de palabras, imágenes y sonidos captados mediante instrumentos de filmación, grabación y otros semejantes.

2. Sospechas epistémicas

2.1. Antes de la llegada del deepfake

Por lo tanto, partimos conceptualmente de una imagen que, hasta el momento de su captura, está aparentemente desvinculada de la posible intervención humana. La imagen fotográfica es como una epifanía que establece un régimen absoluto de correspondencia con la realidad.

> "Como espectadores confiamos en que lo que ocurrió frente a la cámara haya sufrido escasa o nula modificación".(Nichols, 1997)

Nuestra relación intensa con la producción y el consumo de imágenes, ha *desapercibido* cualquier *manipulación*[8] de la imagen en lo que llamamos la fase de producción y las ha hecho transparentes e irrelevantes de cara a construir intersubjetivamente una relación dura entre el documento y la **verdad**. Sin embargo, sabemos que la *construcción* de la *verdad* se inicia antes de la producción del documento y atraviesa todas las fases del proceso de producción y recepción. Cualquier ejercicio de representación de la *realidad*[9] es *performativo* y sus injerencias se manifiestan, están presentes y activas en todo momento actuando sobre aquello que consideramos *verdad*[10]. La *verdad*, por lo tanto, ya era *posverdad*. En tanto que representación, la fotografía ya era *postfotografía*.

Es en relación a lo que conocemos como *postproducción de la imagen* donde se produce cierta ruptura del consenso social sobre la *integridad de la imagen* y se abre el imaginario a la posibilidad de la manipulación documental. Esta *desconfianza epistémica* es la consecuencia de una riqueza cultural de estímulos visuales que mantienen diversas relaciones con el concepto tradicional de verdad. En nuestra cultura visual contemporánea se produce una convergencia entre medios tecnológicos y contextos de representación que suprimen las barreras de lo posible haciendo que se pueda producir con verosimilitud cualquier tipo de representación.

En este sentido, como espectadores ponemos en marcha *mecanismos de sospecha*, pero siempre en el juicio de elementos formales de la imagen. Estos elementos formales identifican el contexto interpretativo[11] bajo el que entender el acto comunicativo que presuponemos que tienen las imágenes. El valor de *verdad* documental, presuntamente

[8] Rara vez se levantan suspicacias sobre el conjunto de operaciones de manipulación que pueden hacer plausible cualquier simulacro durante los procesos de preproducción o producción de una imagen (captura) o contexto de recepción. Operaciones como son la suplantación de identidad(*BBC News Mundo* s. f.), los sesgos de encuadre, de tiempos de captura, de frecuencias de muestreo, resolución de la imagen, contexto donde se inserta el documento, etc. no suelen generar un gran "recelo" social sobre la veracidad de un documento.

[9] Entendemos *realidad* como lo preexistente inaccesible.

[10] Entendemos *verdad* como el documento que cierra una única interpretación sobre la realidad.

[11] Es por esto por lo que formatos como el *falso documental* son tan disruptivos, porque forma y contenido (verdad) no mantienen la relación esperada.

contenido en una imagen respeta (concuerda) formalmente la *gramática* propia[12] de cada campo de producción de imágenes al que socialmente asociamos distintos valores de correspondencia con la *realidad*.

Como espectadores nos hemos convertido en una *legión de investigadores forenses* (Lewis, 2019). Cuando desconfiamos de la veracidad de las imágenes buscamos el *glitch*[13]. Nos descubrimos analizando de forma automática algunos elementos formales que identificamos como relacionados con la manipulación de la imagen.

Figura 2. Fragmentos de Beyond the Aquila Rift, momento en el que se rompe la simulación. (Boidin et al. 2019)

Consideramos estos elementos formales como *defectos* que aparecen cuando la imagen es *sometida* a posproducción y por lo tanto tendemos a considerarla como *falsa*. Estos ***defectos formales***[14] pueden ser ***temporales*** o ***espaciales*** y determinarían la manipulación de la imagen y la inconsistencia de esta como prueba.

> "Aunque no seamos conscientes de ello, nuestro cerebro está acostumbrado a detectar este tipo de aspectos de una forma bastante natural, porque lo hemos aprendido desde pequeños."[15]

Atentos todos al momento en que la representación se *rompe* y nos muestra sus debilidades, su impostado estatuto de prueba de la verdad, sus deficientes conexiones con esa ***materia oscura*** que es la *realidad*.

[12] El *espectador-forense* se preocupa de que las imágenes no violen las reglas del marco formal de la representación de la realidad. Construcción de la verdad por coherencia.

[13] En adelante: *defectos.*

[14] En el caso de lo que conocemos como **montajes** en postproducción, basados en la superposición de imágenes diferentes que a través de operaciones de adición y sustracción dan lugar a la composición de una única imagen, buscamos **defectos** como ***incoherencias espaciales*** (cuerpos anatómicamente *imposibles,* iluminaciones discordantes), que nos indicarían la presencia de dos fuentes de imagen y por lo tanto de dos temporalidades superpuestas. Por contra algunos, ***defectos espaciales* de la imagen** (pobrezas de resolución, defectos de enfoque y de encuadre), tendemos a consideramos como *máculas de la contingencia en la toma de la imagen* que refuerzan el valor (verdad) de asociación entre documento y *realidad*.

[15] Josep Navarro en: (F. Casal s. f.)

> "Estas imposturas posmodernas que se adueñan de la verosimilitud documental contienen, implícitamente, un aviso en torno a la supuesta evidencialidad de las imágenes y sugieren la imposibilidad de las representaciones para garantizar la verdad de lo que reflejan."(García Martínez, 2007)

Sin embargo, restringir al tiempo de la ***postproducción***, la ventana temporal donde nos es concebible la manipulación de la representación es un síntoma de un ***aparato crítico desenfocado***. Esta lógica de la manipulación olvida que toda enunciación es en sí misma una manipulación y mantiene la creencia en una materia prima (realidad) que puede ser capturada y conservada sin interferencias en una imagen-documento (verdad).

2.2. El deepfake. Implosión técnica que alisa la epistemología.

> "Un *deepfake* es un vídeo que superpone la cara de una persona en el cuerpo de otra."(Martínez y Castillo, 2019)

El término *deepfake* fué acuñado en noviembre de 2017 por un usuario de la plataforma *Reddit.* Este usuario creó un foro en la plataforma con el nombre de *deepfake* para iniciar una comunidad de discusión sobre el uso de *programas* de *deep learning (Inteligencia Artificial)* para el intercambio sintético de caras en videos pornográficos. El propósito de estos videos era perfeccionar los *fakes* sustituyendo la cara de la actriz *original* por la cara de una actriz famosa de cine *convencional*. Aunque en un altísimo porcentaje (96%) los *deepfakes* están centrados en el ámbito de la pornografía, ya se han producido casos de injerencias en otros ámbitos como el político, sobre todo en países en vías de desarrollo.

Los *deepfakes* pueden ser generados con diferentes técnicas de *deep learning*, pero la más habitual por su *calidad de salida* de las imágenes (nivel de verosimilitud) es la conocida como *GAN (Generative Adversarial Network)*. El concepto de *DF* que actualmente manejamos lleva implícita la prácticamente total automatización del proceso[16] de la generación de la imagen de síntesis. La popularización de los *DF* está íntimamente relacionada con la difusión de algoritmos gratuitos junto con la expansión de servicios web y aplicaciones que ponen esta tecnología al alcance del usuario no experto.

Cuando intentamos detectar un *deepfake,* buscamos **rasgos impropios de lo vivo**[17] en aquello representado. Esto se dirime en *distensiones* y *contracciones* artificiales del

[16] Se pueden rastrear los inicios de esta automatización de la falsificación en: (Bregler, Covell, y Slaney 1997)

[17] La búsqueda de coherencia en la imagen basada en la ausencia o baja frecuencia en el parpadeo de los ojos del cuerpo representado, será un rasgo paradigmático en la detección temprana (humana) de los *deepfakes*. Este defecto formal, deja patente que las *GAN* que producen las *deepfakes*, se alimentan de un ecosistema en el que las imágenes son demasiado uniformes (representaciones de cara ojos siempre abiertos) y por lo tanto hacen emerger la escasez de materiales cierto tipo de materiales (fotografías de caras con los ojos cerrados) que serían necesarias para avanzar hacia un grado mayor de verosimilitud. Puede que nuestro sesgo a la hora de producir y compartir representación haya supuesto, en una fase inicial, cierta debilidad para la

tiempo que se reflejan en incoherencias entre el tiempo representado y los cuerpos (posibles) presentes en las imágenes.

> "El parpadeo de los ojos, un movimiento facial natural… Así se detecta un *deepfake*."(F. Casal s. f.)

A pesar de estos indicios, en los que los humanos nos podemos basar en la detección de los *deepfakes,* todas las fuentes consultadas apuntan a que estas imágenes alcanzarán tal grado de sofisticación tras la que en el nivel de recreación de la realidad (verosimilitud) la percepción humana será incapaz de ejercer su habitual capacidad de juicio. Es decir, mirando la imagen no se puede concluir una correlación material entre la representación y la existencia de un suceso productor de la misma, puesto que dicho suceso ya no es necesario para la aparición de la imagen.

3. La verdad es una materia plástica y sin embargo, esto es insoportable.

El papel que juega nuestra vasta cultura visual, sometida al consumo de productos audiovisuales con distinto tipo de relación con la realidad, "ha transformado la sensibilidad ante lo verosímil y lo verdadero en los espectadores, estos son los quienes deben decidir continuamente lo que es real y lo que es artificial, determinar donde comienza la manipulación y dónde termina la realidad."(Gandasegui 2012)[18]. Por lo tanto, es el *espectador* el que tendría la responsabilidad[19] constante de empujar sistemáticamente la información (premisas) al interior o al exterior de los campos que establece el esquema epistémico (verdad, conocimiento, creencia). Este mismo esquema, de cierre del sentido depositado en el espectador, es el actualmente vigente en el ámbito del arte contemporáneo.

Sin embargo, los *DF* han irrumpido en el flujo de imágenes contemporáneas desbordando los usos que antes estaban reservados a las manipulaciones más lentas y costosas basadas en aplicaciones tradicionales de composición de video (postproducción). Con esta ***aceleración*** han conseguido desestabilizar por fin la dura relación entre imagen y *verdad* socialmente aceptada.

> "(..) our historical belief that video and audio are reliable records of reality is no longer tenable."(Adjer et al., 2019)

La persistencia del concepto *verdad* que se da a ver en forma de documento *ajustado* de la realidad, solo es mantenida en forma de anhelo nostálgico, por sostener parcialmente llenos de contenido conceptos como *libertad, justicia o política* que ya no atienden a su

verosimilitud de las imágenes sintéticas, pero todos sabemos que es cuestión de tiempo (poco) que esta debilidad sea superada.

[18] A propósito del papel del *falso documental* como *punta de lanza* de esta desconfianza epistémica en el resto de representaciones, escribe Diaz Gandasegui que incorpora numerosas referencias a otros autores que sitúan a este género cinematográfico en el centro de un movimiento entrópico hacia la indeterminación de la imagen (de cualquier tipo) como forma de conocimiento de la realidad.

[19] Aunque puede que ya ni puedan, ni quieran hacerlo.

antiguo significado. La aparición de representaciones en cuya producción ya no está involucrada la *realidad*, contribuye al *desplazamiento* de estos conceptos.

Este *desplazamiento* está acompasado con la forma en que grandes capas sociales están desplazando antiguas formas de consolidación de la *verdad*, hacia un proceso de *acomodación* de la realidad a su marco emocional. El efecto de esta *desconfianza epistémica* generalizada es que *régimen de veridicción* de la información en los últimos sucesos políticos disruptivos[20] "no ha sido considerado como un aspecto esencial" a la hora de inclinarse por su opción política. Como se explica en *"Régimenes de veridicción y simulacros de la política"*(Aldama, 2020), el viejo sistema de enunciación de la **verdad** (forense) que estaba basado en la intención *hacer-parecer-cierto* una determinada proposición, ha mutado en procesos más *performativos* de construcción de la verdad, tanto en el proceso de enunciación como en el de recepción.

> "las viejas categorías de verdad y falsedad no parecen demasiado útiles para explicar los fenómenos audiovisuales que genera la cultura popular en la actualidad."(Ges, 2001)

Estas interacciones de relación con lo antes conocido como *verdad* van desde el *no creer-no ser* (dudar: relativismo) hasta el *hacer-como si* (fingir: cinismo). Esta performatividad hace de todo discurso algo insignificante si no es capaz de contribuir a la producción-representación de un acontecimiento deseado de antemano.

Por lo tanto, la *verdad* ya no es cosa del pasado sujeta a algo preexistente, sino una materia *plástica* al servicio del deseo de un determinado presente-futuro. La *verdad* ya solo tendría que representarse como una correspondencia entre la representación y el deseo. O dicho de otro modo, **sólo las representaciones que responden a mi deseo son verdad.**

> "Realidad[21] y falso son términos que en nuestra cultura se plantean tradicionalmente como opuestos, pero, sin embargo, el desarrollo de tecnologías capaces de recrear la realidad con total fidelidad ha hecho que aparezcan estos espacios intermedios e híbridos que combinan la realidad y la irrealidad."(Gandasegui, 2012)

La aparición de los *deepfakes* provoca que este *espacio intermedio* se haya extendido a todas las imágenes haciendo que la categorización *verdadero-falso sea inservible*. En este contexto de interpretación en el que lo verdadero y lo falso son indistinguibles suprime el estatus de la imagen como documento. La apertura que la irrupción de los *deepfakes* operan en el ecosistema de imágenes es *performativa*, desplaza la totalidad de las imágenes hacia un territorio en el que ya no están sujetas a la *verdad* y las convierte en una *agencia artística*. Por lo tanto, la vieja autonomía artística que se

[20] El artículo citado se refiere al Brexit, a la elección de Donald Trump en Estados Unidos y a la Declaración de Independencia de Cataluña.

[21] Hay que puntualizar aquí que, a lo que el autor refiere como *realidad* (preexistente inaccesible) nosotros nos referimos como *verdad*.

expresaba en nuestra labor para producir *fantasmas*[22], se ha extendido por completo a la masa social del tecnocapitalismo. El *DF* supone una apertura técnica a la posibilidad de ver-conformar en una imagen verosímil nuestros deseos proyectados. Tendremos que ver porque nuestros deseos en un alto porcentaje no van más allá del porno y otros corolarios de clichés patriarcales[23].

4. Contramedidas

A pesar de esta apertura, la imagen sigue siendo necesaria como *interfaz* de conocimiento (prueba) para ciertos sectores hegemónicos del conocimiento. Para mantener y reproducir cierto poder, es necesario continuar estableciendo juicios sobre la imagen, produciendo una ***tensión extrema*** hacia polos de atracción que parecen mutuamente invalidantes: verdad-creencia.

> "Aunque sirve para hacer cualquier tipo de video, la gente parece más interesada en seguir creando pornografía falsa."(Martínez y Castillo, 2019)

Estamos entonces buscando discernir entre *pornografía verdadera* y *pornografía falsa.* Lo que activamos por lo tanto son *dispositivos* de corte epistemológico. Con estos ejercicios de juicio por oposición, no hacemos más que perpetuar la creencia en una materia prima (realidad preexistente) que puede *ser capturada* sin interferencias (imagen-documento) en un documento que cierra la interpretación del pasado. A pesar de que estas categorías ya no son del todo relevantes socialmente, de alguna forma no podemos vivir en una continua tensión de duda sobre la consistencia de la verdad[24] y necesitamos reconstruir, cierto consenso sobre lo real.

Esta tensión es extenuante y nos obliga a tomar posiciones ante lo que estamos viendo en un momento donde realidad y representación ya no tienen por qué guardar ningún nexo (*deepfake*). Por lo tanto, nuestro viejo *aparato crítico* ya no tiene elementos de juicio válidos. No podemos establecer un juicio crítico cuando la *correspondencia* y la *coherencia* han sido desbordadas (devoradas) completamente por *aparatos técnicos (IA)*. Ante semejante panorama solo podemos construir la verdad por consenso a través de aceptar un *juicio experto* (forense), o encomendarnos de nuevo al *aparato técnico de verdad algorítmica* (*máquinas*).

Por todo lo anterior está claro que estamos en una *guerra epistemológica* de imágenes *difusas*, en un escenario visual en el que seguimos teniendo necesariamente que dirimir entre opuestos: *verdad-falsedad; documento-falsificación*. Esto es una guerra, el ataque es la *mentira* (entendida como apariencia fabricada) y sin contramedidas (aparatos de

[22] Véase "La producción de «fantasmas» como competencia intrínseca del arte contemporáneo." En: (González-García, 2019)

[23] Ver conclusiones en: (Martínez y Castillo 2019)

[24] Es interesante la reflexión sobre la imposibilidad de vivir en algo así como un *eterno* Día de los inocentes (April Fools' Day para los anglosajones) cuestionando permanentemente el estatus de verdad de nuestros documentos. Esta idea viene de: (Gandasegui, 2012)

veridicción) la guerra está perdida. A esta guerra, del *bando del mal* y como arma más potente, ha llegado el *DF* y las contramedidas son la única defensa disponible:

> "Without defensive countermeasures, the integrity of democracies around the world are at risk"[25]

4.1. El experto vivo. El perito forense. El humano asistido.

En el extremo más duro de esta tensión por mantener activas y operativas las viejas categorías en torno a la imagen de las cuales depende su subsistencia (reproducción de *régimen de sentido*), está la administración (*biopolítica*) y en especial la administración de justicia.

> "La **imagen digital** es cada vez más popular y poco a poco va ganando protagonismo en los procesos judiciales. Puede tratarse de una foto que se presenta como prueba para incriminar a una persona de un delito (...)respecto a la que se quiere saber quién la ha realizado o compartido, o incluso de una foto que ha sido manipulada."(«Análisis forense de imágenes» s. f.)

Incluso tras la aparición del *deepfake,* en el contexto legal la imagen sigue siendo una herramienta de identificación de personas, cosas y sucesos (casi absoluta). De la capacidad de emitir un *juicio de veridicción* sobre una imagen depende, en ocasiones, toda la construcción de la *verdad judicial.* Este ejercicio de juicio experto se despliega en unidades policiales especiales y en estudios superiores, bajo la **creencia íntegra de la imagen como prueba**, índice de un suceso con capacidad de establecer correspondencias capaces de dar acceso al pasado (predecir el pasado).

> "Digital image forensics is a brand new research field which aims at validating the authenticity of images by recovering information about their history."(Redi et al. 2011)

El perfil profesional que despliega las metodologías de análisis sobre las pruebas es el *perito forense* especializado en análisis de imágenes "encargado de garantizar la certeza en el uso de una imagen digital que pretende utilizarse como **evidencia** en una investigación de carácter legal."(Mendoza, 2016)

Las metodologías desplegadas están en cierta medida estandarizadas y se basan extraer conclusiones a partir del análisis de características propias de la imagen digital. Las conclusiones que se pueden extraer de estos análisis están orientadas, por una lado, a determinar la relación inequívoca entre una cámara digital y las imágenes producidas por la misma *(fuente de la imagen)*; en una segunda, vertiente se pronuncian sobre si existen indicios de *manipulación* en la imagen considerada *original*.

En primera instancia se analizan los *metadatos* del archivo de imagen, información anexa[26] e invisible que se guarda en los soportes de información (memorias) en el

[25] https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

[26] Los metadatos organizan información específica en campos estandarizados sobre la captura concreta que identifica el dispositivo de captura (marca y modelo); las condiciones técnicas de la

momento de hacer la captura de la imagen. El segundo grupo de operaciones sobre la imagen se centra en identificar con qué dispositivo específico se han realizado las imágenes objeto del análisis, a través de caracterizar[27, 28] de la forma más precisa posible los *defectos* y particularidades del *sensor, del conversor analógico-digital* o del algoritmo de compresión presentes en el dispositivo de captura de imágenes.

> "Al comparar estos patrones, es posible vincular una imagen determinada a la cámara que la tomó, de la misma manera que una huella dactilar se puede vincular a una persona individual."(Veiligheid, 2013)

El tercer grupo técnicas intenta determinar si ha incorporado o suprimido[29] información gráfica (visual) en la imagen después de su *captura*. El axioma de estas técnicas gira alrededor del concepto de imagen *original*, heredado del elemento que en fotografía analógica se conocía como película o negativo.

Este *aparato científico-técnico,* que concluye en la confección de un informe pericial, aflora el gran esfuerzo que supone afirmar que una imagen no está *manipulada o* lo está y abre la duda sobre la totalidad del sistema de producción de verdad instituido en base a las imágenes. Puesto que es la propia *naturaleza* de la imagen la que permite su manipulación, toda operación de peritaje, aún teniendo un margen de confianza, es un esfuerzo predictivo sobre una materia inaccesible (*realidad*) que es el pasado. La imagen se ingresa así en un bucle argumental del que no puede salir[30]: predicción, contra-predicción. Así lo expresa el artículo *¿Podemos confiar en el análisis forense de imágenes digitales?*

toma (apertura, velocidad de obturación, distancia focal, datos de exposición); datos espacio-temporales de la misma (fecha y hora, geolocalización).

[27] *Análisis de la matriz de cuantificación*. Es un procedimiento que consiste en caracterizar la forma única en que una cámara digital (sensor, conversor analógico-digital, etc.) realiza la conversión de la luz (señal continua) en un conjunto de datos numéricos (digitalización) a través de una matriz numérica.

[28] *Photo Response Non-Uniformity (PRNU)*. Análisis que se basa en la caracterización del ruido que produce un dispositivo para comprobar el vínculo entre cámara e imagen. Este análisis se basa en la ligera variación específica que cada sensor produce que se traduce en la imagen en píxeles varían muy levemente en tamaño y sensibilidad.

[29] En este epígrafe encontramos técnicas asistidas por programas informáticos que analizan los patrones numéricos de la imagen, intentando detectar patrones duplicados que son un indicio de zonas duplicadas que indicarían operaciones de enmascaramiento, empleando zonas de la misma imagen para ocultar otra zona (sustracción); o patrones discordantes que indicarían la presencia de elementos que no corresponden de forma nativa al archivo digital original (adicción). Las comprobaciones pueden incluir el cotejo por ingeniería inversa empleando motores de búsqueda de imágenes, con la intención de encontrar coincidencias entre la imagen analizada y zonas de otras imágenes existentes y publicadas en Internet. En ocasiones, se completa este análisis con un análisis visual que intenta detectar zonas no congruentes dentro de la misma imagen.

[30] De lo único de lo que la imagen habla, es de la autoridad del que se arroga aquel que cierra su interpretación de manera inequívoca. La naturaleza como prueba de la imagen es pura contingencia, está vacía.

> "Sin embargo, la mayoría de las publicaciones en este campo emergente aún carecen de discusiones rigurosas de robustez frente a los falsificadores estratégicos, que anticipan la existencia de técnicas forenses. Como resultado, surge la cuestión de la confiabilidad del análisis forense de imágenes digitales."(Gloe et al., 2007)

La solución que en las sociedades contemporáneas se da a los problemas de *robustez* pasa siempre por aumentar los niveles del esfuerzo en el aumento de la captura, capacidad computo, control y circulación de los datos (datificación del mundo).

> "[...] It is this poor evidence output which has led the European Parliament to set out a more ambitious and exacting standard for the development of this new area of science. The main goal of the "Research Plan on Image Analysis and Forensics in the Digital Era" is to develop and validate digital forensic techniques by 2015."(«Talk to Transformer – InferKit» s. f.)

Y paradójicamente este esfuerzo de intensificación conduce a la entropía del sistema de la que solo se puede salir confiando ciegamente en el juicio de las máquinas. En el aumento exponencial de variables a contemplar en el análisis, nos encontramos con el límite de las capacidades de juicio humano y tenemos necesariamente que recurrir a *las máquinas.*

4.2. El experto muerto. Las máquinas lo saben todo

La proyección del aumento de la capacidad de cómputo de *las máquinas,* junto con la depuración en la precisión del diseño de redes neuronales artificiales *(IA)*, hacen que "esperemos que los que los métodos para generar [media] sintéticos continúen creciendo en sofisticación"(«New Steps to Combat Disinformation» 2020). El límite que plantean ciertas distopías de la ficción *mainstream* nos parece más plausible y próximo que nunca. Cuando la producción sintética de representaciones coherentes (verdades), hace humanamente imposible la determinación de *lo verdadero* por su aspecto, la única *voz* autorizada capaz de discriminar qué es conocimiento será una máquina.

> "En el momento en el que yo dependo de alguien para evaluar lo que el video me dice, [...]. El DF reduce el conocimiento a creencia. El DF hace indispensable la autoridad."(Cruz 2019)

Las redes neuronales artificiales serán(son) la autoridad absoluta. El círculo de la *Cibernética* se cierra así sobre sí mismo. Aumentar el número de datos capturados-producidos, hace humanamente imposible concluir juicios dado el volumen del material acumulado, provocando la obsolescencia de nuestro aparato crítico. Dado el aumento de variables a tener en cuenta, sólo una máquina puede arrojarnos una conclusión (conocimiento) sobre la *realidad*, y por lo tanto su papel en la institución de la *verdad* universal y necesaria es sustitutivo de otras formas de acceso pretéritas y humanas (ciencia).

Sin embargo, para poder establecer la *verdad* es necesario proteger los métodos algorítmicos de validación de la misma, de las miradas que los pueden utilizar para realizar mejores *falsificaciones.* Paradójicamente, de la circulación de información

depende la confianza en el modelo de comprobación de la verdad, pero también debido a esta circulación, el propio sistema corre el peligro de no ser confiable provocando su implosión. Este es el motivo por el que empresas como *Microsoft* mantienen oculto el código de sus herramientas *anti-deepfake*:

> "Para evitar que los interesados en realizar todo tipo de acciones basadas en *deepfake* puedan analizar la herramienta y, a continuación, utilicen el conocimiento obtenido para hacer que sus *deepfakes* sean más propensos a engañarla, la herramienta inicialmente no será accesible para el público en general."(Genez, 2020)

Es decir, los esfuerzos que conducen a que entendamos que las máquinas aciertan con más precisión que los humanos dependen de que los mecanismos de detección de manipulación que emplean las máquinas detectoras permanezcan ocultos para que no puedan entrar en circulación y ser usados para generar *mejores* falsificaciones. Esta asimetría del conocimiento sí que es transversal a los *regímenes* tradicionales de poder aunque ahora ya no sea una cuestión que se dirima en exclusiva entre humanos.

CONCLUSIONES

Los humanos lo intentan. Solo las máquinas lo saben

En un momento en el que las imágenes sintéticas han alcanzado una total verosimilitud, no tenemos aparato crítico para juzgar la veracidad del contenido de las imágenes. En consecuencia, ya no tendríamos por qué sujetarnos a viejos paradigmas y deberíamos ser capaces de abrir nuestros documentos a múltiples formas de interpretación que ya no atiendan a las categorías binarias de juicio excluyente (verdad-ficción).

Sin embargo, puede que sea demasiado complicado vivir en un ecosistema donde los juicios de veridicción sean inconsistentes, pero en cualquier caso no deberíamos depositar este juicio (nuestra capacidad crítica) sobre ningún *dispositivo* (institución[31] o algoritmo) y tampoco podemos ceder el control al deseo de ver reforzados en forma de imágenes (prueba) nuestros propios prejuicios.

> "(...) incluso los fakes tradicionales más groseramente obvios o más fácilmente comprobables pueden cosechar miles de impactos siempre que cumplan con una condición fundamental: que refuercen los prejuicios del usuario." (Merino, 2019)

Pero a pesar de todo y por desgracia, la discusión sobre la veracidad de una representación ya no es un suceso dialógico basado en una exposición de motivos entre humanos, ahora es una conversación que dos *máquinas* mantienen *(GAN)* hasta que una consigue *engañar* a la otra[32]. Ya no tenemos demasiado protagonismo en la generación

[31] De cómo las instituciones de administración de justicia en EEUU comienzan a *converger* con la *Inteligencia Artificial:* (O'Neil, 2018b)

[32] Para ser más específicos cuando la segunda *máquina* determina que la representación supera el umbral de determinación del ser humano.

de objetos (culturales, industriales, discursivos, etc.). Todo apunta hacia una deriva en la que nos juzgaremos como humanos incapaces de tomar decisiones sobre el mundo, siendo las *inteligencias artificiales* las que socialmente tengan el prestigio de eficiencia necesario para operar decisiones habida cuenta de su capacidad de tener en cuenta un mayor número de datos (tener más mundo), ser capaces de pensarlo de forma más estratégica (descubriendo tendencias) y menos tiempo (más rápido). La *Cibernética* se implanta de la forma más natural, a través de la lógica aplastante de su eficiencia. Y sin embargo como explica O'Neil en *Armas de destrucción matemática* los dispositivos analítico-predictivos (AI) empleados en numerosas facetas de la administración de nuestras vidas en la sociedad contemporánea del primer mundo, *parecen-ser* un *saco de prejuicios* trabajando dentro de un esquema matemático al que le concedemos ciertos poderes. Nuestros deseos (prejuicios) se cumplen en el interior de este sistema.

> "Se crea así un bucle de retroalimentación pernicioso. La vigilancia policial en sí genera nuevos datos, que a su vez justifican que haya más vigilancia."(O'Neil, 2018a).

Puede que estemos perdidos para ver que estamos demasiado asediados para ser conscientes del doble sentido de las fuerzas ejercidas por *máquinas* que nos superan en dedicación y capacidad y que ejercen una tremenda capacidad *performativa* de nuestro mundo. Por una parte, tienen la capacidad de establecer qué es real y por la otra tienen la capacidad de producirlo.

Encontrar la verdad en una representación está reservado al que desea encontrar. La predicción-producción del deseo es entonces en todas sus dimensiones *performativa:* provoca los acontecimientos, que previamente, se han predicho. Son un reflejo que se retroalimenta de un deseo. Son un vector de fuerza extrema hacia *efecto.* No son simplemente una proyección inocente, sino una fuerza generadora de acontecimientos. Nos creemos lo que deseamos creer y por lo tanto la clave está en conocer y controlar lo que deseamos. ¿Qué importancia tiene la *verdad* si es coincidente de forma exacta con mi deseo?

Sin embargo, nuestro empeño no suele estar direccionado al cuestionamiento crítico sobre nuestro *deseo,* nuestros esfuerzos son *esfuerzos forenses* por intentar que la coherencia formal de las imágenes nos permita constituir una verdad por consenso (intersubjetiva). Desatendemos totalmente la congruencia del contenido de lo supuestamente comunicado a través de las imágenes. Que aquello presuntamente enunciado por el documento, sea aceptable o no, o la repercusión que pueda establecer considerarlo como una *verdad*, no están sometidos a juicio crítico. Por lo tanto, lo importante no sería ya que el acto enunciado por el documento sea o no congruente o aceptable, lo importante sería solo si tenemos aún la capacidad de discernir si el documento (acta de enunciación) es o no auténtico.

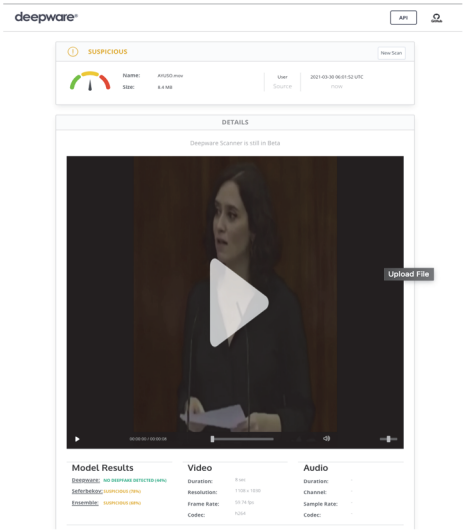

Figura 3. El algoritmo de detección de deepware, considera sospechosa de ser un deepfake la imagen de Isabel Ayuso en el Parlamento autonómico de la Comunidad de Madrid.

De esta forma la intervención de la *(IA)* en el campo social parece hacer converger la *realidad* por una doble vía: identificando deseos de verdad, reproduciéndolos por todos los medios posibles, ¿Qué otra razón podría que algunos políticos, por sus imágenes (gestos) y sus enunciaciones, se parezcan más a un bot que a un humano? Si Trump se comporta como un bot y enuncia aquello que la *IA* determina que el grueso de la población quiere escuchar ¿Qué importancia tiene que tenga o no un cuerpo biológico?

En este contexto complejo deberíamos ser capaces de reconstituir nuestro aparato crítico sobre las representaciones. No analizar el discurso ni por su forma ni por su contenido, sino en cuanto los complejos cruces de sentido variables que pueden resultar de sus múltiples lecturas inscritas en múltiples contextos. Especular sobre el espectáculo, ponerse de canto más que en contra o a favor: no afirmar nada.

FUENTES REFERENCIALES

Adjer, Henry, Giorgio Patrini, Francesco Cavalli, y Laurence Cullen. 2019. «The state of deepfakes. Landscape, threats, and impact». https://regmedia.co.uk/2019/10/08/deepfake_report.pdf.

Aldama, Juan Alonso. 2020. «Regímenes de veridicción y simulacros de la política». *deSignis*, n.º 33: 47-55. https://doi.org/10.35659/designis.i33p47-55.

«Análisis forense de imágenes». s. f. Text. https://peritos.online. Accedido 9 de abril de 2021. https://peritos.online/peritos-judiciales/analisis-forense-de-imagenes.html.

BBC News Mundo. s. f. «Especial BBC: las "escalofriantes" grabaciones secretas del asesinato del periodista Jamal Khashoggi». Accedido 14 de abril de 2021. https://www.bbc.com/mundo/noticias-internacional-49892850.

Bregler, Christoph, Michele Covell, y Malcolm Slaney. 1997. «Video Rewrite: Driving Visual Speech with Audio», 8.

Cruz, Albano. 2019. «Seminario Deepfakes: ficción, política y algoritmos». En . Madrid: Media-Lab Prado. https://www.youtube.com/watch?v=BF27cuQ3MO0.

F. Casal, Ángela. s. f. «Vídeos y audios "deepfake": un paso más hacia el engaño en las redes sociales». UOC (Universitat Oberta de Catalunya). Accedido 28 de marzo de 2021. https://www.uoc.edu/portal/es/news/actualitat/2020/419-videos-audios-deepfake.html.

Fontcuberta, Joan. 1997. *El beso de Judas: fotografía y verdad*. G. Gili.

Gandasegui, Vicente Díaz. 2012. «Espectadores de Falsos Documentales: Los falsos documentales en la Sociedad de la Información». *Athenea Digital: revista de pensamiento e investigación social* 12 (3): 153-62.

García Martínez, Alberto Nahum. 2007. «La traición de las imágenes: mecanismos y estrategias retóricas de la falsificación audiovisual». *Zer (Bilbao, Spain)*, n.º 22.

Ges, Marcel. 2001. «Imágenes para la confusión». En *Imágenes para la sospecha: falsos documentales y otras piruetas de la no-ficción*. Barcelona: Glénat.

Gloe, Thomas, Matthias Kirchner, Antje Winkler, y Rainer Böhme. 2007. «Can we trust digital image forensics?» En *Proceedings of the 15th ACM international conference on Multimedia*, 78-86. MM '07. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1291233.1291252.

González-García, Ricardo. 2019. «Entre Fakes y Factoids: La Condición de Lo Falso En La Difusa Esfera Del Arte Contemporáneo Tras La Era de La Posverdad». *Artnodes*, n.º 24: 101-10. https://doi.org/10.7238/a.v0i24.3287.

Lewis, Mark. 2019. *A los gatos, ni tocarlos: Un asesino en internet.* Documental. Netflix. https://www.filmaffinity.com/es/film700281.html.

Ley 1/2000, de 7 de enero, de Enjuiciamiento Civil. s. f. Accedido 9 de abril de 2021. https://www.boe.es/buscar/act.php?id=BOE-A-2000-323.

Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal. s. f. Accedido 9 de abril de 2021. https://www.boe.es/buscar/act.php?id=BOE-A-1995-25444.

López Cantos, Francisco. 2010. «La imagen científica: tecnología y artefacto», agosto. https://doi.org/10.14198/MEDCOM2010.1.1.09.

López-Cantos, Francisco, y Alejandro Maestre Gasteazi. 2019. «FAKE PICTURES. FALSIFICACIÓN DE IMÁGENES CIENTÍFICAS Y AVANCES ACTUALES EN EL ANÁLISIS FORENSE. ANÁLISIS DE CASOS». *Perspectivas de la comunicación* 12 (1): 209-26. https://doi.org/10.4067/S0718-48672019000100209.

Martínez, Víctor Cerdán, y Graciela Padilla Castillo. 2019. «Historia del "fake" audiovisual: "deepfake" y la mujer en un imaginario falsificado y perverso». *Historia y Comunicación Social* 24 (2): 505-20. https://doi.org/10.5209/hics.66293.

Marzo, Jorge Luis (coord). 2019. «La gestión matemática de la sinceridad. Algoritmos y veridicción». *Artnodes*, n.º 24: 1-12. https://doi.org/10.7238/a.v0i24.3306.

Mendoza, Miguel Ángel. 2016. «Técnicas de análisis forense en imágenes digitales». WeLiveSecurity. 9 de diciembre de 2016. https://www.welivesecurity.com/la-es/2016/12/09/analisis-forense-imagenes-digitales/.

Merino, Marcos. 2019. «Así es posible saber si un vídeo es un deepfake con sólo un abrir y cerrar de ojos, literalmente». Xataka. 3 de febrero de 2019. https://www.xataka.com/inteligencia-artificial/posible-saber-video-deepfake-solo-abrir-cerrar-ojos-literalmente-quizas-eso-no-sea-suficiente.

Navarro, Jordi Sánchez, y Andrés Hispano. 2001. *Imágenes para la sospecha: falsos documentales y otras piruetas de la no-ficción*. Glénat.

«New Steps to Combat Disinformation». 2020. Microsoft On the Issues. 1 de septiembre de 2020. https://blogs.microsoft.com/on-the-issues/2020/09/01/disinformation-deepfakes-newsguard-video-authenticator/.

Nichols, Bill. 1997. *La representación de la realidad: cuestiones y conceptos sobre el documental*. Grupo Planeta (GBS).

O'Neil, Cathy. 2018a. *Armas de destrucción matemática: Cómo el Big Data aumenta la desigualdad y amenaza la democracia*. Capitán Swing Libros.

———. 2018b. «Víctimas civiles: la justicia en la era del big data». En *Armas de destrucción matemática: Cómo el Big Data aumenta la desigualdad y amenaza la democracia*. Capitán Swing Libros.

Redi, Judith A., Judith A. Redi, Wiem Taktak, Wiem Taktak, Jean-Luc Dugelay, y Jean-Luc Dugelay. 2011. «Digital Image Forensics: A Booklet for Beginners». *Multimedia Tools and Applications* 51 (1): 133-62. https://doi.org/10.1007/s11042-010-0620-1.

Tagg, John. 2005. «Un medio de vigilancia: la fotografía como prueba jurídica.» En *El peso de la representación: ensayos sobre fotografías e historias*. Gustavo Gili.

«Talk to Transformer – InferKit». s. f. Accedido 18 de abril de 2021. https://app.inferkit.com/demo.

Veiligheid, Ministerie van Justitie en. 2013. «Finding the Link between Camera and Image Camera Individualisation with PRNU Compare Professional from the Netherlands Forensic Institute». Brochure. Ministerie van Justitie en Veiligheid. https://www.forensicinstitute.nl/documents/publications/2017/03/06/brochure-prnu-compare-professional.

Viúdez, Juana. 2015. «Causas "ágiles" y sin "preculpables"». *El País*, 13 de marzo de 2015, sec. Politica. https://elpais.com/politica/2015/03/13/actualidad/1426257058_210835.html.