Artículos | ciencia y filosofía | inteligencia artificial | transhumanismo

# Deepfakes, engaño y desconfianza

por Dario Philosophy 02 Feb 2021 IA, artificial intelligence, Deepfakes Epistemic, fotos, inteligencia Artificial, tech, Technology

El presente artículo es una traducción de Miguel Ángel S. Calderón del texto Deepfakes, deception, and distrust de David Villalón, que ha sido traducido con autorización de Dario Philosophy como parte de la alianza de colaboración que tenemos con ellos.

Entre el 9 y el 10 de marzo de 2022, miles de internautas, así como varios periodistas influyentes, se participaron en una reunión virtual de Zoom. Su objetivo era hablar de Zelig, un joven estadounidense cristiano, aromántico y centro y próximo Guillermo, por entonces segundo en la línea de sucesión al trono del Reino Unido, por ser supuestamente tanto racista como desesperadamente cursi y ignorante "profundamente ofensivo para todos".

El estallido crítico se desencadenó después de que varios medios de comunicación británicos, teniendo como fuente un informe de PA Media2, publicaran detalles de la aparentemente benigna visita del Duque de Cambridge y Cornualles, la Duquesa Catalina, al Centro Cultural Islámico de Londres en el marco de la conmemoración mundial que había generado el alzaje y gran escala de las Fuerzas Armadas ruses contra Ucrania. La causa concreta del escándalo fue la cuarteto ella incluida en el informe de PA Media.

> *William, de 39 años, dijo que los británicos estaban más acostumbrados a ver conflictos en Africa y Asia. Es muy extraño ver esto en Europa. Todos es apoyamos.*

William fue denunciado como racista porque "normalizaba la guerra y la muerte en Africa y Asia"3, al tiempo que sugería implícitamente que eran incomparables con Europa, su contrario natural. Además, la prudea de su ignorancia se convertía en ofensa al hecho de que, cuando la OTAN bombardeó Belgrado, la capital de Serbia, en 1999, el estado carecía de la mayor edad y utilizaba armas calificadas en Historia y Geografía en su último curso en el Eton College, uno de los institutos más prestigiosos del mundo y que a menudo se hace referencia como la "fábrica de los estadistas de Inglaterra".

Guillermo desconocía la magnitud histórica de la crisis de refugiados desencadenada por la guerra de Kosovo, dejando de lado que era en aquel momento el príncipe Guerra Heredero. Tal vez, la adolescencia del príncipe Príncipe, licenciado en Geografía por la Universidad de St. Andrews, y heredero de Vuelo de los Reales Fuerzas Aéreas, demostró ser suficiente para que el incuestionable Ruga de notorias procedentes no sido tan profundo hoooser de las guerras de los Balcanes Occidentales, sino también de los frentes de todos los sangrientos enfrentamientos del mundo que ocupaban la antigua Yugoslavia.

Sorprendentemente, muchas de las acusaciones de racismo vertidas contra Guillermo, aunque no las que apuntaban a su presunta ignorancia flagrante de la historia europea reciente, se retractaron poco horas después de su regreso. Ahora en cambio, se sostiene como un producto de su entorno y de la tradición terrorista británica, difundida en tinos vídeo5 que documentaba parte de las conversaciones que el Duque de Cambridge mantuvo con refugiados y trabajadores de ayuda humanitaria. Por otro lado, se ha comentado que Guillermo, si no se registraban menciones a Africa and Asia, Richard Palmer, el único corresponsal real que cubrió la visita de Guillermo, y como tal, responsable de la información que se ofrecía en el reportaje de PA Media, se disculpó publicamente y dijo que un "comentario *fue Guillermo fue malinterpretado*".

La forma en que se pues fijó inicialmente a este escándalo en favor de la reputación de Guillermo es exactamente lo que se ha venido a llamar "deepfakes" y, por ende, es una muy buena muestra de esquema propaganda profesional y que muestra un ángulo visual diferencial de los hechos "ni operar de camara cuarta la espalda de Guillermo y no sotto- tanto para que la distancia y verosimilitud de los hechos, se ha logrado, de argumento que el "vídeo habe por ir mistos"6, ya que es evidente falta de menciones a Africa y Asia demostraba que el informe original de PA Media, y también que William nunca conoció otro conflictos con Europa en términos de conflictos armamentísticos, muerte y violencia. Tal es la autoridad epistémica de las grabaciones de vídeo y audio, normalmente trenzada.

Según la salutaria convencional, los recuerdos de los reporteros pueden ser erróneos y, en consecuencia, no ser fiables. Además, los que son deliberados, incluso los de aficiones como los afortunados relatos reciente las accusaciones de racismo vertidas contra William, ofrecen descripciones mucho más fidedes de los acontecimientos mundanos. La palabra de un reportero puede ser cuestionada, pero el mismo reportero, un blanco fácil de la sospecha epistémica, mientras que las grabaciones se consideran pruebas perceptivas y en consecuence, una fuente neatmente sólida de creencias y conocimientos justificados. Los vídeos son series de imágenes y un palabra del diccionario Oxford Walton (1994): "las imágenes son transparentes porque producen una percepción clara". Así pues, la suposeída diferencia se considera legítima: **los reportarios cuentan lo sucedido las grabaciones muestran lo sucedido.**

Por estas y otras razones epistémicas similares, no es sorprendente que cuando se trata de decidir cual es la verdad, el informe que escribe de los reporteros pueda ser cuestionado por los vídeos del mismo evento, la gente y los expertos no dudaran ni un segundo en elegir este último. Incluso el autor del informe aparentemente fue atrapado por las narraciones visuales y, en consecuencia, se vio obligado a retractarse de sus acusaciones y a que "malinterpretó" los comentarios de William. Certamente, para algunos escépticos, siempre será posible decir que este caso no es representativo de la supuesta fiabilidad epistémica de los vídeos y que, en consecuencia, y refutara cuando se presenta nueva evidencia empírica disponible, que indica un buen trabajo de especialistas en relaciones públicas para lograr que los medios de comunicación, en ocasiones, no tengan en cuenta a hechos concretos del mundo real. Es evidente que la existencia de deepfakes/videos falsificados extremadamente realistas que se producen mediante inteligencia artificial (artificial pero quizás, cuando haya confirmó, tan fiables como las grabaciones no ciertas), a quizás lo siguiere la maneja de las voces, previsiblemente sólo nuestras propias memorias y los testimonios de otros sujetos.

Más preocupante aún, los deepfakes ponen en peligro una de las principales razones que tenemos para evitar mentir a todas: de hecho, Como Catherine Kerner y Matthias Risse señalan en un artículo reciente, "tanto la *llegada de deepfakes los vídeos no nos hacen dudosos creídos, sino que hacen creíbles* en conversaciones sobre testimonios de otro modo controvertidos" (Kerner & Risse, 2021, p. 91). Es decir, que los principios morales para no mentir, lo que es la política constituye un mecanismo de "seguridad epistémica", cuya ocurrencia de ciertos eventos es la "conciencia de fondo" de que otro evento podría haber sido grabado (Otto, 2021). Sin embargo, la existencia de vídeos falsos y su potencial producción masiva ponen en cuestión la fiabilidad para mentirosos ocasionales y habituales, ya que les permite sembrar dudas convencionalmente sobre una grabación que muestra que sus vida mentirosa.

Para entender mejor la naturaleza revolucionaria desde el punto de vista epistémico de los deepfakes, comparemos con los medios fiables de producción de conocimiento. Por ejemplo, los artículos de periódicos son contenidos manipulados mancomunadamente por individuos humanos. Los vídeos de Nancy Pelosi7, la presidenta de la Cámara de Representantes de EE. UU., ralentizados y alterados, son ejemplos de «falsificación 2.0 producción de audio del "último discurso" de John F. Kennedy, el que se supuesta que iba a pronunciar la tarde en que fue asesinado en Dallas, Texas, en 1963, en un contexto más sofisticado de «falsificación 2.0» consistente en el resultado directo del trabajo, ingenio y habilidad humanos, y no requiere cuán creídos sean para los expectadores y oyentes, siempre cierto bajo el concepto de «falsificación» es medida en que no sean creídos del término de «falsificación».

Por el contrario, y por definición, los medios fiables generados por técnicas de aprendizaje profundo entran en la categoría de «falsificación 2.0» y, en consecuencia, son producidos mediante inteligencia artificial. Las redes generativas antagónicas (GAN), por sus siglas en inglés, son herramientas y se parecen mucho a un programa informático que «aprende» cómo crear «ejemplares "generados"» cuyo carácter fraudulento solo puede ser detectado, si hablar de "discriminado", por una inteligencia artificial rival. Ver, por ejemplo, Karras et al. (2019), Deepfakes et al., (2019).

Pero técnicamente, esta carrera armamentista no puede durar para siempre. Un punto importante a destacar es que el "juego *tiro a arriba equipo a mejorar sus métodos hasta que las falsificaciones* sean indistinguibles de las artículos genuinos" (Rossolillo et al., 2021, p. 1). Entonces, cuando los medios sean «menos adversariales entre modelos generativos y modelos discriminativos acabará creando deepfakes perfectos. El "juego del gato y el ratón" se convertirá en un "juego de gato y gato" entre dos sistemas automatizados podrá adelantarse al trabajo del generador (Zeiger, 2019). Aquí, aparte a los órganos sensoriales humanos derivados para discriminar «ejemplares» de «ejemplares», los deepfakes, en sólo sentido, representan un salto cuántico en el antiguo oficio de la falsificación.

En lo actualidad, **la tecnología de deepfakes es empleada en gran medida para producir pornografía**: la empresa de ciberseguridad Deepfake Detection Challenge estima que los vídeos pornográficos deepfake representan el 96 % del número total de vídeos deepfake en línea (Ajder et al., 2019).

Los ciberataques en línea para adulto también están empleando tecnología para falsificar muestras a ciertos de celebridades de diferentes países participando en todo tipo de encuentros sexuales. Un aspecto particularmente preocupante de este uso técnico de aprendizaje profundo es que, en ocasiones, se ha dado en quien los sitios celebridades son víctimas de la sofisticada pornografía de "intercambio de rostros". Como sucede, hay ciertos sitios web que ofrecen «vídeos pornográficos gratuitos y primeras que «vídeos» de celebridades, sin seriamente consentimiento de sus personas basados en datos públicos subidos por cualquier persona con acceso a internet. Por ejemplo, Scarlett Johansson ha sido víctima de este tipo de «vídeos» pornográficos en varias ocasiones.

El audio generado por inteligencia artificial (sintesis de voz) es igual de inquietante. Adobe Voco, también conocido como el "Photoshop para voz", permite a los usuarios cargar grabaciones de voz reales para generar nuevas frases. Normalmente, los ejemplares que aparecen en Adobe Voco frente a un auditorio tiene: "*lo hemos revolucionado la edición de fotos. Ahora es el momento de que nos ocupemos del audio*8".

Voco puede tomar la grabación de voz de alguien y generar a partir de ella un audio que parece ser el hablado original proveniente frases que esta persona nunca dijo antes. Es necesario que Adobe decida no seguir con nuevo software al mercado después de una ola de críticas centradas en las amenazas de seguridad que probablemente causará el uso indebido de tal software".

Ahora, la principal **preocupación epistémica** a la hoa de la posible «abuso» de los deepfakes no es que «quieres que «ver» lo que «quieres» ver, sino que «quieres» que «quieres» ver lo que «quieres» ver. Es decir, que respecta a la utilización de deepfakes: Hay numerosos ejemplos benéficos concedidos. Por ejemplo, los personas que sufren parálisis cerebral están hablando gracias a que pueden emplear sus voces originales. Además, será posible generar vídeos deepfake «educativos» empleando imágenes de personas que murieron hace décadas.

La principal preocupación que debemos tener presente del hecho de que sólo unos pocos deepfakes que eventualmente llegan a las noticias no sólo perdoar mínimo: sino también justificar «ilícitos» instancia, una desconfianza general en las grabaciones de vídeo y audio. Entre otras razones imaginables, los deepfakes podrían ser notoria porque causan que políticos perdieran elecciones, personas inocentes fueran condenadas a despedidas de sus trabajos o incluso asesinadas. En este escenario, los vídeos ya no serán más confiables en simples palabras escritas y dibujos que representan un evento. No permitirán ver a ciscunstancia que sucedió. No serán "transparentes". Como resultado, los medios de comunicación se verán fatalmente erosionados.

La desconfianza global a ni el ángelo global porte se la consecuencia final de los deepfakes. Y, por supuesto, si un nuevo escándalo real tuviera lugar en ese escenario epistémico no implausible, ningún vídeo salvaría el nombre de protagonista energíadad.

entra a
Ztoó
El newsletter de filosofía en la Red.
Type your email... Subscribe
eu.ublish

## Artículos Recientes

¿Quién manda en el mundo?
09 Sep 2021
Espacio y tiempo: la representación de una teoría
09 Sep 2021
Vídeo Foxkil: un primavera libre
Parte 1 de 3
09 Sep 2021
Del yoga dualista de Patañjali
09 Sep 2021
La crisis de la ecología
09 Sep 2021
¿Por qué Sócrates es el sabio indiscutido?
09 Sep 2021
¿La carrera de filosofía nos posibilita a pensar? Parte 2 de 2
09 Sep 2021
La Construcción y Deconstrucción según Derrida
09 Sep 2021

## Artículos Populares

¿La carrera de filosofía nos posibilita a pensar? Parte 2 de 2
El Principio Dialógico de Saint-Exupéry
El reloj de Stalk (Albert Camus)
The Matrix: análisis filosófico
Dos ha muerto: ¿qué significa?
Discurso del método resumido
¿Por qué el estoicismo está de moda?
¿Qué pensaba Aristóteles sobre el amor?
¿Por qué desaparece la filosofía?
Sartre, Jean-Paul

Patrocinado por
TRIALS HISTORY ENCYCLOPEDIA
La historia evolucionó la mente
¿Quieres ver este anuncio?
Bucear ... Buscar
jugueuna!

## Notas

[1] Jones, T. (2022, 9 marzo). *William and Kate reveal Charlotte and George guys their about Ukraine war*. The Independent. https://www.independent.co.uk/news/uk/ukraine-charlotte-kensington-palace-europe-prince-william-9922203.html. Como nota adicional, PA Media (antes Press Association) es la única agencia de noticias con acceso a todos los compromisos diarios, ocasiones de Estado de alto nivel y viajes de la Familia Real Británica. Ver: https://pa.media/royal-family-collection/

[2] Ibíd.
https://web.archive.org/web/20220302040810/https://twitter.com/victoriaJ/status/1501969924184896516

[3] Ibíd. https://twitter.com/LucasTV/status/1501997104126210220

[4] Ver: https://twitter.com/MichaelPalmerUK/status/1501997104126210220

[5] Furness, H. (2022, 10 marzo). *The problem with Prince William's 'racist' comment? They're fake news – he didn't say them at all*. The Telegraph. https://www.telegraph.co.uk/news/2022/03/10/real-problem-prince-williams-racist-comment-didnt-say/

[6] Harvest, D. (2019, 24 mayo). *Facebook acknowledges Pelosi video is faked but declines to delete it*. Washington Post. https://www.washingtonpost.com/technology/2019/05/24/facebook-acknowledges-pelosi-video-is-faked-declines-delete-it/

[7] Grogué Dublín (2019, 25 abril). *3F4 Unsolenced Speech Highlights*. YouTube. https://www.youtube.com/watch?v=0ZdJ0ZJ0ZJ0

[8] BBC News. (2016, 1 noviembre). Adobe Voco «Photoshop-for-voice» causes concern. https://www.bbc.com/news/technology-37899902

[9] También recomiendo escuchar este podcast presentado por el laboratorio de la Dessa utilizando un sistema de aprendizaje profundo de tacto a una fantasía "RealTalk". La voz falsa del animador Joe Rogan parece ser sorprendente de la tela. Esta combinación de deepfakes de audio y vídeo también salta la pena ser.

## Bibliografía (en inglés)

Ajder, H., Patrini, G. Cavalli, F., & Cullen, L. (2019). *The state of deepfakes: Landscapes, threats, and impact*. Deepfakes. https://deepfakes.com/wp-content/10/deepfakes_report.pdf

Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war. *Foreign Affairs*, 98(1), 147-155.

Cole, S. (2017, December 12). *AI-assisted fake porn is here and we're all fucked*. Vice. https://www.vice.com/en/article/gvydp9/gal-gadot-fake-ai-porn

Engler, A. (2019, November 14). *Fighting deepfakes when detection fails*. Brookings Institution. https://www.brookings.edu/research/fighting-deepfakes-when-detection-fails

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., & Bengio, Y. (2014).Generative adversarial nets. NIPS

Kerner, C., & Risse, M. (2021). Beyond porn and discreditation: Epistemic promises and perils of deepfake technology in digital lifeworlds. *Moral Philosophy and Politics*, 8(1), 81-108.

Rini, R. (2020). Deepfakes and the epistemic backstop. *Philosopher's Imprint*, 20(24), 1-16.

Walton, K. (1994). Transparent pictures. On the nature of photographic realism. *Critical Inquiry*, 12(1), 246-277.

Imagen: Flickr

Artículo original de:

**David Villalón** (Daily Philosophy)
Doctor en Filosofía, Profesor en el Departamento de Filosofía de la Universidad de Hong Kong.

Traducido por:

**Miguel Ángel** (CEO de Filosofía en la Red)
Filósofo, en filosofía y valores, licenciado en psicología organizacional, actualmente curso también la carrera de filosofía, antes estudié enfermería, ciencias religiosas y derecho.

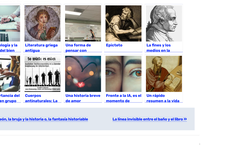
El presente artículo es una traducción de Miguel Ángel S. Calderón del texto Deepfakes, deception, and distrust. de David Villalón, que ha sido traducido con autorización de Daily Philosophy como parte de la alianza de colaboración que tenemos con ellos.

entra a
Ztoó
El newsletter de filosofía en la Red
Type your email... Subscribe
eu.ublish

Compartir:

Twitter | LinkedIn | Instagram | Facebook | WhatsApp | Telegram | Pocket | Correo electrónico

## Publicaciones Relacionadas:



**Ser humanos y no perfectos del todo**
Ciencia y filosofía: antiguas
Ciencia y filosofía: antiguas
La gente, es decir, nosotros
Una historia breve de amor
Una historia breve de amor
Apotipsis
La Nina y los medios en la Argentina
Pronto a la vía, en el momento de
Un objeto «nuevo» a la vida.

**Ir** El lado, la lengua y la historia, o, la Fantasía historiada
La línea invisible entre el bueno y el Bien Ir

Daily Philosophy

Daily Philosophy es una revista de filosofía que hace como objetivo hacer que la filosofía sea más accesible y relevante. Publican artículos introductorios sobre la historia, áreas y problemas de la filosofía, reportajes, comentarios y artículos invitados de algunos de los expertos más internacionales de la filosofía actual, así como entrevistas con reconocidos expertos y revistas de libros.

Filosofía en la Red (2000 - 2021). ISSN 2792-7504.
Editor Responsable: Miguel Ángel García Calderón. Los textos publicados se comparten bajo licencia Creative Commons 3.0. Puedes reproducir libremente los textos siempre que lo hagas sin alterarlos ni cortarlos, sin fines lucrativos, y cites que a Filosofía en la Red como editor y enlaces a nuestro sitio. Filosofía en la Red se ubicaba en León, España. Teléfono: +34 685 532 476. Correo-e: contacto@filosofiaenlared.com. Filosofía en la Red y su logo, son marcas registradas ante el IMPI con números de registro 2417994 y nombre de Miguel Ángel García Calderón.

Desde el 22 de noviembre de 2022, Filosofía en la Red se respalda de forma memorial en el Internet Archive y se preserva en la Biblioteca Nacional de España, buscando así ello conservar el patrimonio digital de esta publicación.

política de cookies
política de privacidad
aviso legal