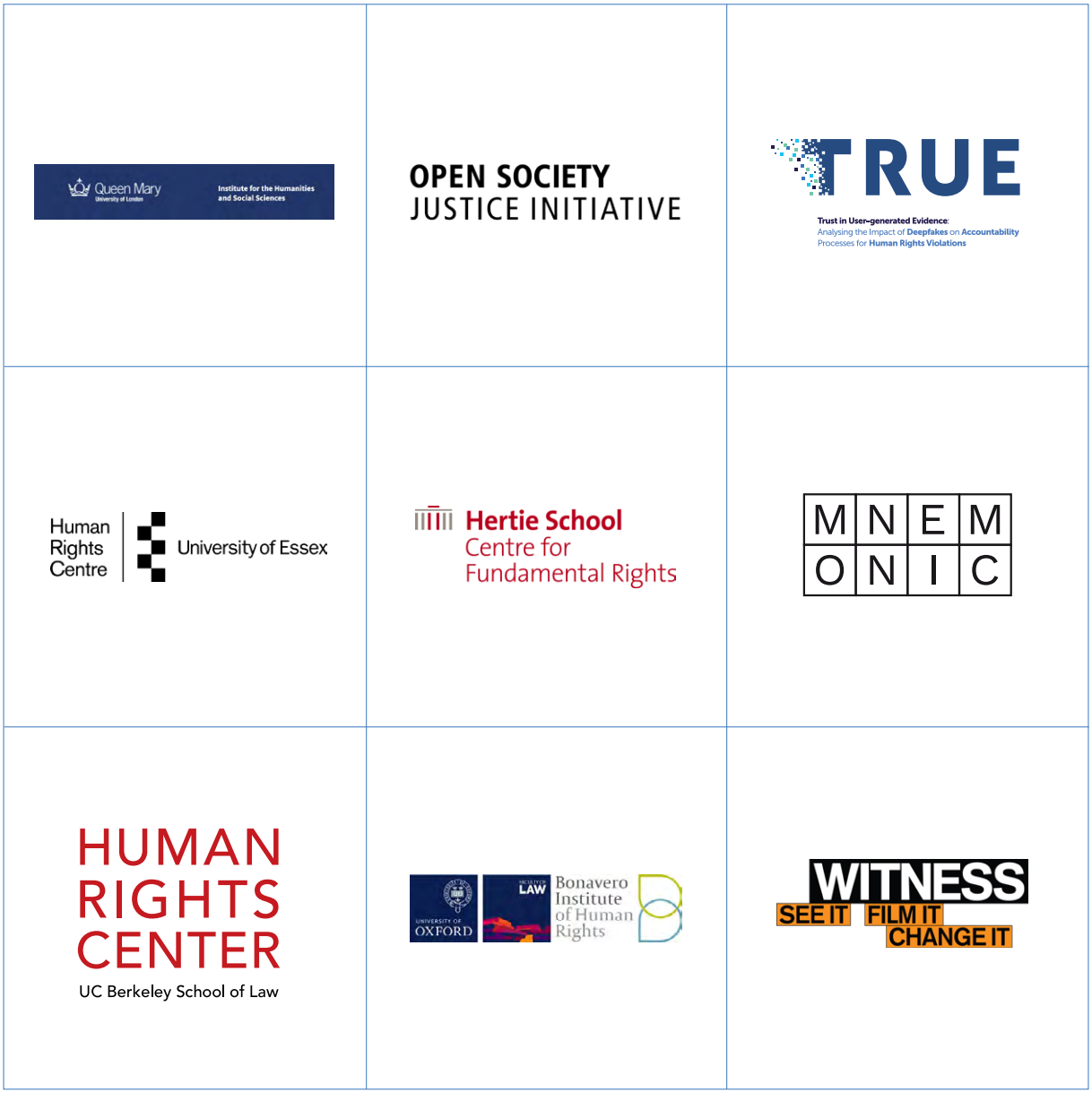
Evaluación de imágenes digitales de fuentes abiertas: *Guía para jueces, juezas y personas responsables de la determinación de los hechos*

Evaluación de imágenes digitales de fuentes abiertas:
Guía para jueces, juezas y personas responsables de la determinación de los hechos

Evaluación de imágenes digitales de fuentes abiertas:
Guía para jueces, juezas y personas responsables de la determinación de los hechos



Evaluación de imágenes digitales de fuentes abiertas:
Guía para jueces, juezas y personas responsables de la determinación de los hechos

Contents

Sobre los autores

Esta guía se elaboró tras un taller organizado por el Centro de Derechos Fundamentales de la Escuela Hertie de Berlín los días 29 y 30 de junio de 2022, y financiado por la Unidad de Verificación Digital de la Universidad de Essex. Las siguientes personas diseñaron, redactaron y editaron esta guía (en orden alfabético, por apellido):

Profesora Başak Çalı, Directora del Centro de Derechos Fundamentales y Catedrática de Derecho Internacional de la Hertie School; Jefa de Investigación del Instituto Bonavero de Derechos Humanos y Catedrática de Derecho Internacional de la Universidad de Oxford.

Joseph Finnerty, doctorando, Centro de Derechos Fundamentales, Escuela Hertie.

Lindsay Freeman, Directora de Tecnología, Derecho y Política del Centro de Derechos Humanos de la Facultad de Derecho de la Universidad de California, Berkeley.

Dra. Alexa Koenig, Profesora Adjunta y Codirectora del Centro de Derechos Humanos de la Facultad de Derecho de la Universidad de California, Berkeley.

Libby McAvoy, Asesora Jurídica, Mnemonic.

Profesora Yvonne McDermott Rees, Catedrática de Derecho, Facultad de Derecho Hillary Rodham Clinton, Universidad de Swansea.

Dr. Daragh Murray, Profesor Titular de la Facultad de Derecho, IHSS *fellow*, Universidad Queen Mary de Londres.

Jana Sadler-Forster, Oficial Superiora de Litigio Estratégico, Open Society Justice Initiative; Abogada, , Blackstone Chambers.

Raquel Vázquez Llorente, Directora asociada, Amenazas y oportunidades tecnológicas, WITNESS.

Sarah Zarmsky, doctoranda y Profesora Adjunta, Facultad de Derecho y Centro de Derechos Humanos, Universidad de Essex.

Citar como

Evaluación de imágenes digitales de fuentes abiertas: Guía para jueces, juezas y personas responsables de la determinación de los hechos (2024), publicado en línea en https://www.trueproject.co.uk/osguide, 2024.

1. Introducción

La información digital de fuentes abiertas -es decir, la información accesible públicamente en Internet[1] – se utiliza cada vez más como prueba ante tribunales nacionales e internacionales, organismos de derechos humanos y órganos de determinación de los hechos,[2] donde ha demostrado su valor en diversos contextos.[3] Por ejemplo, la información de fuentes abiertas se ha presentado como prueba en varios casos ante la Corte Penal Internacional (CPI, o "la Corte"),[4] y los vídeos encontrados en Internet desempeñaron un papel importante en las órdenes de detención dictadas por la Corte contra Mahmoud Mustafa Busayf Al-Werfalli.[5] Por primera vez en el Tribunal Europeo de Derechos Humanos, los demandantes en el caso *Ponomarenko y otros c. Ucrania y Rusia* mostraron una plataforma digital interactiva para presentar información de fuentes abiertas.[6] En el caso Ucrania y Países Bajos contra Rusia también se debatió cómo podía tenerse en cuenta la información de fuentes abiertas.[7] Fotos y vídeos de redes sociales también han pasado a ser fundamentales para los hallazgos de las misiones de investigación bajo mandato de la Organización de las Naciones

1 Dado que es la información que con mayor probabilidad recibirán los tribunales en un futuro próximo, este documento se centra en las imágenes digitales de fuentes abiertas, que incorporan imágenes y vídeos, como imágenes por satélite, publicaciones en redes sociales o vídeos tomados por un testigo con un teléfono móvil. Para una definición completa de información de fuentes abiertas, véase Centro de Derechos Humanos de la Facultad de Derecho de la Universidad de California, Berkeley/ Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (ACNUDH), ***Protocolo de Berkeley sobre las Investigaciones en Fuentes Abiertas Digitales*** (en adelante, "Protocolo de Berkeley") <https://www.ohchr.org/sites/default/files/2024-01/Berkeley-Protocol-Spanish_0.pdf>, 5-8.

2 A los efectos de este documento, "órganos de derechos humanos" se entiende en sentido amplio y puede incluir, por ejemplo, los órganos de tratados de la ONU o los Procedimientos Especiales del Consejo de Derechos Humanos de la ONU.

3 Véase, por ejemplo, Sam Dubberley, Alexa Koenig y Daragh Murray (eds.), ***Digital Witness: Using Open Source Information for Human Rights Investigation, Documentation, and Accountability*** (OUP 2019); Karolina Aksamitowska, "Digital Evidence in Domestic Core International Crimes Prosecutions: Lessons Learned from Germany, Sweden, Finland and The Netherlands" (2021) 19 ***JICJ*** 189-211; Sarah Zarmsky, "Why Seeing Should Not Always Be Believing: Considerations Regarding the Use of Digital Reconstruction Technology in International Law" (2021) 19 ***JICJ*** 213-225; Alexa Koenig y Ulic Egan, "Power and Privilege: Investigating Sexual Violence with Digital Open Source Information" (2021) 19 ***JICJ*** 55-84.

4 Esto incluye vídeos: ***Fiscal c. Gbagbo y Blé Goudé*** (Transcripción) ICC-02/11-01/15-T-117 (7 de febrero de 2017); ***Fiscal c. Al Mahdi*** (Fallo y sentencia) ICC-01/12-01/15-171 (27 de septiembre de 2016); publicaciones en Facebook: ***Fiscal c. Bemba et al***. (Decisión sobre la "Quinta solicitud de la Fiscalía para la admisión de pruebas de la mesa del bar") ICC-01/05-01/13-1524 (14 de diciembre de 2015); ***Fiscal c. Bemba et al***. (Quinta solicitud de la Fiscalía para la admisión de pruebas de la mesa de abogados) ICC-01/05-01/13-1498 (30 de noviembre de 2015), §§ 17-18; ***Fiscal c. Yekatom y Ngaïssona*** (Transcripción) ICC-01/14-01/18-T-023 (29 de marzo de 2021), 69; imágenes: Fiscal c. Said (Transcripción) ICC-01/04-01/21-T-004 (12 de octubre de 2021), 17, e imágenes de satélite: ***Fiscal c. Al Hassan*** (Transcripción) ICC-01/12-01/18-T-027 (21 de septiembre de 2020).

5 ***Fiscal c. Al-Werfalli*** (Orden de detención) ICC-01/11-01/17-2 (15 de agosto de 2017), §§ 11-22; ***Fiscal c. Al-Werfalli*** (Segunda orden de detención) ICC-01/11-01/17-13 (5 de julio de 2018), §§ 17-18. Véase además, Emma Irving, 'And So It Begins...Social Media Evidence in an ICC Arrest Warrant' (***Opinio Juris***, 17 de agosto de 2017) <http://opiniojuris.org/2017/08/17/and-so-it-begins-social-media-evidence-in-an-icc-arrest-warrant/>.

6 ***Ponomarenko y otros contra Ucrania y Rusia***, TEDH, Ap. No. 60372/14. Pendiente. La plataforma está disponible en: <https://ilovaisk.forensic-architecture.org/>.

7 ***Ucrania y Países Bajos c. Rusia***, Decisión de admisibilidad, TEDH, App. Nos. 8019/16, 43800/14, 28525/20, 30 de noviembre de 2022, § 472.

Unidas (ONU),[8] y el enjuiciamiento nacional de crímenes internacionales.[9]

Como forma de prueba relativamente nueva, la información digital de fuentes abiertas puede resultar desconocida para muchos/as profesionales del Derecho. Por consiguiente, este documento ofrece una visión general de las principales técnicas de investigación de fuentes abiertas digitales. El fin es ayudar a los jueces, las juezas y personas responsables de la determinación de los hechos, en su propia evaluación de la información de fuentes abiertas digitales, cuando haya sido presentada por una parte en el procedimiento o por un tercero, u obtenida a través de un informe externo.[10] Es importante destacar que este documento no aborda cómo llevar a cabo investigaciones de fuentes abiertas digitales.[11] El único propósito de esta guía es ayudar en la evaluación de la credibilidad, fiabilidad y valor probatorio de la información de fuentes abiertas digitales. Se explican algunas técnicas de investigación de fuentes abiertas digitales, pero sólo para dar una idea del proceso de investigación.

El Protocolo de Berkeley sobre Investigaciones Digitales de Fuentes Abiertas ("Protocolo de Berkeley") ofrece un amplio marco sobre cómo llevar a cabo investigaciones digitales de fuentes abiertas.[12] El Protocolo de Berkeley establece las normas profesionales que deben aplicarse en la identificación, recopilación, conservación, análisis y presentación de la información digital de fuentes abiertas en las investigaciones penales internacionales y de derechos humanos. Incluye normas internacionales para llevar a cabo investigaciones en línea sobre presuntas violaciones del derecho penal internacional, humanitario y de los derechos humanos. También proporciona orientación sobre metodologías y procedimientos para recopilar, analizar y preservar la información digital de manera profesional, legal y ética.

Este documento se basa en el Protocolo de Berkeley para apoyar a los jueces, juezas y personas responsables de la determinación de los hechos en su evaluación de la información de fuentes abiertas. Debido a su importancia para los mecanismos de rendición de cuentas y justicia, este documento se centra únicamente en imágenes y vídeos digitales de fuentes abiertas, y es coherente con las definiciones, principios

8 Como ejemplo, véase Daragh Murray, Yvonne McDermott y Alexa Koenig, "Mapping the Use of Open Source Research in UN Human Rights Investigations" (2022) 14 ***Journal of Human Rights Practice*** 554-581.

9 Tribunal de Apelación de La Haya, causa n.º 22/001283-21 (6 de diciembre de 2022); Tribunal de Apelación de Suecia Occidental, Fiscal Jefe c. ***Hassan Mostafa Al-Mandlawi y Al Amin Sultan (sentencia, 30 de*** marzo de 2016); Tribunal de Distrito de Södertörn, Fiscal c. ***Mouhannad Droubi (sentencia***, 26 de febrero de 2015); Tribunal de Distrito de Örebro, Fiscal c. ***Saeed (sentencia, 19 de febrero de 2019);*** Tribunal de Distrito de La Haya, asuntos n.º 09/748012-19 y 09/748012-19-P (sentencia, 29 de junio de 2021); Tribunal de Distrito de La Haya, asunto n.º 09/748001-19 (sentencia, 16 de julio de 2021).

10 A efectos de la presente guía, los términos "***pruebas digitales de código abierto***" y "pruebas de código abierto" pueden utilizarse indistintamente, únicamente por motivos de legibilidad. No obstante, la atención se centra explícitamente en las pruebas digitales de código abierto basadas en imágenes.

11 Para una visión general de las técnicas de investigación, véase: ***Dubberley et al.***, nota 3 supra. Para cursos o información sobre investigaciones de fuentes abiertas, véanse: Amnistía Internacional, "Online Course on Open Source Human Rights Investigations" <https://advocacyassembly.org/en/partners/amnesty>; Institute for International Criminal Investigation, "Open Source Investigations Course" <https://iici.global/course/open-source-investigation-foundational/>.

12 ***El Protocolo de Berkeley*** fue elaborado por el Centro de Derechos Humanos de la Universidad de California Berkeley y la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, y supuso una serie de consultas con expertos internacionales.

y técnicas descritos en el Protocolo de Berkeley.

Las reglas de admisibilidad diferirán dependiendo de la jurisdicción, así como el posible requerimiento de prueba pericial y, de ser así, qué tipo de experto. Esto también dependerá en gran medida de los hechos específicos del caso. Esta guía se organiza en torno a una serie de cuestiones clave que un tribunal o un organismo de determinación de los hechos puede tener que abordar en su evaluación de la información digital de fuentes abiertas, incluida la determinación de la autenticidad de la imagen digital y el análisis de los metadatos pertinentes, como la fuente, la ubicación y la información temporal. Para cada cuestión, la guía define aquellos términos y técnicas que son relevantes, y ofrece ejemplos con el fin de informar el proceso de evaluación que llevan a cabo los jueces, las juezas y aquellos/as responsables de la determinación de los hechos. En cada sección hay un recuadro de "Puntos Clave", que ofrece un resumen de la información para facilitar una consulta rápida. También se incluye un glosario de términos técnicos, con hipervínculos y resaltado en negrita.

Esta guía pretende ayudar en la evaluación de cualquier material presentado, de modo que los mecanismos de rendición de cuentas puedan capitalizar el potencial de la información digital de fuentes abiertas. Estamos convencidos de que la información de fuentes abiertas seguirá siendo muy valiosa para la rendición de cuentas. La guía pretende ser ilustrativa y no exhaustiva. Aunque se abordan las principales técnicas de investigación de código abierto, surgen continuamente nuevas técnicas.

Punto Clave: Esta guía pretende ayudar a los jueces, las juezas y otros/as responsables de la toma de decisiones, en su evaluación de la información digital de fuentes abiertas, explicando algunas de las técnicas de investigación más comunes. No es una guía sobre cómo llevar a cabo investigaciones de fuentes abiertas digitales.

2.

¿Qué es la información digital de fuentes abiertas y cómo enfocar su evaluación?

El Protocolo de Berkeley define la información digital de fuentes abiertas como "información que cualquier miembro del público puede observar, adquirir o solicitar, sin necesidad de un estatus legal especial o acceso no autorizado".[13] La información digital de fuentes abiertas es "información públicamente disponible en formato digital, que generalmente se adquiere de Internet".[14] En el contexto de la rendición de cuentas, la información digital de fuentes abiertas consiste principalmente en publicaciones en redes sociales, imágenes, vídeos, documentos y grabaciones de audio en Internet, imágenes por satélite y datos publicados por el gobierno.

Para los jueces, las juezas u otras personas encargadas de determinar los hechos, hay una serie de factores que deben tenerse en cuenta a la hora de evaluar el valor probatorio de la información digital de fuentes abiertas. Por "verificación" se entiende la evaluación de toda la información disponible asociada al material. El proceso de verificación de la información de fuentes abiertas implica una combinación de diferentes técnicas, como la geolocalización, la cronolocalización y el análisis de metadatos,[15] y no se limita a una técnica específica. Al evaluar la información de fuentes abiertas es importante examinar la metodología de investigación empleada.

Puntos Clave: La evaluación de la información digital de fuentes abiertas se centra en garantizar que se ha llevado a cabo un proceso de verificación adecuado. Las técnicas de verificación utilizadas variarán inevitablemente en cada caso. Puede ser útil tener en cuenta que cada técnica forma parte de un rompecabezas corroborativo, y los/as investigadores/as deben descartar posibilidades alternativas.

13 *Protocolo de Berkeley*, § 1.

14 *Id.*

15 Estas técnicas se tratan más adelante, en la sección 4.

3.

¿Cuáles son las características distintivas de la información digital de fuentes abiertas?

En su mayor parte, las imágenes digitales de fuentes abiertas deben abordarse del mismo modo que cualquier otra forma de prueba, teniendo en cuenta factores existentes como la corroboración y la fiabilidad de la fuente. Sin embargo, hay algunas consideraciones adicionales que deben tenerse en cuenta:

- En primer lugar, las imágenes digitales de fuentes abiertas pueden no incluir indicios tradicionales de autenticidad, como información sobre la persona que grabó el material o detalles sobre el dispositivo original en el que se grabó. Y lo que es más importante, un usuario puede publicar contenidos que él mismo no grabó.
- En segundo lugar, y tal y como suele ocurrir con las cuentas de las redes sociales, las personas asociadas a una de estas cuentas pueden ser anónimas o desconocidas. Por ejemplo, algunas plataformas de redes sociales no exigen a los/as usuarios/as que faciliten su nombre real, pueden permitir que los/as usuarios/as cambien su nombre de usuario/a repetidamente, y/o que varias personas publiquen en una misma cuenta.
- En tercer lugar, la naturaleza de los entornos digitales permite la rápida difusión de grandes volúmenes de material, y la persona que publicó el material por primera vez puede ser desconocida.
- En cuarto lugar, como se explica en el apartado 4, los contenidos pueden no ser auténticos en diferentes sentidos. Las herramientas para generar o editar contenidos son ahora mucho más accesibles y pueden utilizarse sin formación profesional ni programas informáticos complejos. Además, a diferencia de las pruebas físicas, los contenidos digitales pueden manipularse a distancia.

Puntos Clave: En general, la información digital de fuentes abiertas debe abordarse de la misma manera que cualquier otra forma de prueba. Sin embargo, hay algunos atributos únicos a los que merece la pena prestar atención. Los/as proponentes de las pruebas han de abordar aquellas cuestiones que surgen a raíz de la forma en la que la información de fuentes abiertas es característicamente diferente de otras formas de prueba.

4.

Cuestiones clave al evaluar la información digital de fuentes abiertas

Esta sección identifica las cuestiones clave que deben tenerse en cuenta a la hora de evaluar la autenticidad y fiabilidad de la información digital de fuentes abierta. En el contexto de las investigaciones de fuentes abiertas, la verificación es el proceso mediante el cual se evalúa la exactitud y validez de la información. Las imágenes digitales se consideran auténticas y fiables una vez que se ha demostrado que representan lo que se afirma que representan. Al evaluar cómo se han verificado y autenticado las imágenes digitales, hay que tener en cuenta si el análisis del/de la investigador/a evalúa: (A) el contenido de las propias imágenes, (B) los metadatos, (C) la fuente, (D) la ubicación y (E) la hora; además de cómo hace dicha evaluación.

Hay una amplia gama de razones por las que los contenidos en línea pueden no ser lo que pretenden, por ejemplo:[16]

- **Atribución errónea de lugar, fecha, hora o descontextualización:** Aún cuando el contenido pueda representar hechos reales, es posible que la hora o el lugar de una foto o un vídeo se hayan atribuido erróneamente o que el contenido se haya sacado de contexto. Por ejemplo, un vídeo en el que supuestamente se mostraban ataques turcos en el norte de Siria circuló por múltiples medios de comunicación importantes en 2019. Sin embargo, poco después se descubrió que el vídeo era en realidad de un campo de tiro de Kentucky, en Estados Unidos.[17]
- **Contenidos editados (shallowfakes):** Puede ser que fotos o vídeos editados hayan sido presentados como contenido original, con o sin conocimiento de dicha edición. Por ejemplo, el contenido ha podido ser pueden cortado, o tal vez se hayan aplicado filtros, añadido o eliminado elementos, o acelerado o ralentizado los fotogramas del vídeo. Este tipo de contenido se conoce como "shallowfakes").[18] Un ejemplo de ello es un vídeo de Nancy Pelosi, Presidenta de la Cámara de Representantes de Estados Unidos,

16 Claire Wardle, 'Noticias falsas. It's complicated'. (***Primer Borrador de Noticias***, 16 de febrero de 2017) <https://firstdraftnews.org/articles/fake-news-complicated/>.

17 Heather Murphy, 'ABC Apologizes for Showing Video from U.S. Gun Range in Report on Syria' (***The New York Times***, 14 de octubre de 2019) <https://www.nytimes.com/2019/10/14/business/media/turkey-syria-kentucky-gun-range.html>.

18 Ashley Stoll, "Shallowfakes and Their Potential for Fake News" (***Washington Journal of Law, Technology & Arts,*** 13 de enero de 2020) <https://wjlta.com/2020/01/13/shallowfakes-and-their-potential-for-fake-news/>.

que se ralentizó para que pareciera que estaba ebria y que arrastraba las palabras. El vídeo fue posteriormente desmentido. [19]

- **Metadatos modificados:**

  - » Metadatos modificados o eliminados automáticamente: Es posible que los metadatos adjuntos al contenido hayan sido modificados automáticamente por una plataforma cuando el contenido es cargado. Por ejemplo, WhatsApp – como la mayoría de las redes sociales y plataformas de comunicación digital – elimina la mayoría de los metadatos de los contenidos subidos a la plataforma.

  - » Metadatos modificados o borrados manualmente: Los metadatos adjuntos al contenido pueden ser modificados, con o sin conocimiento del/la usuario/a, mostrando una ubicación, un dispositivo de grabación o un registro de fecha y hora incorrecto. También puede ser que los metadatos se hayan borrado total o parcialmente. Esta modificación puede haberse realizado a través de un editor de metadatos, o mediante una función integrada en determinados sistemas operativos, y puede haberse llevado a cabo por distintos motivos. [20]

- **Contenidos escenificados:** Es posible que el contenido haya sido escenificado utilizando actores y decorados de cine o televisión. Por ejemplo, en 2014 en relación con el conflicto de Siria, se hizo viral un vídeo de un niño rescatando a una niña bajo disparos de armas, titulado "Syrian Hero Boy". El vídeo, inicialmente presentado como auténtico,[21] fue descubierto como una grabación realizada por un grupo de cineastas y rodado con actores en un plató de Malta.

- **Contenidos generados o manipulados por Inteligencia Artificial (deepfakes o** medios sintéticos**):** A medida que la tecnología de Inteligencia Artificial (IA) se hace más accesible, puede que esta sea utilizada para generar o editar audio, fotos y vídeos.[22] Los deepfakes son un ejemplo de técnicas de generación de medios sintéticos basadas en la IA. Son una nueva forma de manipulación audiovisual que permite crear simulaciones realistas de la cara, la voz o las acciones de alguien. Por ejemplo, en 2022 circuló por las redes sociales un vídeo deepfake del presidente ucraniano

19 Hannah Denham, 'Another fake video of Nancy Pelosi goes viral on Facebook' (***Washington Post***, 3 de agosto de 2020) <https://www.washingtonpost.com/technology/2020/08/03/nancy-pelosi-fake-video-facebook/>.

20 La modificación puede realizarse con fines engañosos o por otros motivos; en algunos casos, por ejemplo, puede ser necesaria la redacción de los metadatos para preservar el anonimato. Para obtener información sobre los procesos de edición de metadatos, véase Casey Schmidt, "Revamp your information with these unique metadata editors" (***Canto***, 2 de febrero de 2021) <https://www.canto.com/blog/metadata-editor/>; Mauro Huculak, "How to edit image metadata on Windows 10" (***Windows Central***, 10 de enero de 2017) <https://www.windowscentral.com/how-edit-picture-metadata-windows-10>.

21 BBC News, '#BBCTrending: Syrian 'hero boy' video faked by Norwegian director' (***BBC News***, 14 de noviembre de 2014) <https://www.bbc.com/news/blogs-trending-30057401>.

22 WITNESS, ***Deepfakes*** (2022), disponible en línea en <https://www.mediafire.com/file/421ov54c77t04tq/Backgrounder_Deepfakes_2022.pdf/file>.

Zelensky pidiendo a sus tropas que se rindieran.[23] La tecnología de medios sintéticos también permite a los usuarios añadir o eliminar objetos fácilmente, alterar las condiciones de fondo, crear una imagen de una persona que no existe, o generar una imagen de un acontecimiento u objeto a partir de una descripción de texto, entre otras funciones.[24] Los contenidos generados o editados por IA pueden ser difíciles de detectar, y requieren el análisis de un experto en síntesis de IA o en análisis forense de medios. Las herramientas que afirman poder identificar deepfakes no siempre son precisas, y no se debe confiar únicamente en ellas a la hora de examinar contenidos sobre los que haya sospechosa de generación o manipulación de IA. En estas situaciones, también ha de tenerse en cuenta el contexto en el que se encuentran estas imágenes y la corroboración de los hechos y otros datos relevantes.[25] La determinación de cuándo un contenido fue creado, puede proporcionar información sobre las herramientas de generación o modificación disponibles en aquel momento.

No obstante, incluso los contenidos editados o no auténticos pueden tener valor probatorio.[26] El contenido puede manipularse sin intención de inducir a error. Por ejemplo, se puede cortar un vídeo y unirlo a otro, sin que el editor pretenda sugerir que las dos partes sucedieron secuencialmente. O, incluso si la intención es inducir a error, puede haber aspectos de la foto o el vídeo que tengan valor probatorio, como la fecha o la hora en que se grabó, o el propio contenido si es propaganda. También puede referirse a otros factores, como los elementos mentales de un delito (mens rea). Por supuesto, este tipo de información debe abordarse con la debida cautela. Los/as investigadores/as deben aplicar las consideraciones habituales para el análisis de la información digital de fuentes abiertas, tal como se indica en el Protocolo de Berkeley, con el fin de atribuir el valor adecuado, en su caso, a las imágenes digitales potencialmente no auténticas.

Puntos Clave: Las imágenes digitales de fuentes abiertas pueden no ser siempre lo que pretenden ser, por múltiples razones, entre ellas: atribución errónea, edición, modificación de metadatos, puesta en escena, y el uso de inteligencia artificial. Muchos de estas intervenciones pueden identificarse mediante técnicas de verificación adecuadas. A pesar de ello, las imágenes digitales no auténticas pueden tener valor probatorio.

23 Bobby Allyn, 'Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn' (***NPR***, 16 de marzo de 2022) <https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia?t=1660657155956>.

24 Open AI, 'DALL-E: Creating Images from Text' (***OpenAI***, 5 de enero de 2021) <https://openai.com/blog/dall-e/>.

25 Sam Gregory, "The World Needs Deepfake Experts to Stem This Chaos" (***Wired***, 24 de junio de 2021) <https://www.wired.com/story/opinion-the-world-needs-deepfake-experts-to-stem-this-chaos/>.

26 Véase, por ejemplo, ***Fiscal v. Nahimana, Barayagwiza & Ngeze***, Sentencia, Tribunal Penal Internacional para Ruanda, Caso nº ICTR-99-52-T, 3 de diciembre de 2003, § 274.

A. Información sobre el contenido

Aunque la información de fuentes abiertas puede presentarse de forma diferente a las formas más tradicionales de pruebas fotográficas o de vídeo, el contenido (es decir, la información que aparece en la foto o el vídeo) debe analizarse de la misma manera. Al evaluar el informe de un/a investigador/a, deben tenerse en cuenta dos componentes.

En primer lugar, el proceso seguido al examinar el contenido. El/la investigador/a debe seguir las normas profesionales para la recopilación, el análisis y la conservación de pruebas digitales de fuentes abiertas, tal como se indica en el Protocolo de Berkeley. También debe ser transparente en lo que respecta a cualquier sesgo o limitación conocidos de su trabajo, y debe haber intentado compensar los sesgos cognitivos y técnicos en la medida de lo posible.[27] El/la investigador/a debe indicar si ha puesto ha prueba hipótesis alternativas o si ha considerado otros métodos para interpretar o cuestionar su trabajo.

En segundo lugar, si el análisis y las conclusiones del/de la investigador/a se ajustan a sus conocimientos. Por ejemplo, en algunos casos puede ser necesario que un/a investigador/a consulte a un experto técnico, en la materia o en la industria (como patólogos/as forenses, botánicos/as, o expertos/as médicos/as oen armamento, militares o geoespaciales). En otros casos, puede ser necesario que un/a investigador/a consulte a personas con conocimientos específicos del contexto, ya que investigadores/as sin la experiencia adecuada en el contexto o la materia representadapueden pasar por alto pistas contextuales que refutan claramente las conclusiones, o no cuestionar adecuadamente los sesgos, las suposiciones, o la información errónea o desinformada. La persona que propone la prueba debe consultar a expertos cuando sea necesario y no hacer afirmaciones sobre lo que se representa en las imágenes que queden fuera de sus conocimientos. Si el idioma de la fuente original difiere del idioma del informe, deberá tenerse en cuenta la exactitud de la traducción.

Puntos Clave: El contenido digital de fuentes abiertas se analiza del mismo modo que las fotos o vídeos tradicionales, con dos puntos de especial escrutinio. En primer lugar, debe evaluarse el proceso del/de la investigador/a para asegurarse de que ha actuado con la diligencia debida al analizar el contenido. En segundo lugar, las conclusiones del/de la investigador/a deben adecuarse a sus conocimientos y experiencia.

27 Yvonne McDermott, Alexa Koenig y Daragh Murray, "Open Source Information's Blind Spot: Human and Machine Bias in International Criminal Investigations" (2021) 19 *JICJ* 85-105.

B. Metadatos

Los metadatos son datos que describen y proporcionan información sobre contenidos específicos, como la foto o el vídeo que se está evaluando.[28] Hay dos posibles conjuntos principales de metadatos : los metadatos adjuntos en el momento de la creación, edición o distribución; y los metadatos añadidos por los/as investigadores/as como parte del proceso de análisis o preservación. Cada uno de ellos puede proporcionar información diferente. [29]

Metadatos adjuntos en el momento de la creación, edición o distribución de los contenidos

Los metadatos incorporados en el momento de la creación del contenido digital pueden incluir la hora, la fecha y el lugar de captura, así como otra información tal que el tipo de dispositivo en el que se grabó el contenido. La creación de metadatos varía según el tipo de dispositivo en el que se creó el contenido, y depende en gran medida de cómo esté configurado dicho dispositivo, o de si la plataforma a la que se subió "elimina" automáticamente los metadatos.

Hay una serie de factores que pueden provocar variaciones en los metadatos, por ejemplo:

A. El registro de fecha y hora: Puede verse afectado si el dispositivo está en una configuración de zona horaria "por defecto";

B. Las coordenadas GPS aproximadas: Pueden verse afectadas por factores como el número de torres de telefonía móvil en las inmediaciones y su ubicación, o el nivel de cobertura del proveedor de red en la zona; y

C. Los metadatos derivados: Por ejemplo, algunos teléfonos móviles interpretan la altitud a partir de otros metadatos.

Además, normalmente es necesario usar una herramienta de visualización de metadatos para poder extraerlos e interpretarlos. Dependiendo del software que se utilice, los resultados pueden ser ligeramente diferentes, como ilustran los siguientes ejemplos (Figura A). En el caso de los contenidos generados o editados por IA, algunas herramientas pueden incluir detalles sobre el software que creó o modificó la imagen o el audio, o el modelo generativo que se utilizó (Figura B).

28 *Protocolo de Berkeley*, § 184.

29 Cabe señalar que los metadatos también pueden modificarse o crearse de otras maneras. Por ejemplo, los metadatos pueden modificarse deliberadamente después de la creación, por ejemplo para alterar la hora de grabación. Igualmente, los metadatos han podido ser añadidos automáticamente si el contenido ha sido editado con un paquete de edición de fotos o vídeos.

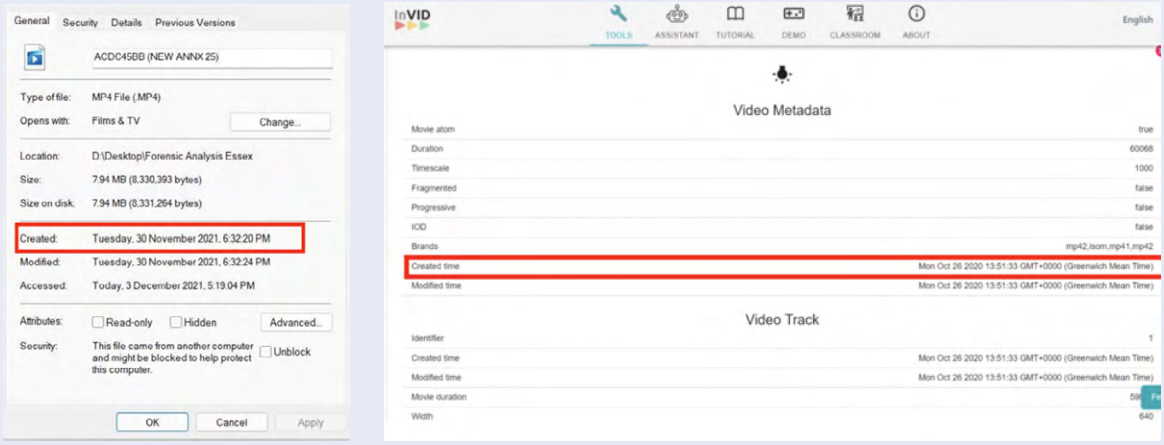


Figura A: Capturas de pantalla de dos resultados diferentes usando distintas herramientas de visualización de metadatos, extraídos por la Unidad de Verificación Digital de la Universidad de Essex. Los metadatos corresponden a un vídeo que muestra un suceso en el Lekki Tollgate de Nigeria. La imagen superior muestra los metadatos extraídos mediante el Explorador de archivos de Microsoft, con una fecha y hora de creación del 30 de noviembre de 2021 a las 18:32 horas. La imagen inferior muestra los metadatos extraídos con InVid Toolkit, que indican una fecha y hora de creación del 26 de octubre de 2020 a las 13:51 horas. Los/as investigadores/as atribuyeron esta discrepancia al hecho de que los metadatos del Explorador de archivos de Microsoft (fechados el 30 de noviembre de 2021) representaban el momento en que el archivo se cargó en el ordenador, mientras que los metadatos de InVid (fechados el 26 de octubre de 2020) correspondían al momento real de la grabación.

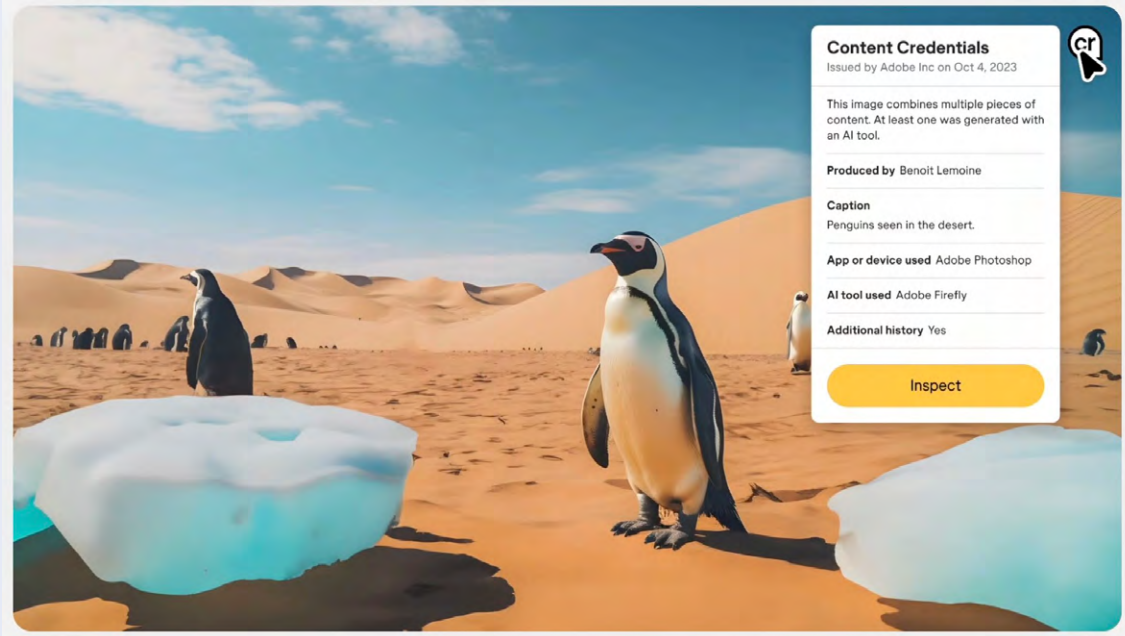


Figura B: Ejemplo de credencial de contenido que explica los métodos utilizados para crear una imagen mediante Inteligencia Artificial. Fuente: https://blog.adobe.com/en/publish/2023/10/10/new-content-credentials-icon-transparency

La precisión de los metadatos también puede depender de la configuración del usuario, especialmente en aquellos dispositivos como las cámaras de circuito cerrado de televisión (CCTV) en las que la hora debe introducirse manualmente y, por tanto, es fácil que sea errónea.

Otro problema para los/as investigadores/as de fuentes abiertas es la falta de metadatos, ya que a menudo son "eliminados" cuando se publican en las redes sociales o se envían a través de aplicaciones de mensajería como WhatsApp. Dado que gran parte del contenido relevante para esta guía se obtiene de las redes sociales, es probable que los metadatos originales no se adjunten a los informes presentados por los/as investigadores/as u otros actores.

Metadatos adjuntos en el proceso de la carga de contenido o mientras está en línea

Aunque los metadatos de los archivos originales suelen eliminarse durante el proceso de carga a Internet, pueden añadirse metadatos útiles a los contenidos en el momento de la carga y durante el tiempo que existen en línea. Como se indica en los subapartados D y E, los detalles sobre la hora y la ubicación registrados con el contenido durante el proceso de carga pueden proporcionar a los investigadores pistas adicionales para su evaluación de esta información clave. Además, las interacciones en línea con el contenido, como comentarios, comparticiones, etc., pueden proporcionar a los/as investigadores/as información útil.

En los últimos años han surgido numerosas herramientas que añaden metadatos a una imagen. Estos estándares de procedencia pueden rastrear cómo se ha creado el contenido multimedia, así como las modificaciones que se hayan llevado a cabo de forma que sea muy difícil alterar la firma criptográfica sin dejar rastro del intento. A efectos de verificación, las herramientas más valiosas son las que incluyen metadatos cifrados criptográficamente en el punto de grabación o creación del contenido, en lugar de en una fase posterior (ya que la imagen podría haber sido manipulada entretanto). Estas herramientas suelen denominarse "tecnologías de captura controlada". El diseño de este software puede variar y la integridad de sus metadatos no debe darse por sentada. Del mismo modo, el hecho de que una imagen carezca de metadatos añadidos criptográficamente no significa que no sea fiable o que no pueda autenticarse por otros medios.

Metadatos añadidos por el/la investigador/a

Los metadatos también pueden ser añadidos por los/a investigadores/as, tras obtener el contenido, como parte del proceso de análisis o de preservación. Por ejemplo, los/as investigadores/as pueden añadir información que represente sus propias interpretaciones del contenido, como podrían ser datos sobre el tipo de acontecimiento (por ejemplo, "ataque aéreo" o "tortura"). Como parte del proceso de preservación, los/as investigadores/as también pueden añadir el registro de fecha y hora (que indica cuándo el/as investigador/a recibió los datos o una estimación del momento del suceso descrito) o un valor "hash". Un valor hash criptográfic es una forma única de identificación digital (una cadena alfanumérica) que confirma,

mediante el uso de criptografía, que el contenido recogido no ha sido modificado desde el momento en que se calculó el hash.[30] Se pueden asignar valores hash a un elemento para ayudar a establecer que no ha sido manipulado desde el momento en que se aplicó el hash hasta el momento en que se presenta ante un tribunal u otro organismo de determinación de los hechos. Si se modifica una imagen digital, aunque sea ligeramente, se obtendrá un valor hash completamente nuevo.

Marcas de agua invisibles

En contenido sintético, las marcas de agua invisibles se pueden añadir a nivel de píxel o codificar en la frecuencia de audio. Aunque son imperceptibles para el ojo o el oído humano, pueden ser detectadas por programas informáticos entrenados para ello. Su edición o eliminación requiere conocimientos técnicos. Un analista forense de imágenes puede confirmar si un vídeo o una imagen han sido creados mediante inteligencia artificial.

Puntos Clave: Los metadatos deben considerarse parte de una imagen general corroborativa. Los metadatos pueden ser útiles para construir una hipótesis sobre la hora o la ubicación de una foto o un vídeo, o sobre si el contenido ha sido editado, pero deben ser evaluado. Si los metadatos no existen o son incorrectos, no significa necesariamente que el contenido no es fiable. En el caso de contenido sintético, las marcas de agua invisibles pueden proporcionar más contexto sobre una imagen con la ayuda deherramientas de detección, siempre y cuando éstas hayan sido adecuadamente entrenadas.

C. Información sobre la fuente

Tradicionalmente, los testigos declaran sobre la fuente de una foto o un vídeo. Sin embargo, dada la naturaleza del entorno digital, esto puede no ser posible en el caso de imágenes en línea de fuentes abiertas. Por ejemplo, el nivel de contenido digital disponible podría hacer imposible la comparecencia de testigos para cada foto o vídeo, asimismo la persona que subió o compartió el contenido podría ser anónima o haber fallecido. Es importante destacar que la mayoría de las técnicas de investigación descritas en esta sección no confirman la identidad de la fuente de una foto o vídeo. Más bien, son útiles para determinar las características clave de la fuente (posibles afiliaciones políticas, ubicación física aparente, o una conexión habitual con un evento) que pueden facilitar una evaluación de su fiabilidad.

A la hora de evaluar la fuente de una imagen digital, deben tenerse en cuenta las funciones de varios actores: el/la creador/a (quien grabó el contenido original), el/la que subió el contenido a Internet, el/la que lo compartió (quien lo distribuyó en línea o a través de chats de grupos de mensajería) y el/la compilador/a o editor/a

30 *Protocolo de Berkeley*, § 155(h).

(quien puede haber reunido varios vídeos en uno, o modificado el contenido de alguna manera). Estas funciones pueden solaparse. Por ejemplo, el/la creador/a puede ser también quien haya cargado el contenido en línea.

En algunos casos, pueden existir indicadores de la identidad de la persona que realiza la carga. Por ejemplo, algunas cuentas de redes sociales pueden estar "verificadas",[31] lo que sugiere la identidad probable del autor de la carga. Sin embargo, el nivel de verificación varía significativamente según la red social de la que se trate. La "verificación" se refiere aquí a quién publica el contenido, no al contenido en sí. No debe asumirse que el contenido publicado por una cuenta "verificada" es de facto creíble o fiable.

Dada la naturaleza del entorno digital, la fuente puede ser anónima o pseudoanónima. Una cuenta pseudoanónima es aquella que por ejemplo representa a una red de personas que graban y comparten contenidos, normalmente relacionados con una causa o un conflicto determinados. Un ejemplo es "Raqqa is Being Slaughtered Silently" (Raqqa está siendo masacrada en silencio), un grupo de activistas que publican regularmente sobre las actividades del ISIS en Raqqa (Siria).[32] En algunos casos, las cuentas anónimas pueden ser cuentas de imitadores (aquellas que se hacen pasar por una persona o entidad famosa) o bots (cuentas que generan publicaciones automáticamente). Para evaluar una fuente anónima, es útil analizar el comportamiento de la cuenta. Esto es, si la cuenta publica regularmente sobre un conflicto o causa, y si ese contenido es coherente dentro del contexto general (por ejemplo, si contiene información sobre regiones conocidas o partes en un conflicto), o si la fuente demuestra afiliación o sesgo político. Otro indicador de fiabilidad podría ser si otras cuentas verificadas y creíbles siguen a esa cuenta.

Por supuesto, es posible que fuentes que no publican regularmente sobre una situación puedan, no obstante, publicar una única pieza de contenido creíble y relevante. Este fue el caso de un vídeo que mostraba el asesinato de dos mujeres y dos niños en Camerún en 2018, que se compartió ampliamente en los medios de comunicación.[33] La verificación de este vídeo condujo a la detención y condena definitiva de los soldados implicados en las ejecuciones.[34]

Puntos Clave: Existen múltiples formas de evaluar la fuente de una imagen digital. Sin embargo, dada la naturaleza del entorno en línea que permite compartir imágenes en múltiples ocasiones, y que las cuentas pueden ser

31 Cuando se verifica una cuenta en una plataforma de redes sociales (como Facebook, Twitter, Instagram o YouTube), significa que la plataforma ha confirmado (según sus propias normas) que el perfil pertenece a la persona o empresa a la que representa. Las plataformas tienen en cuenta diversos factores a la hora de determinar si verifican una cuenta, como la identificación mediante documento de identidad o una dirección de correo electrónico oficial, la cobertura de noticias, el recuento de seguidores y la actividad de la cuenta.

32 Véase, por ejemplo, la página de Facebook "Raqqa is Being Slaughtered Silently الرقة تذبح بصمت', disponible en <https://m.facebook.com/Raqqa.SI/?__tn__=%2Cg>.

33 Nick Turse, 'Cameroon is a close U.S. ally-and its soldiers carried out a shocking execution of women and children' (***The Intercept***, 26 de julio de 2018) <https://theintercept.com/2018/07/26/cameroon-executions-us-ally/>.

34 BBC News, 'Cameroon soldiers jailed for killing women and children' (***BBC News***, 21 de septiembre de 2020) <https://www.bbc.co.uk/news/world-africa-54238170>.

anónimas, puede resultar difícil determinar con absoluta certeza la fuente de una imagen o un vídeo digital encontrado en línea. En los casos en que no pueda identificarse la fuente, ello no significa que el contenido no sea fiable o carezca de valor probatorio.

D. Información sobre la ubicación

Los/as investigadores/as utilizan diversas técnicas para determinar dónde se tomó una foto o se grabó un vídeo. La geolocalización se refiere a "la identificación o estimación de la ubicación de un objeto, una actividad o el lugar desde el que se generó un elemento".[35] El proceso de geolocalización pretende determinar dónde se creó el contenido. Un informe de geolocalización debe incluir una presentación transparente de las piezas de información que se utilizaron para identificar la ubicación geográfica. La precisión obtenida depende de una serie de variables.

Los metadatos de una foto o un vídeo pueden incluir una geoetiqueta (coordenadas GPS), que puede utilizarse como punto de partida para determinar dónde se tomó la grabación. Sin embargo, es posible que falten los metadatos originales (especialmente si la foto o el vídeo se publicaron en las redes sociales, que eliminan los metadatos) o que se hayan alterado, por lo que solo deben utilizarse como parte de una imagen corroborativa más general.

En su forma más simple, la geolocalización consiste en cotejar las características geográficas visibles en el contenido de una imagen (ya sean estructuras naturales o construidas por el hombre) con un lugar real, utilizando imágenes por satélite u otro material de referencia conocido, como Google Street View. Por lo general, cuantas más características únicas estén presentes y puedan compararse con los datos de referencia, mayor será el grado de confianza de que la foto o el vídeo se tomaron en ese lugar concreto.

Búsqueda inversa de imágenes o vídeos

Una búsqueda inversa consiste en cargar una imagen o fotogramas de un vídeo en un motor de búsqueda, de modo que el algoritmo de búsqueda pueda identificar otras copias de la misma imagen o de imágenes similares en Internet. Una búsqueda inversa de imágenes puede revelar si una imagen o vídeo se publicó en línea antes de la fecha de su supuesta creación. Esto puede demostrar que una imagen o vídeo no es lo que su autor afirma que es. Sin embargo, la falta de coincidencias no demuestra de forma concluyente que la imagen sea creíble. Por ejemplo, las imágenes anteriores pueden haber sido retiradas de Internet o no haber sido subidas nunca.

La limitación de una búsqueda inversa de imágenes es que sólo explora dentro

35 *Protocolo de Berkeley*, § 190.

de la base de datos de un motor de búsqueda, que es un pequeño porcentaje del contenido en Internet. No incluye, por ejemplo, materiales de la deep web (que no está indexada para los motores de búsqueda) o de la dark web (la parte de Internet a la que sólo se puede acceder mediante software especializado, como el navegador Tor). Es posible que una búsqueda inversa de imágenes en el momento de la investigación no arroje ningún resultado, pero el mismo proceso de búsqueda ejecutado en una fecha posterior -como en el momento de un procedimiento judicial- puede dar resultados. Las bases de datos de los motores de búsqueda crecen constantemente para abarcar más contenidos, y lo que se indexa puede variar significativamente de un motor de búsqueda a otro. A medida que aumenta el volumen de medios sintéticos distribuidos en línea, esto puede afectar a los archivos audiovisuales, sesgando los resultados. Asimismo, material generado o editado por IA también puede arrojar resultados al realizar una búsqueda inversa de imágenes.

En algunos casos, una búsqueda inversa de una foto o un fotograma clave de un vídeo puede dar como resultado una coincidencia directa con un lugar concreto. Esto ocurre principalmente cuando la imagen incluye una calle conocida, un punto de referencia, u otra estructura fácilmente identificable.

En última instancia, una búsqueda inversa de imágenes de un elemento descubierto en Internet puede utilizarse para evaluar si la foto o el vídeo que está analizando el/la investigador/a fue la primera publicación en línea conocida de la imagen. Las búsquedas inversas de imágenes también pueden utilizarse para descartar determinadas ubicaciones si la imagen ya ha aparecido en línea en una base de datos de búsquedas, y ya se ha confirmado que corresponde a una ubicación distinta de la que se está investigando.

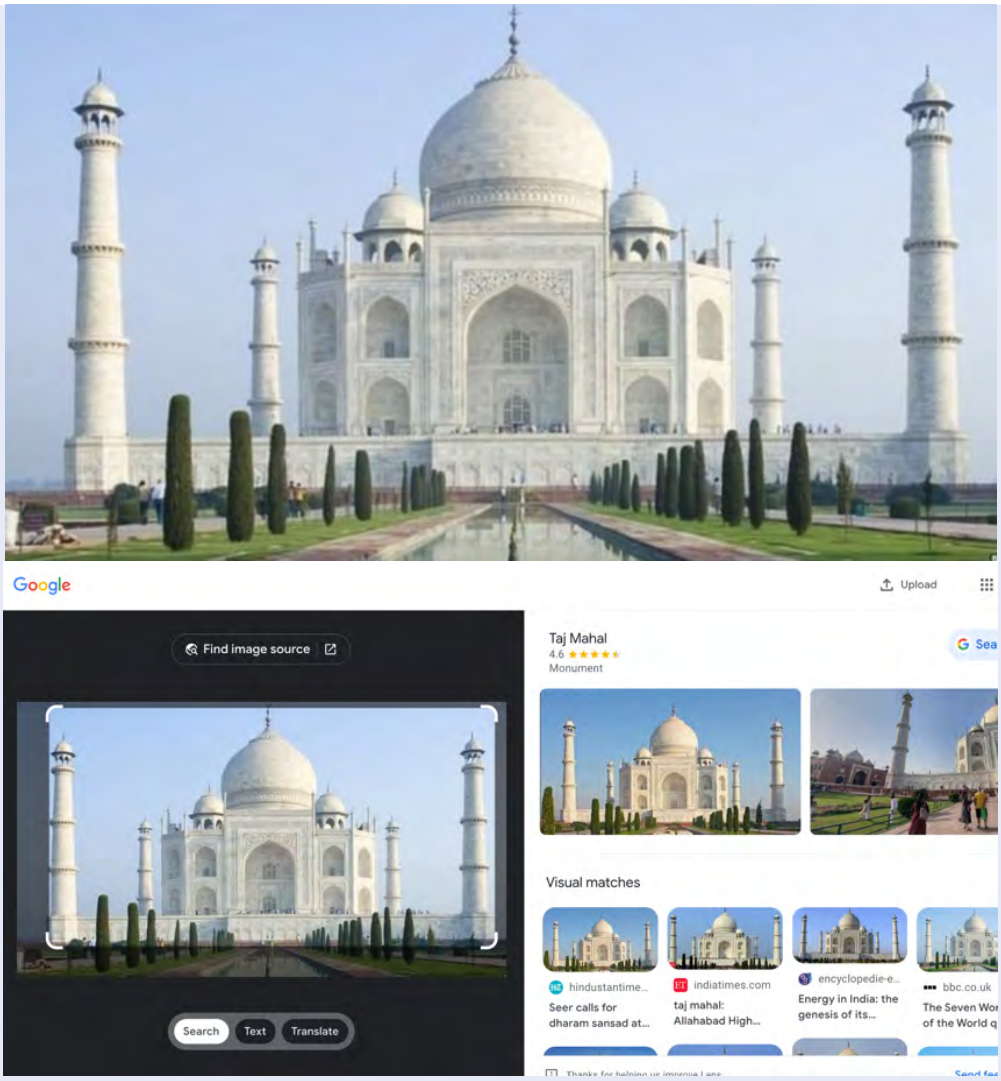


Figura C: Fotografía del Taj Mahal y captura de pantalla de la búsqueda inversa de imágenes en Google. La foto se cargó para buscarla mediante la función "buscar por imagen" de Google y arrojó el resultado anterior, que proporcionaba el nombre del monumento.

Utilizar pistas de la imagen

Para determinar dónde se tomó una imagen digital, a veces hay pistas dentro de la imagen que pueden utilizarse como indicios. Detalles visibles relacionados con la ubicación, como negocios y nombres de calles, pueden ser buscados en línea. Otra información comúnmente visible también puede indicar la ubicación: las matrículas de los coches suelen ser específicas de un lugar, al igual que los uniformes de la policía o el ejército, mientras que las farolas de las calles también pueden diferir significativamente de un lugar a otro. Este tipo de características en la foto o el vídeo pueden ayudar a reducir las posibilidades razonables de dónde se creó el contenido.

Un/a investigador/a suele anotar una captura de pantalla de la foto o el vídeo con

recuadros de colores para resaltar determinadas características, como edificios, árboles o cadenas montañosas visibles, que se utilizan como parte del análisis. Una vez identificadas estas características, los/as investigadores/as pueden relacionarlas con una ubicación utilizando una o varias de las siguientes técnicas.

Street view o mapas en 3D

Street view u otros mapas en 3D pueden utilizarse para cotejar características cuando aquellas estén disponibles, y pueden ser útiles cuando los edificios o puntos de referencia conocidos ya están etiquetados.

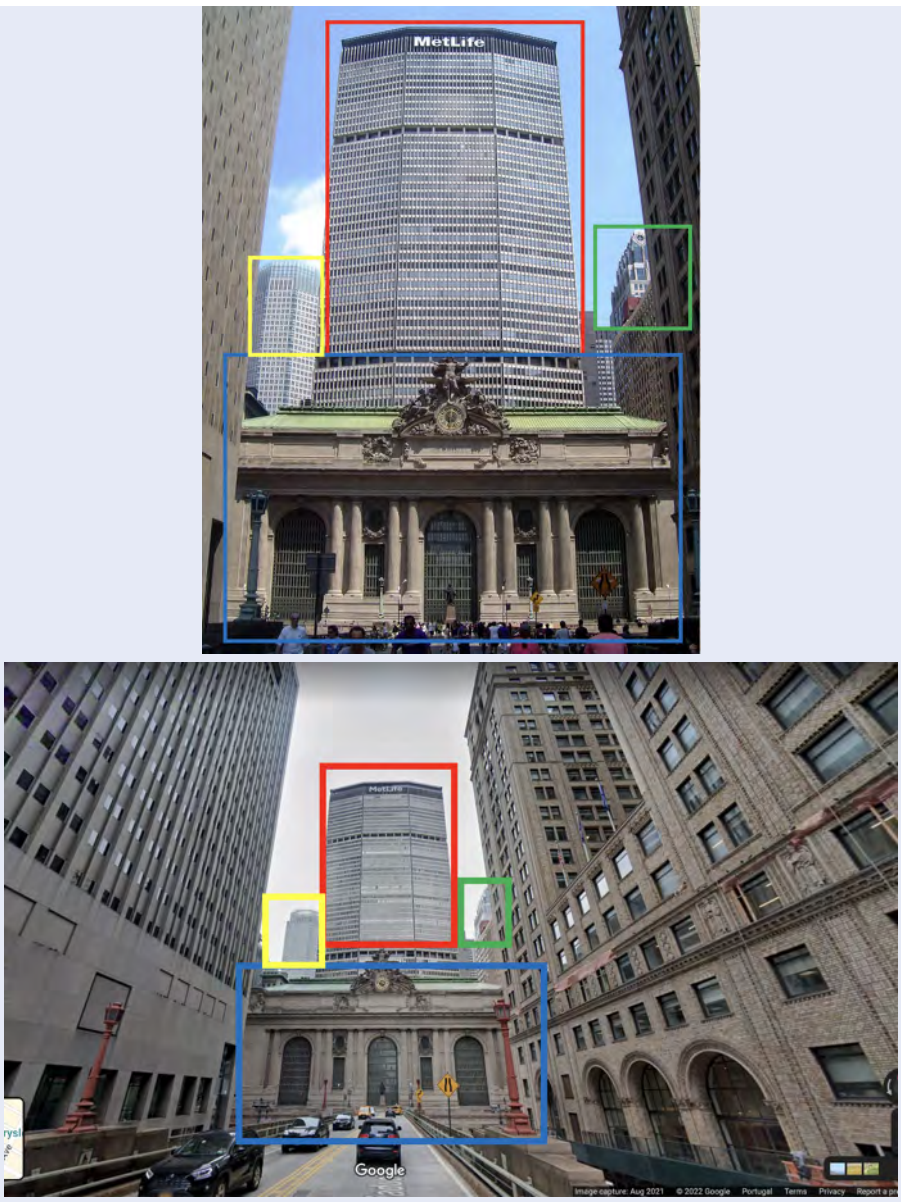


Figura D: Ejemplo de uso de la función Google Street View para hacer la correspondencia de una fotografía de la Grand Central Station de Nueva York tomada de Google Images (imagen superior) con imágenes a vista de calle en Google Maps (imagen inferior).

Imágenes por satélite

Las imágenes por satélite pueden utilizarse para comparar elementos a vista de pájaro. Para ello se suelen utilizar marcadores, como recuadros de colores, que muestran qué estructuras de una foto o un vídeo coinciden con qué puntos de las imágenes por satélite.

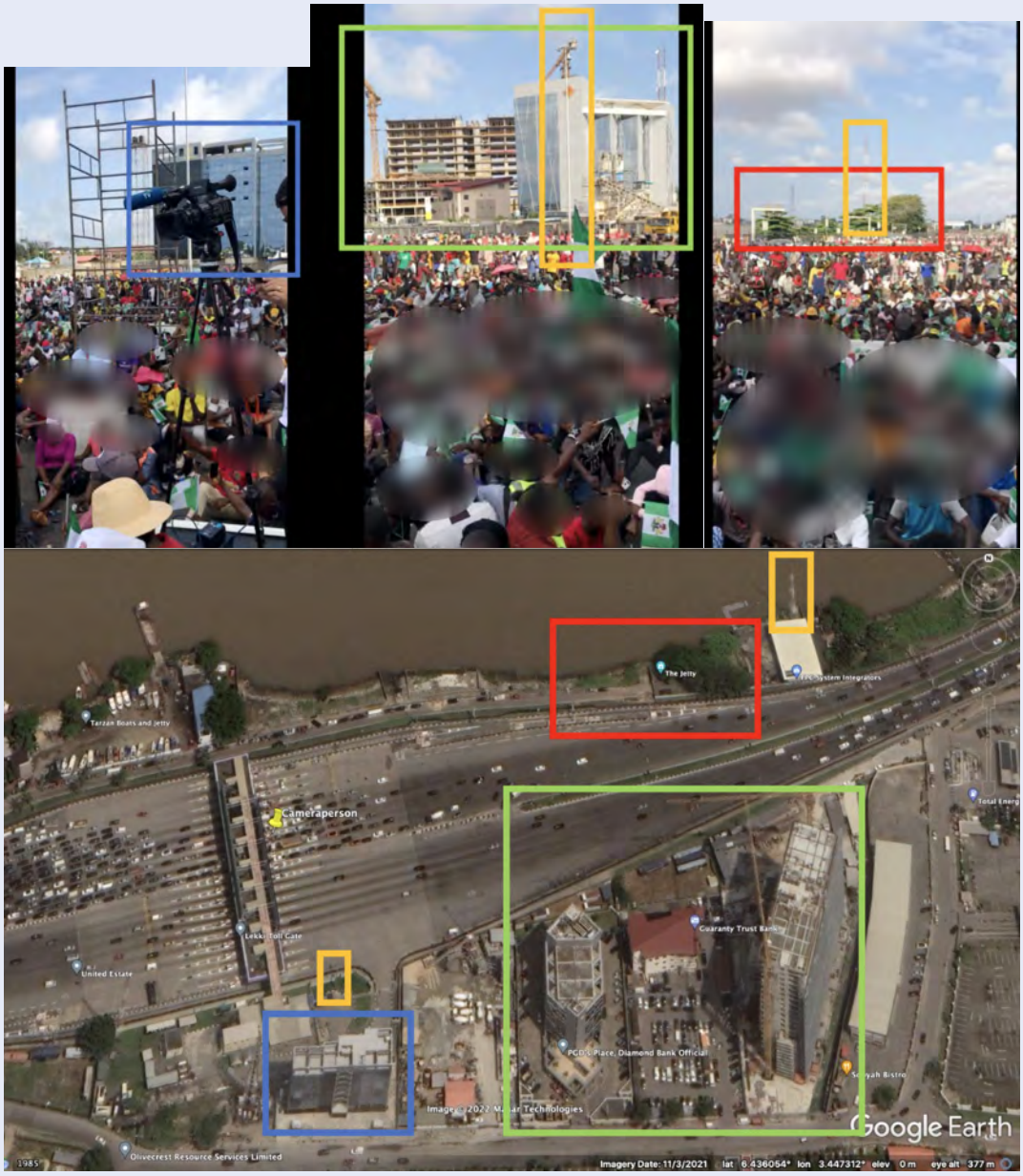


Figura E: Geolocalización de un vídeo de un incidente en el Lekki Tollgate de Nigeria realizada por la Unidad de Verificación Digital de la Universidad de Essex. Los/as investigadores/as anotaron las imágenes con recuadros de colores para indicar dónde las estructuras en el vídeo parecen coincidir con las estructuras visibles en las imágenes por satélite de Google Earth Pro.

Cartografía del terreno

La cartografía del terreno consiste en buscar características topográficas, como cadenas montañosas, y cotejarlas con imágenes de satélite, lo que puede ser útil si no hay una vista de la calle o pocas estructuras artificiales en las imágenes captadas, o si sólo se dispone de imágenes de mala calidad.

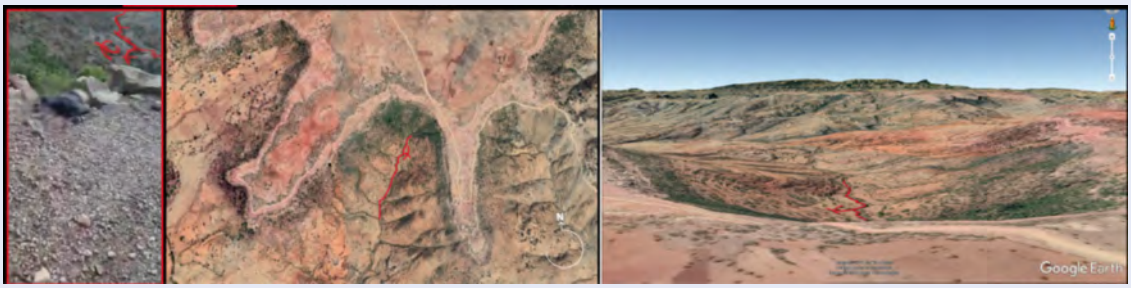


Figura F: Ejemplo de cartografía del terreno extraída de la geolocalización por el Laboratorio de Pruebas Ciudadanas de Amnistía Internacional de un vídeo de un incidente en Mahbere Dego (Etiopía). Las montañas identificadas con líneas rojas en el fondo del vídeo se cotejaron con imágenes de satélite de Google Earth Pro.

Retos de la geolocalización

La geolocalización suele llevar mucho tiempo y puede resultar muy compleja. Por lo tanto, es importante comprender sus limitaciones. El informe de investigación que acompaña a la geolocalización debe articular claramente la metodología adoptada y explicar las limitaciones del análisis.

Por ejemplo, la geolocalización suele implicar la comparación del contenido con imágenes por satélite, cuya disponibilidad y calidad pueden variar. Dependiendo del punto de la Tierra, las imágenes por satélite pueden ser de menor calidad debido a restricciones (por ejemplo, cuando los gobiernos impiden a las empresas de imágenes por satélite poner a disposición imágenes de alta resolución para determinadas zonas, como ocurrió anteriormente en Gaza).[36] Las imágenes por satélite también pueden estar obstruidas por la nubosidad o contener zonas deliberadamente bloqueadas por los gobiernos (por ejemplo, en el momento de redactar este informe, China ha bloqueado zonas de la región de Xinjiang en Baidu Maps).[37] Si las imágenes se filmaron en interiores, la geolocalización por comparación con imágenes de satélite puede resultar imposible; en su lugar, se habrán de utilizar otros métodos, como la búsqueda inversa de imágenes y la búsqueda basada en pistas. Debido a uno o varios de estos factores, o simplemente a la ausencia de características definitorias en la grabación (por ejemplo, un vídeo filmado en el mar sin puntos de referencia visibles), la geolocalización puede ser muy difícil o

36 Christopher Giles y Jack Goodman, 'Israel-Gaza: Why is the region blurry on Google Maps?' (***BBC News***, 17 de mayo de 2021) < https://www.bbc.co.uk/news/57102499>.

37 Alison Killing, Megha Rajagopalan y Christo Buschek, 'Blanked-Out Spots On China's Maps Helped Us Uncover Xinjiang's Camps' (***Buzzfeed News***, 27 de agosto de 2020) <https://www.buzzfeednews.com/article/alison_killing/satellite-images-investigation-xinjiang-detention-camps>.

imposible. En los casos en que la geolocalización no sea factible, es importante que el/la investigador/a explique por qué no ha podido geolocalizar el contenido.

Puntos Clave: Existen múltiples métodos que los/as investigadores/as pueden utilizar para determinar dónde se tomó una foto o un vídeo, entre los que se incluyen: el análisis de metadatos, las búsquedas inversas de imágenes y fotogramas de vídeos, el uso de pistas en la imagen, el cotejo de características en imágenes por satélite, y la cartografía del terreno. Toda técnica tiene sus limitaciones. Evaluaciones de localización que son fiables tendrán en cuenta y utilizarán múltiples métodos de análisis.

E. Información horaria

Por lo general, es más difícil determinar cuándo se tomó una foto o un vídeo que dónde se tomó, aunque el proceso de investigación implica métodos similares. Al evaluar cuándo se generó una foto o un vídeo, el/la investigador/a suele fijarse primero en los metadatos para determinar si contienen una fecha de creación. Sin embargo, en muchos casos, como se ha comentado anteriormente, los metadatos pueden ser inexactos o inexistentes, y los investigadores se verán obligados a utilizar otras técnicas para cronolocalizar una foto o un vídeo. Por cronolocalización se entiende "la corroboración de las fechas y horas de los acontecimientos representados en una información, normalmente imágenes visuales".[38] A continuación se exponen algunas técnicas populares de cronolocalización.

Fecha y hora de carga

Una publicación en redes sociales siempre tendrá asociado un registro de fecha y hora. Estos datos pueden ser útiles para la cronolocalización. Es importante destacar que el registro de fecha y hora indica cuándo una cuenta específica subió el contenido, pero no necesariamente se corresponde con cuándo se creó dicho contenido. Por lo tanto, las marcas de fecha y hora pueden utilizarse como indicadores de un marco temporal de cuándo tuvo lugar un acontecimiento, y pueden ser un punto de partida útil para la cronolocalización, pero no se puede confiar en ellas para indicar el momento exacto en el que tuvo lugar el suceso.

Además, el registro de fecha y hora puede variar en función de la plataforma en la que se publicó el contenido, ya que los sitios web de redes sociales operan en zonas horarias diferentes o siguen protocolos distintos para registrar la fecha y la hora. Por ejemplo, las plataformas pueden mostrar las publicaciones en la zona horaria local del espectador, o en la zona horaria donde se encuentra la empresa, y no en la zona horaria del lugar donde se tomó originalmente la foto o el vídeo.

38 *Protocolo de Berkeley*, § 191.

Identificación de pistas dentro de una imagen

A veces hay pistas en las imágenes que pueden ayudar a determinar la fecha o la hora en que se tomó una foto o un vídeo. Siempre ha sido así, pero ahora puede resultar más fácil acceder a estas pistas. Por ejemplo, si hay detalles como carteles en el fondo de una imagen, street view puede ayudar a determinar cuándo se colocaron esos carteles. Esto ocurrió cuando el ex-asesor de Trump, George Papadopoulos, fue acusado de haber abandonado el país después de que se publicara una foto suya en Londres mientras estaba siendo investigado por el FBI.[39] Los/as periodistas pudieron rebatir el momento en el que la imagen había sido tomada gracias a indicios en el fondo de la imagen, incluido un cartel en un poste de la luz; descubriendo que la foto tenía cuatro años de antigüedad y no era contemporánea.[40] Otras características de la foto o el vídeo, como la presencia de adornos navideños, también podrían ayudar a acotar la época en que se creó el contenido. Por supuesto, estos detalles pueden haber sido editados dentro o fuera del contenido.

Análisis de sombras

El análisis de sombras reconoce que las sombras proyectadas por objetos y personas pueden indicar la posición del sol en el momento de la captura. Si se conocen la ubicación y la fecha, la longitud y la dirección de las sombras creadas por el sol que aparecen en la grabación pueden indicar la hora aproximada en la que se tomó la foto o el vídeo. Una calculadora de sombras puede estimar la longitud y dirección aproximadas de la sombra de una estructura basándose en la posición del sol de acuerdo al lugar, fecha y hora introducidos por el/la investigador/a.[41] Sin embargo, el análisis temporal con estas herramientas es inexacto. Cuando es posible realizarlo, el análisis de sombras proporciona la mejor estimación que el/la investigador/a ha podido realizar de la ventana de tiempo en la que se ha tomador una foto o un vídeo, pero no es un cálculo exacto.

39 George Bowden y Jack Sommers, 'London Photo Of George Papadopoulos Was Taken At least Four Years Ago' (***HuffPost***, 31 de octubre de 2017) <https://www.huffingtonpost.co.uk/entry/george-papadopoulos-twitter-donald-trump-london-picture_uk_59f8793ae4b09b5c2568ff4c>.

40 ***Ibid.***

41 Youri van der Weide, 'Using the Sun and the Shadows for Geolocation' (***Bellingcat***, 3 de diciembre de 2020) <https://www.bellingcat.com/resources/2020/12/03/using-the-sun-and-the-shadows-for-geolocation/>.

Figura G: Cronolocalización realizada por el Sector035 de una foto tomada en el Sendero Nacional de Israel, en Tel Aviv. En la primera foto, se identificó la sombra y, a continuación, se amplió su longitud en línea con los edificios del fondo (segunda y tercera fotos). A continuación se utilizó SunCalc para estimar la hora correspondiente. (https://medium.com/quiztime/lining-up-shadows-2351ae106cec)

Análisis meteorológico histórico

Los informes meteorológicos históricos pueden utilizarse para corroborar el momento en que se tomó una foto o se grabó un vídeo, por lo general demostrando que las condiciones meteorológicas que aparecen en las imágenes coinciden con las registradas en la fecha supuesta. Sin embargo, los análisis meteorológicos

históricos pueden ser poco fiables, y las condiciones meteorológicas de las fotos o vídeos pueden ser alteradas con tecnología algorítmica y deepfakes. [42]

Búsqueda inversa de imágenes y fotogramas de vídeos

La búsqueda inversa de imágenes también puede utilizarse para la cronolocalización. La búsqueda inversa de imágenes puede ser útil para comprobar hipótesis sobre cuándo se creó una foto o un vídeo. Por ejemplo, si el pie de foto afirma que fue tomada en una fecha específica (por ejemplo, diciembre de 2017), pero una búsqueda inversa de imágenes muestra que la imagen ya existía en línea antes (por ejemplo, febrero de 2014), esto indicaría que la foto no fue tomada realmente en la fecha alegada. Sin embargo, si la búsqueda inversa de imágenes no arroja resultados, esto no significa que el contenido no existiera previamente, ya que las bases de datos de búsqueda sólo representan un pequeño porcentaje de la información en línea. En general, las búsquedas inversas de imágenes pueden proporcionar información útil para comprobar las hipótesis de los/as investigadores/as cuando arrojan resultados, pero si no los hay, no debe suponerse que la foto o el vídeo no existían previamente.

Comparación de imágenes por satélite

Las comparaciones de imágenes por satélite pueden utilizarse para la cronolocalización cuando hay cambios en una zona a lo largo del tiempo que son visibles en las imágenes por satélite. Por ejemplo, las imágenes de satélite históricas pueden utilizarse para delimitar un marco temporal en el que se construyeron o destruyeron edificios.[43] Visualizando imágenes de distintos periodos de tiempo, los/as analistas pueden ver cuándo aparecen o desaparecen nuevos detalles. Por ejemplo, las imágenes por satélite pueden indicar cuándo se quemaron edificios o se profanaron monumentos culturales,[44] o pueden mostrar tierra removida, lo que podría indicar la presencia de una fosa común. Sin embargo, es importante destacar que las imágenes por satélite (especialmente las que están disponibles públicamente, como las de Google Earth Pro) no suelen tener imágenes históricas claras para cada fecha concreta y, por lo general, sólo permiten reducir el marco temporal a unos pocos meses.

42 Samantha Cole, 'Watch an Algorithm Turn Winter Into Summer in Any Video' (***Vice,*** 5 de diciembre de 2017) <https://www.vice.com/en/article/xwvz9a/watch-an-algorithm-turn-winter-into-summer-in-any-video-image-to-image-translation>.

43 Sam Dubberley y Joe Freeman, 'Killings, corruption, land grabs: human rights violations against the Rohingya today' (***Amnistía Internacional,*** 25 de agosto de 2020) <https://citizenevidence.org/2020/08/25/rohingya-verification/>.

44 Benjamin Strick, 'Geolocalización de la destrucción de infraestructuras en Camerún: A Case Study of Kumbo and Kumfutu' (***Bellingcat,*** 21 de noviembre de 2018) <https://www.bellingcat.com/resources/case-studies/2018/11/21/geolocation-infrastructure-destruction-cameroon-case-study-kumbo-kumfutu/>; SITU Research, 'Plataforma digital de la CPI: Timbuktu, Mali' <https://situ.nyc/research/projects/icc-digital-platform-timbuktu-mali>.

Figura H: Captura de pantalla de la plataforma digital desarrollada por SITU Research para la Corte Penal Internacional que muestra la destrucción de sitios del patrimonio cultural en Malí. Las imágenes muestran el emplazamiento del mausoleo de El Kounti antes y después de la destrucción. (http://icc-mali.situplatform.com/).

Puntos Clave: Hay múltiples métodos que los/as investigadores/as pueden utilizar para proporcionar una estimación de cuándo se tomó una imagen digital, incluyendo (pero no limitado a) el análisis de metadatos, sombras, pistas en las e imágenes, del clima a través del tiempo, de la hora y la fecha de carga, de las búsquedas inversas de imágenes y vídeos, y de las comparaciones de imágenes de satélite. Toda técnica tienen sus limitaciones. Las evaluaciones fiables de cuándo se tomó una imagen digital tendrán en cuenta y utilizarán múltiples métodos de análisis.

5.

Conclusión

Los tribunales, los órganos creados en virtud de tratados de derechos humanos y otros organismos de determinación de los hechos utilizan cada vez más imágenes digitales de fuentes abiertas. Este contenido puede constituir material altamente probatorio para la evaluación forense y judicial de presuntas violaciones del Derecho internacional de los derechos humanos, del Derecho internacional humanitario y del Derecho penal internacional, y pueden ser introducidas tanto por la acusación como por la defensa en los juicios penales para apoyar sus argumentos. Sin embargo, la creciente prevalencia de fotografías y vídeos de fuentes abiertas como prueba puede acarrear riesgos de interpretación errónea o de confianza infundada, ya sea en los materiales de fuentes abiertas o en el análisis de los/as investigadores/as. Las técnicas y descripciones contenidas en esta guía están diseñadas para ayudar a los jueces, las juezas y a los/as investigadores/as a evaluar los materiales digitales de fuentes abiertas. Al mismo tiempo, este tipo de material suele formar parte de un conjunto más amplio de pruebas ante el tribunal o el órgano de investigación. En última instancia, la información digital de fuentes abiertas debe evaluarse con arreglo a las mismas normas probatorias generales que se aplican en la institución o el tribunal del que se trate, y con sujeción a las cargas y normas probatorias establecidas por dicha institución o el tribunal.

Glosario

Búsqueda inversa de imágenes/vídeos: Una búsqueda inversa consiste en cargar una imagen o un vídeo en un motor de búsqueda para que el algoritmo de búsqueda pueda identificar otras copias de la misma imagen o de imágenes similares en Internet. La limitación de una búsqueda inversa de imágenes es que sólo escanea dentro de la base de datos de un motor de búsqueda, que incluye un pequeño porcentaje del contenido actualmente en Internet. No incluye, por ejemplo, materiales de la deep web (que no está indexada para motores de búsqueda como Google) o de la dark web (la parte de Internet a la que sólo se puede acceder mediante software especializado, como el navegador Tor).

Cronolocalización: corroboración de la fechas y hora de un acontecimiento, generalmente representado en una imagen visual. Por ejemplo, puede ser posible determinar la hora del día en que se tomó una fotografía examinando la longitud de las sombras producidas por la luz solar, junto con otros indicadores.

Dark web: la parte de Internet a la que sólo se puede acceder mediante software especializado, y que permite a los usuarios y operadores de sitios web permanecer en el anonimato y sin ser rastreados.

Deep web: la parte de Internet que no está indexada y, por tanto, no es accesible a través de los motores de búsqueda.

Geolocalización: identificación o estimación de la ubicación de un objeto o una actividad, o de la ubicación desde la que se generó un elemento. Por ejemplo, puede ser posible determinar la ubicación desde la que se tomó un vídeo o una fotografía descargados de Internet utilizando técnicas de geolocalización. Dichas técnicas podrían incluir, por ejemplo, la identificación de características geográficas únicas en una fotografía con su ubicación real en un mapa.

Información de fuentes abiertas: información que cualquier miembro del público puede observar, comprar o solicitar, sin necesidad de un estatus legal especial o acceso no autorizado.

Información digital de fuentes abiertas: información disponible públicamente en formato digital, que generalmente se adquiere en Internet.

Inteligencia artificial (IA): rama de la informática dedicada a desarrollar programas para que las máquinas aprendan a reaccionar ante variables desconocidas y se adapten a nuevos entornos.

Material o contenido sintético: ambién denominados medios generativos, se definen como contenidos visuales, auditivos o multimodales generados o modifi-

cados por algoritmos (normalmente mediante inteligencia artificial). A menudo, estos resultados son realistas, no serían identificables como sintéticos por una persona normal, y pueden simular artefactos, personas o acontecimientos.

Metadatos: son datos sobre datos. Contienen información sobre un archivo electrónico y suelen incluir las características y el historial de un archivo, como su nombre, tamaño, y fechas de creación y modificación. Los metadatos pueden describir cómo, cuándo, y por quién se recopiló un archivo digital, y cómo, cuándo, y por quién se creó, accedió, modificó y formateó un archivo digital.

Pseudonimización: el tratamiento de datos personales de tal forma que la información ya no puede atribuirse a un interesado concreto sin utilizar información adicional.

Valor hash criptográfico: cálculos que pueden ejecutarse en cualquier tipo de archivo digital para generar una cadena alfanumérica de longitud fija que puede utilizarse como prueba de que un archivo digital no ha sido modificado desde que se realizó el hash de ese contenido. Esta cadena seguirá siendo la misma cada vez que se ejecute el cálculo mientras el archivo no cambie.

Verificación: se refiere al proceso de establecer la exactitud o validez de la información que se ha recopilado en línea. La fuente, el contenido y el elemento o archivo digital deben considerarse colectivamente y compararse para comprobar su coherencia.

Agradecimientos

La traducción y producción fue financiada por la Cuenta de Aceleración de Impacto ESRC de la Universidad de Swansea y el proyecto TRUE de la Universidad de Swansea, financiado por UKRI Frontier Research Grant EP/X016021/1. El trabajo también contó con el apoyo del Instituto de Humanidades y Ciencias Sociales (IHSS) de la Universidad Queen Mary de Londres.

Deseamos expresar nuestro agradecimiento a las siguientes personas, que aportaron valiosos comentarios sobre diversos aspectos del texto en distintas fases del proceso de redacción: **Dato' Shyamala Alagendra,** abogada penal internacional; **Hadi Al Khatib**, Director General, Mnemonic; **Siobhán Allen**, Abogada Senior, Global Legal Action Network (GLAN); **shirin anlen**, Tecnóloga de Medios, WITNESS; **Pavlo Bogachenko,** Associado Senior, DLA Piper; **Su Excelencia la Jueza Solomy Bossa,** Jueza, Sala de Apelaciones, Corte Penal Internacional; **Jacobo Castellanos,** Coordinador, WITNESS; **Camille Chabot**, investigador, Centro de Derechos Humanos de la Facultad de Derecho de UC Berkeley y Universidad de Pekín; **Su Excelencia la Jueza Margaret De Guzman**, Jueza, Mecanismo Residual Internacional para Tribunales Penales; **Dr Jeff Deutch**, Investigador Principal, Mnemonic; **Sam Dubberley**; **Michael Elsanadi**, Investigador de Fuentes Abiertas, Mnemonic; **Jessica Gavron**, Directora Jurídica, Centro Europeo para la Defensa de los Derechos Humanos; **Dr Matthew Gillett,** Profesor Titular, Facultad de Derecho y Derechos Humanos, Universidad de Essex; **Jonathan Hak KC; Anne Hausknecht**, Estudiante de Doctorado, Proyecto TRUE, Universidad de Swansea; **Peter Haynes KC**; **Profesor Laurence R. Helfer,** Profesor de Derecho, Duke University, Miembro del Comité de Derechos Humanos de la ONU; **Gabriele Juodkaite-Granskiene**, Jueza, Tribunal Supremo de Lituania; **Koen Kluissen**, Detective Inspector e Investigador de Fuentes Abiertas (Crímenes Internacionales), Fiscalía de los Países Bajos; **Su Excelencia la Jueza Joanna Korner**, Jueza, Corte Penal Internacional; **Profesor Philip Leach,** Catedrático de Derecho de los Derechos Humanos, Middlesex University London; **Kateryna Latysh,** MSCA4Ukraine Postdoctoral Fellow, Vilnius University y Profesora Asociada, Yaroslav Mudryi National Law University; **Nema Milaninia**, Asesora Especial del Embajador en Misión Especial de EE.UU. para la Justicia Penal Global; **Dearbhla Minogue,** Abogada Senior, Global Legal Action Network (GLAN); **Judy Mionki**, Asesora Jurídica, Corte Penal Internacional, Oficial de Enlace para la Región de África, Comité Jurídico de Derechos Humanos, Asociación Internacional de Abogados; **Yvonne Ng**, Gerente del Programa de Archivos, WITNESS; **Raluca Racusan**; **Su Señoría el Juez Keith Raynor**, Juez, Inglaterra y Gales; **Yaroslavna Sychenkova,** consultora independiente; **Profesor Yuval Shany**, Cátedra Hersch Lauterpacht de Derecho Internacional, Universidad Hebrea de Jerusalén; **Konstantina Stavrou,** doctoranda, Universidad de Viena; **Benjamin Strick**, Centre for Information Resilience; **Hryhorii Zhurakivskyi**, Misión de Asistencia de la Unión Europea a Ucrania.

Gracias también a todos/as los/as jueces y juezas y antiguos/as jueces y juezas que aportaron comentarios pero deseaban permanecer en el anonimato.

Evaluación de imágenes digitales de fuentes abiertas:
Guía para jueces, juezas y personas responsables de la determinación de los hechos