Évaluation de l'imagerie numérique provenant de sources libres d'accès:
Un guide pour les juges et les enquêteurs

Évaluation de l'imagerie numérique provenant de sources libres d'accès:
Un guide pour les juges et les enquêteurs

Évaluation de l'imagerie numérique provenant de sources libres d'accès: *Un guide pour les juges et les enquêteurs*

Évaluation de l'imagerie numérique provenant de sources libres d'accès:
Un guide pour les juges et les enquêteurs

Contents

A propos des auteurs

Ce guide a été préparé à la suite d'un atelier organisé par le Centre for Fundamental Rights à la Hertie School de Berlin les 29 et 30 juin 2022, et financé par la Digital Verification Unit de l'Université d'Essex. Les personnes suivantes ont conçu, rédigé et corrigé ce guide (par ordre alphabétique):

Professeur Başak Çalı, directrice du Centre pour les droits fondamentaux et professeure de droit international, Hertie School et directeur de recherche à l'Institut Bonavero des droits humains et professeur de droit international, Université d'Oxford.

Joseph Finnerty, doctorant, Centre pour les droits fondamentaux, Hertie School.

Lindsay Freeman, directrice de la technologie, du droit et de la politique, Centre des droits humains, Université de Californie, Berkeley, Faculté de droit.

Alexa Koenig, professeure adjointe et co-directrice du Centre des droits humains de l'Université de Californie, Berkeley, Faculté de droit.

Libby McAvoy, conseillère juridique, Mnemonic.

Yvonne McDermott Rees, professeure de droit, Hillary Rodham Clinton School of Law, Swansea University.

Daragh Murray, maître de conférences, Faculté de Droit et IHSS Fellow, Queen Mary University of London.

Jana Sadler-Forster, agent principal du contentieux stratégique, Open Society Justice Initiative et avocate, Blackstone Chambers.

Raquel Vazquez Llorente, directrice associée, menaces et opportunités technologiques, WITNESS.

Sarah Zarmsky, doctorante et maîtresse de conférences, École de droit et Centre des droits humains, Université d'Essex.

Citer comme

Évaluation de l'imagerie numérique provenant de sources libres d'accès: Un guide pour les juges et les enquêteurs (2024), publié en ligne sur https://www.trueproject.co.uk/osguide, 2024.

1.

Introduction

Les informations numériques provenant de sources libres d'accès – c'est-à-dire les informations accessibles au public sur l'internet[1] – sont de plus en plus utilisées comme éléments de preuve devant les tribunaux nationaux et internationaux, les organismes de défense des droits humains et les organismes d'enquête,[2] , où elles se sont révélées précieuses dans divers contextes.[3] Par exemple, des informations provenant de sources libres d'accès ont été présentées comme preuves dans un certain nombre d'affaires devant la Cour pénale internationale (CPI, ou 'la Cour')[4] et des vidéos trouvées en ligne ont joué un rôle important dans les mandats d'arrêt délivrés par la Cour à l'encontre de Mahmoud Mustafa Busayf Al-Werfalli.[5] Pour la première fois devant la Cour européenne des droits de l'homme, les requérants dans *l'affaire Ponomarenko et autres c. Ukraine et Russie* ont présenté une plate-forme numérique interactive pour présenter des informations provenant de sources libres d'accès.[6] *L'affaire Ukraine et Pays-Bas c. Russie* a également interrogé la manière dont les informations provenant de sources libres d'accès pouvaient être prises en considération.[7] Les photos et les vidéos provenant des réseaux sociaux sont également devenues des éléments essentiels pour les conclusions des missions

1 Comme il s'agit des informations les plus susceptibles d'être reçues par les tribunaux dans un avenir proche, ce document se concentre sur l'imagerie numérique de source ouverte, en incorporant des images et des vidéos, telles que l'imagerie satellitaire, les publications sur les médias sociaux ou les vidéos prises par un témoin sur un smartphone. Pour une définition complète des informations de source ouverte, voir Human Rights Center de la faculté de droit de l'Université de Californie à Berkeley/Haut-Commissariat des Nations Unies aux droits de l'homme (HCDH), Protocole de Berkeley sur l'utilisation de sources ouvertes numériques dans les enquêtes (ci-après, "Protocole de Berkeley") <https://www.ohchr.org/sites/default/files/2024-01/Berkeley-Protocol-French_0.pdf>, 5-8.

2 Aux fins du présent document, l'expression "organismes de défense des droits humains" est entendue au sens large et peut inclure, par exemple, les organes de traités des Nations unies ou les procédures spéciales du Conseil des droits de l'homme des Nations unies.

3 Voir par exemple Sam Dubberley, Alexa Koenig, et 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 et Ulic Egan, "Power and Privilege: Investigating Sexual Violence with Digital Open Source Information" (2021) 19 *JICJ* 55-84.

4 Il s'agit notamment de vidéos: *Le Procureur c. Gbagbo et Blé Goudé* (Transcription) ICC-02/11-01/15-T-117 (7 février 2017); *Le Procureur c. Al Mahdi* (Jugement et sentence) ICC-01/12-01/15-171 (27 septembre 2016); des posts Facebook: Le *Procureur c. Bemba et consorts* (Décision relative à la "Cinquième requête de l'Accusation aux fins d'admission d'éléments de preuve provenant de la table du barreau") ICC-01/05-01/13-1524 (14 décembre 2015); *Le Procureur c. Bemba et consorts* (Cinquième requête de l'Accusation aux fins d'admission d'éléments de preuve provenant de la table du barreau) ICC-01/05-01/13-1524 (14 décembre 2015); *Le Procureur c. Bemba et consorts.* (Cinquième requête de l'Accusation pour l'admission d'éléments de preuve à la barre) ICC-01/05-01/13-1498 (30 novembre 2015), §§ 17-18; *Procureur c. Yekatom et Ngaïssona* (Transcription) ICC-01/14-01/18-T-023 (29 mars 2021), 69; images: Procureur c. Said (Transcription) ICC-01/04-01/21-T-004 (12 octobre 2021), 17, et images satellite: *Procureur c. Al Hassan* (Transcription) ICC-01/12-01/18-T-027 (21 septembre 2020).

5 *Le Procureur c. Al-Werfalli* (Mandat d'arrêt) ICC-01/11-01/17-2 (15 août 2017), §§ 11-22; Le *Procureur c. Al-Werfalli* (Second mandat d'arrêt) ICC-01/11-01/17-13 (5 juillet 2018), §§ 17-18. Voir également Emma Irving, 'And So It Begins... Social Media Evidence in an ICC Arrest Warrant' (*Opinio Juris*, 17 août 2017) <http://opiniojuris.org/2017/08/17/and-so-it-begins-social-media-evidence-in-an-icc-arrest-warrant/>.

6 *Ponomarenko et autres c. Russie*, Cour européenne des droits de l'homme, App. No. 60372/14. En attente. La plateforme est disponible à l'adresse suivante: <https://ilovaisk.forensic-architecture.org/>.

7 *Ukraine et Pays-Bas c. Russie*, Décision de recevabilité, CourEDH, Requêtes n° 8019/16, 43800/14, 28525/20, 30 novembre 2022, § 472. Nos. 8019/16, 43800/14, 28525/20, 30 novembre 2022, § 472.

d'enquête mandatées par les Nations Unies,[8] et les poursuites nationales des crimes internationaux.[9]

En tant que forme de preuve relativement nouvelle, l'information numérique provenant de sources libres d'accès peut être peu familière à de nombreux professionnels du droit. Par conséquent, le présent document donne un aperçu des principales techniques d'enquête sur les sources numériques libres d'accès afin d'aider les juges et les enquêteurs à évaluer les informations provenant de sources numériques libres d'accès, lorsqu'elles ont été soumises par une partie à la procédure ou par un tiers, ou obtenues par le biais d'un rapport externe.[10] Il est important de noter que ce document ne traite pas de la manière de mener des enquêtes sur sources libres d'accès.[11] Le seul objectif de ce guide est d'aider à évaluer la crédibilité, la fiabilité et la valeur probante des informations provenant de sources libres d'accès. Certaines techniques d'enquête sur les sources libres d'accès sont expliquées, mais uniquement pour donner un aperçu du processus d'enquête.

Le protocole de Berkeley sur les investigations à partir d'informations issues de sources libres d'accès ("le protocole de Berkeley") fournit un cadre pour la conduite des enquêtes sur les sources numériques libres d'accès[12] Le Protocole de Berkeley établit les normes professionnelles qui doivent être appliquées dans l'identification, la collecte, la préservation, l'analyse et la présentation d'informations provenant de sources numériques libres d'accès dans le cadre d'enquêtes pénales internationales et d'enquêtes sur les droits humains. Il comprend des normes internationales pour la conduite de recherches en ligne sur des violations présumées du droit pénal international, du droit humanitaire et des droits humains. Il fournit également des conseils sur les méthodologies et les procédures de collecte, d'analyse et de conservation des informations numériques de manière professionnelle, légale et éthique.

Ce document s'appuie sur le Protocole de Berkeley pour aider les juges et autres personnes chargées de l'établissement des faits à évaluer les informations provenant de sources numériques libres d'accès. En raison de leur importance pour les mécanismes de responsabilité et de justice, ce document se concentre uniquement sur l'imagerie numérique provenant de sources numériques libres d'accès (incorporant

8 Pour des exemples, voir Daragh Murray, Yvonne McDermott et Alexa Koenig, "Mapping the Use of Open Source Research in UN Human Rights Investigations" (2022) 14 *Journal of Human Rights Practice* 554-581.

9 Cour d'appel de La Haye, affaire n° 22/001283-21 (6 décembre 2022); Cour d'appel de Suède occidentale, *Procureur général c. Hassan Mostafa Al-Mandlawi et Al Amin Sultan* (arrêt, 30 mars 2016); Tribunal de district de Södertörn, *Procureur c. Mouhannad Droubi* (arrêt, 26 février 2015); Tribunal de district d'Örebro, *Procureur c. Saeed* (jugement, 19 février 2019); Tribunal de district de La Haye, affaires n° 09/748012-19 et 09/748012-19-P (jugement, 29 juin 2021); Tribunal de district de La Haye, affaire n° 09/748001-19 (jugement, 16 juillet 2021).

10 Dans le cadre de ce guide, les termes "preuves numériques à source ouverte" et "preuves à source ouverte" peuvent être utilisés de manière interchangeable, uniquement pour des raisons de lisibilité. L'accent est toutefois mis explicitement sur les preuves numériques à source ouverte basées sur des images.

11 Pour une vue d'ensemble des techniques d'enquête, voir: Dubberley *et al, supra note 3*. Pour des cours ou des informations sur les enquêtes à source ouverte, voir: Amnesty International, "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 *Le protocole de Berkeley* a été élaboré par le Centre des droits humains de l'Université de Californie à Berkeley et le Haut Commissariat des Nations unies aux droits de l'homme, et a fait l'objet d'une série de consultations avec des experts internationaux.

des images et des vidéos) et est cohérent avec les définitions, les principes et les techniques décrits dans le Protocole de Berkeley.

Ce guide s'articule autour d'un certain nombre de questions clés qu'un tribunal ou un organisme d'enquête peut être amené à traiter dans son évaluation des informations provenant de sources numériques libres d'accès, notamment la détermination de l'authenticité de l'image numérique et l'analyse des métadonnées, de la source, de la localisation et de l'heure pertinentes. Pour chaque question, le guide définit les termes et les techniques pertinents et fournit des exemples afin d'informer les juges et les enquêteurs sur leur propre processus d'évaluation. Chaque section comporte un encadré intitulé 'Points clés à retenir', qui résume les informations pour permettre une consultation rapide. Un glossaire des termes techniques pertinents est également inclus, et les termes inclus dans le glossaire sont hyperliés et mis en évidence en caractères gras.

Les différentes juridictions varient quant à leurs règles d'admissibilité et quant à savoir si un témoignage d'expert est requis et, si oui, selon quel type d'expertise. Cela dépendra également fortement des faits spécifiques de l'affaire juridique. Ce guide a pour but d'aider à l'évaluation de tout matériel soumis, afin que les mécanismes de responsabilité puissent capitaliser sur le plein potentiel de l'information numérique provenant de sources libres d'accès. Nous sommes convaincus que l'information provenant de sources numériques libres d'accès continuera à être d'une valeur inestimable pour lutter contre l'impunité. Ce guide se veut illustratif et non exhaustif. Les principales techniques d'enquête sur les sources numériques libres d'accès sont abordées, mais de nouvelles techniques apparaissent continuellement.

Principaux enseignements: Ce guide a pour but d'aider les juges et autres décideurs à évaluer les informations provenant de sources libres d'accès, en expliquant certaines des techniques d'enquête sur les sources libres d'accès les plus courantes. Il ne s'agit pas d'un guide sur la manière de mener des enquêtes sur les sources libres d'accès.

2.

Qu'est-ce que l'information numérique provenant de sources libres d'accès et comment aborder son évaluation ?

Le protocole de Berkeley définit l'information libre d'accès comme 'l'information que tout membre du public peut observer, acheter ou demander, sans avoir besoin d'un statut juridique particulier ou d'un accès non autorisé'.[13] Les informations numériques provenant de sources libres d'accès sont des "informations accessibles au public sous forme numérique, généralement obtenues sur internet".[14] Dans le contexte de la reddition de comptes, les informations numériques provenant de sources libres d'accès comprennent notamment les messages sur les réseaux sociaux, les images, les vidéos, les documents et les enregistrements audios sur Internet, l'imagerie satellitaire et les données publiées par le gouvernement.

Les juges et autres personnes chargées de l'établissement des faits doivent tenir compte d'un certain nombre de facteurs lorsqu'ils évaluent la valeur probante des informations numériques provenant de sources libres d'accès. La "vérification"fait référence à l'évaluation de toutes les informations disponibles associées au matériel. Le processus de vérification des informations provenant de sources numériques libres d'accès implique une combinaison de différentes techniques, telles que la géolocalisation, la et l'analyse des métadonnées,[15] et ne se limite pas à une technique spécifique. Lors de l'évaluation des informations provenant de sources libres d'accès, il est important d'examiner la méthodologie d'enquête employée.

Principaux enseignements: L'évaluation des informations provenant de sources libres d'accès est centrée sur l'assurance qu'un processus de vérification approprié a été mené. Les techniques de vérification utilisées diffèrent inévitablement d'un cas à l'autre. Il convient de garder à l'esprit que chaque technique fait partie d'un puzzle de corroboration et que les enquêteurs doivent exclure toute autre possibilité.

13 *Protocole de Berkeley*, § 1.

14 *Id.*

15 Ces techniques sont examinées plus loin, à la section 4.

3.

Quelles sont les caractéristiques de l'information numérique provenant de sources libres d'accès à connaître?

Pour l'essentiel, l'imagerie numérique provenant de sources libres d'accès doit être abordée de la même manière que toute autre forme de preuve, en tenant compte des facteurs existants tels que la corroboration et la fiabilité de la source. Toutefois, il convient de garder à l'esprit quelques considérations clés supplémentaires:

- Tout d'abord, les images provenant de sources libres d'accès peuvent ne pas comporter les indices traditionnels d'authenticité, tels que des informations sur la personne qui a enregistré le matériel ou des détails sur l'appareil d'origine sur lequel les images ont été enregistrées. Il est important de noter qu'un utilisateur peut publier un contenu qu'il n'a pas lui-même enregistré.
- Deuxièmement, comme c'est souvent le cas avec les comptes de réseaux sociaux, la ou les personnes associées à un compte peuvent être anonymes ou inconnues. Par exemple, certaines plateformes de réseaux sociaux n'exigent pas des utilisateurs qu'ils fournissent leur vrai nom, peuvent permettre aux utilisateurs de changer leur nom d'utilisateur à plusieurs reprises, et/ou plusieurs personnes peuvent publier des messages sur un seul compte.
- Troisièmement, la nature des environnements numériques permet la diffusion rapide de volumes importants de matériel, et la personne qui a posté le matériel pour la première fois peut être inconnue.
- Quatrièmement, comme indiqué dans la section 4, le contenu peut être inauthentique de diverses manières. Les outils permettant de générer ou d'éditer du contenu sont aujourd'hui beaucoup plus accessibles et peuvent être utilisés sans formation professionnelle ni logiciel complexe. Contrairement aux preuves matérielles, le contenu numérique peut être modifié à distance.

Ce qu'il faut retenir: En général, les informations numériques provenant de sources libres d'accès doivent être abordées de la même manière que toute autre type de preuve. Cependant, il existe quelques caractéristiques uniques qui méritent d'être prises en compte. Les partisans de la preuve doivent répondre aux questions soulevées par les différences caractéristiques entre l'information provenant de sources libres d'accès et les autres formes de preuve.

4.

Questions clés à prendre en compte lors de l'évaluation des informations numériques provenant de sources libres d'accès

Cette section identifie les questions clés à prendre en compte lors de l'évaluation de l'authenticité et de la fiabilité de l'information numérique provenant de sources libres d'accès. Dans le contexte des enquêtes portant sur des sources libres d'accès, la vérification est le processus par lequel l'exactitude et la validité des informations sont évaluées. L'imagerie numérique est considérée comme authentique et fiable lorsqu'il a été démontré qu'elle représente ce qu'elle est censée représenter. Lors de l'évaluation de la manière dont l'imagerie numérique a été vérifiée et authentifiée, il convient de se demander si et comment l'analyse de l'enquêteur évalue: (A) le contenu de l'imagerie elle-même, (B) les métadonnées, (C) la source, (D) l'emplacement et (E) l'heure.

Il existe un grand nombre de raisons pour lesquelles le contenu en ligne peut ne pas être ce qu'il est censé être.[16] Il s'agit notamment des raisons suivantes:

- **Mauvaise attribution de lieu, de temps ou décontextualisation:** Même si le contenu dépeint des événements réels, il est possible que l'heure ou le lieu d'une photo ou d'une vidéo soit mal attribué ou que le contenu soit sorti de son contexte. Par exemple, une vidéo censée montrer des attaques turques dans le nord de la Syrie a été diffusée par de nombreux grands médias en 2019. Cependant, peu de temps après, il a été constaté que la vidéo provenait en fait d'un champ de tir du Kentucky, aux États-Unis d'Amérique.[17]
- **Contenu modifié (shallowfakes):** Dans certains cas, des photos ou des vidéos modifées peuvent être présentées comme des contenus originaux. Elles peuvent être coupées, filtrées, des éléments peuvent être ajoutés ou supprimés, ou les images de la vidéo peuvent être accélérées ou ralenties

16 Claire Wardle, "Fake news. It's complicated. (*First Draft News*, 16 février 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 octobre 2019) <https://www.nytimes.com/2019/10/14/business/media/turkey-syria-kentucky-gun-range.html>.

(ce type de contenu édité est connu sous le nom de "shallow-fakes").[18] Par exemple, une vidéo authentique de Nancy Pelosi, présidente de la Chambre des représentants aux États-Unis, a été modifiée pour donner l'impression qu'elle était en état d'ébriété et qu'elle avait des difficultés à s'exprimer. Cette vidéo a été démentie par la suite. [19]

- **Métadonnées modifiées:**

  - Métadonnées automatiquement modifiées ou supprimées: Les métadonnées attachées au contenu peuvent être automatiquement modifiées par une plateforme lorsque ce contenu est téléchargé. Par exemple, WhatsApp – comme la plupart des réseaux sociaux et des plateformes de communication numérique – supprime la plupart des métadonnées des contenus téléchargés sur la plateforme.

  - Métadonnées modifiées ou supprimées manuellement: Les métadonnées attachées au contenu peuvent être modifiées, sciemment ou non, et peuvent indiquer un lieu, un appareil d'enregistrement ou un horodatage incorrect. Elles peuvent également être totalement ou partiellement effacées. La modification peut se faire par l'intermédiaire d'un modificateur de métadonnées ou d'une fonction intégrée à certains systèmes d'exploitation, et elle peut être entreprise à des fins variées. [20]

- **Contenu mis en scène:** Le contenu peut être mis en scène à l'aide d'acteurs et de décors de cinéma ou de télévision. Un exemple de cela s'est produit en 2014 en relation avec le conflit en Syrie, lorsqu'une vidéo d'un jeune garçon sauvant une fille sous les tirs - intitulée "Syrian Hero Boy" et initialement présentée comme authentique - est devenue virale.[21] Il s'est avéré par la suite qu'un groupe de cinéastes était à l'origine de la vidéo, qui ne provenait en fait pas du conflit syrien mais avait été filmée avec des acteurs sur un plateau de tournage à Malte.

- **Contenu généré ou manipulé par l'IA (deepfakes ou médias synthétiques):** À mesure que la technologie de l'intelligence artificielle (IA) devient plus largement accessible, des fichiers audios, des photos et des vidéos numériques provenant de sources libres d'accès peuvent être générés ou modifiés par l'IA.[22] Les "deepfakes" sont un exemple de techniques de génération

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

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

20 La modification peut être entreprise à des fins trompeuses ou pour d'autres raisons; dans certains cas, par exemple, le caviardage des métadonnées peut être nécessaire pour préserver l'anonymat. Pour plus d'informations sur les processus d'édition des métadonnées, voir Casey Schmidt, "Revamp your information with these unique metadata editors" (*Canto*, 2 février 2021) <https://www.canto.com/blog/metadata-editor/>; Mauro Huculak, "How to edit image metadata on Windows 10" (*Windows Central*, 10 janvier 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 novembre 2014) <https://www.bbc.com/news/blogs-trending-30057401>.

22 WITNESS, *Deepfakes* (2022), disponible en ligne à l'adresse suivante: <https://www.mediafire.com/file/421ov54c77t04tq/Backgrounder_Deepfakes_2022.pdf/file>.

de médias synthétiques basées sur l'IA. Il s'agit d'une nouvelle forme de manipulation audiovisuelle qui permet de créer des simulations réalistes du visage, de la voix ou des actions d'une personne. Par exemple, une vidéo deepfake du président ukrainien Zelensky appelant ses troupes à se rendre a circulé en 2022 sur les réseaux sociaux.[23] La technologie des médias synthétiques permet également aux utilisateurs d'ajouter ou de supprimer facilement des objets, de modifier les conditions d'arrière-plan, de créer l'image d'une personne qui n'existe pas ou de générer l'image d'un événement ou d'un objet à partir d'une description textuelle, entre autres caractéristiques.[24] Les contenus générés ou modifiés par l'IA peuvent être difficiles à détecter et nécessiter l'analyse d'un expert en synthèse d'IA ou en criminalistique des médias. Les outils qui prétendent identifier les deepfakes ne sont pas toujours exacts et ne doivent pas être utilisés seuls lors de l'examen d'un contenu suspect; le contexte et la corroboration doivent également être pris en compte.[25] La détermination de la date de création d'un élément de contenu peut donner des indications sur les outils de génération ou de modification de médias qui étaient disponibles à l'époque.

Néanmoins, même un contenu modifié ou inauthentique peut avoir une valeur probante.[26] Le contenu peut être manipulé sans intention de tromper. Par exemple, une vidéo peut être coupée et jointe à une autre vidéo, sans que le rédacteur ait l'intention de suggérer que les deux parties se suivent l'une l'autre de manière séquentielle. Par ailleurs, même si l'intention est de tromper, certains aspects de la photo ou de la vidéo peuvent avoir une valeur probante, comme la date ou l'heure à laquelle la séquence a été enregistrée, ou le contenu lui-même s'il s'agit de propagande. Cela peut également concerner d'autres facteurs tels que les éléments mentaux d'un crime (mens rea). Ce type d'information doit bien entendu être abordé avec la prudence qui s'impose. Les enquêteurs doivent appliquer les considérations habituelles pour l'analyse des informations numériques provenant de sources libres d'accès, telles qu'elles sont décrites dans le protocole de Berkeley, afin d'attribuer la valeur appropriée, le cas échéant, à des images numériques potentiellement inauthentiques.

Principaux enseignements: Les images provenant de sources libres d'accès en ligne ne sont pas toujours ce qu'elles sont censées être, pour de multiples raisons, notamment l'attribution erronée, la manipulation, la modification des métadonnées, la mise en scène et l'utilisation de l'intelligence artificielle pour créer ou modifier le

23 Bobby Allyn, "Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn" (*NPR*, 16 mars 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 janvier 2021) <https://openai.com/blog/dall-e/>.

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

26 Voir, par exemple, *Procureur c. Nahimana, Barayagwiza & Ngeze*, Jugement, Tribunal pénal international pour le Rwanda, Affaire n° ICTR-99-52-T, 3 décembre 2003, § 274.

contenu. Bon nombre de ces raisons peuvent être identifiées à l'aide de techniques de vérification appropriées. L'imagerie numérique inauthentique peut néanmoins avoir une valeur probante.

A. Informations sur le contenu

Bien que les informations provenant de sources libres d'accès puissent être présentées d'une manière différente des formes plus traditionnelles de preuves photographiques ou vidéo, le contenu (c'est-à-dire les informations représentées sur la photo ou la vidéo) doit être analysé de la même manière. Lors de l'évaluation du rapport d'un enquêteur, deux éléments doivent être pris en compte.

Premièrement, le processus suivi lors de l'examen du contenu. L'enquêteur doit respecter les normes professionnelles en matière de collecte, d'analyse et de préservation des preuves provenant de sources libres d'accès, telles que décrites dans le protocole de Berkeley. Il doit également faire preuve de transparence en ce qui concerne les biais ou les limites connus de son travail, et doit avoir tenté de compenser les biais cognitifs et techniques dans la mesure du possible.[27] Le chercheur doit indiquer s'il a testé d'autres hypothèses ou envisagé d'autres méthodes d'interprétation ou de remise en question de son travail.

Deuxièmement, il s'agit de savoir si l'analyse et les conclusions de l'enquêteur correspondent à son expertise. Par exemple, dans certains cas, il peut être nécessaire pour un enquêteur de consulter un expert technique, un expert en la matière ou un expert industriel (comme des médecins légistes, des botanistes ou un expert médical, en armement, militaire ou géospatial). Dans d'autres cas, il peut être nécessaire que l'enquêteur consulte des personnes ayant une connaissance spécifique du contexte. En effet, les enquêteurs n'ayant pas l'expertise appropriée du contexte ou du sujet décrit - comme ceux qui ne connaissent pas personnellement la zone prétendument décrite - peuvent manquer des indices contextuels qui réfutent clairement les conclusions ou ne pas interroger de manière appropriée les préjugés, les suppositions ou les informations erronées. La personne qui présente les preuves doit consulter des experts si nécessaire et ne pas faire d'affirmations sur ce qui est représenté sur les images qui ne relèvent pas de son expertise. Si la langue de la source originale diffère de celle du rapport, l'exactitude de la traduction doit être prise en compte.

Principaux enseignements: Le contenu des sources libres d'accès est analysé de la même manière que les photos ou les vidéos traditionnelles, mais deux points doivent faire l'objet d'une attention particulière. Tout d'abord, il convient d'évaluer le processus de l'enquêteur pour s'assurer qu'il a fait preuve d'une diligence raisonnable lors de l'analyse du contenu. Deuxièmement, les conclusions de l'enquêteur doivent correspondre à ses connaissances et à son expertise.

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

B. Métadonnées

Les métadonnées sont des données qui décrivent et fournissent des informations sur des éléments de contenu spécifiques, tels que la photo ou la vidéo évaluée.[28] Il existe deux principaux ensembles de métadonnées possibles pour chaque élément: les métadonnées attachées au moment de la création, de la modification ou de la distribution, et les métadonnées ajoutées par les enquêteurs dans le cadre du processus d'analyse ou de préservation. Chacun de ces ensembles peut fournir des informations différentes. [29]

Métadonnées attachées au moment de la création, de la modification ou de la distribution du contenu

Les métadonnées intégrées au moment de la création du contenu numérique peuvent inclure l'heure, la date et le lieu de la capture, ainsi que des informations telles que le type d'appareil sur lequel le contenu a été créé. La création de métadonnées varie selon le type d'appareil qui a créé le contenu et dépend en grande partie de la manière dont l'appareil est configuré ou si la plateforme sur laquelle il a été téléchargé "dépouille" (c'est-à-dire supprime) automatiquement les métadonnées.

Un certain nombre de facteurs peuvent entraîner des variations dans les métadonnées:

A. L'horodatage: Elle peut être influencée par le fait que l'appareil est réglé sur un fuseau horaire "par défaut";

B. Les coordonnées GPS approximatives: Celles-ci peuvent être influencées par des facteurs tels que le nombre et l'emplacement des tours de téléphonie mobile à proximité, ou le niveau de couverture du fournisseur de réseau dans la zone; et

C. La lecture de métadonnées dérivées: Par exemple, certains téléphones mobiles interprètent l'altitude sur la base d'autres métadonnées.

En outre, les métadonnées doivent généralement être examinées à l'aide d'un visualiseur de métadonnées afin de les extraire et de les interpréter. Selon la visionneuse de métadonnées utilisée, les résultats peuvent être légèrement différents, comme l'illustrent les exemples suivants (figure A). Pour les contenus générés ou modifiés par l'IA, certains outils peuvent intégrer des détails sur le logiciel qui a créé ou modifié l'image ou l'audio, ou sur le modèle génératif utilisé (figure B).

28 Protocole de Berkeley, § 184.

29 Il convient de noter que les métadonnées peuvent également être modifiées ou créées d'autres manières. Par exemple, les métadonnées peuvent être délibérément modifiées après la création, afin d'altérer l'heure de l'enregistrement, etc. De même, des métadonnées peuvent être ajoutées automatiquement si le contenu est édité à l'aide d'un logiciel d'édition de photos ou de vidéos.

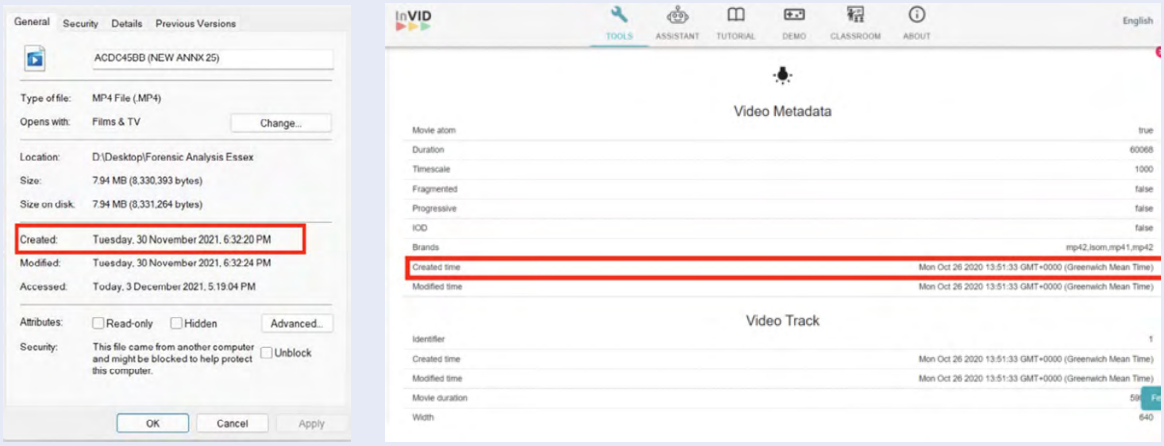

Figure A: captures d'écran de deux sorties de la visionneuse de métadonnées extraites de la vérification, par l'unité de vérification numérique de l'université d'Essex, d'une vidéo décrivant un événement survenu au Tollgate de Lekki au Nigeria. L'image du haut montre les métadonnées extraites à l'aide de l'Explorateur de fichiers Microsoft, avec une date et une heure de création du 30 novembre 2021 à 18 h 32. L'image du bas montre les métadonnées extraites à l'aide de la boîte à outils InVid, indiquant une date et une heure de création du 26 octobre 2020 à 13:51. Les chercheurs ont attribué cette différence au fait que les métadonnées de Microsoft File Explorer (datées du 30 novembre 2021) correspondent au moment où le fichier a été téléchargé sur l'ordinateur, tandis que les métadonnées d'InVid (datées du 26 octobre 2020) correspondent à l'heure réelle de l'enregistrement.

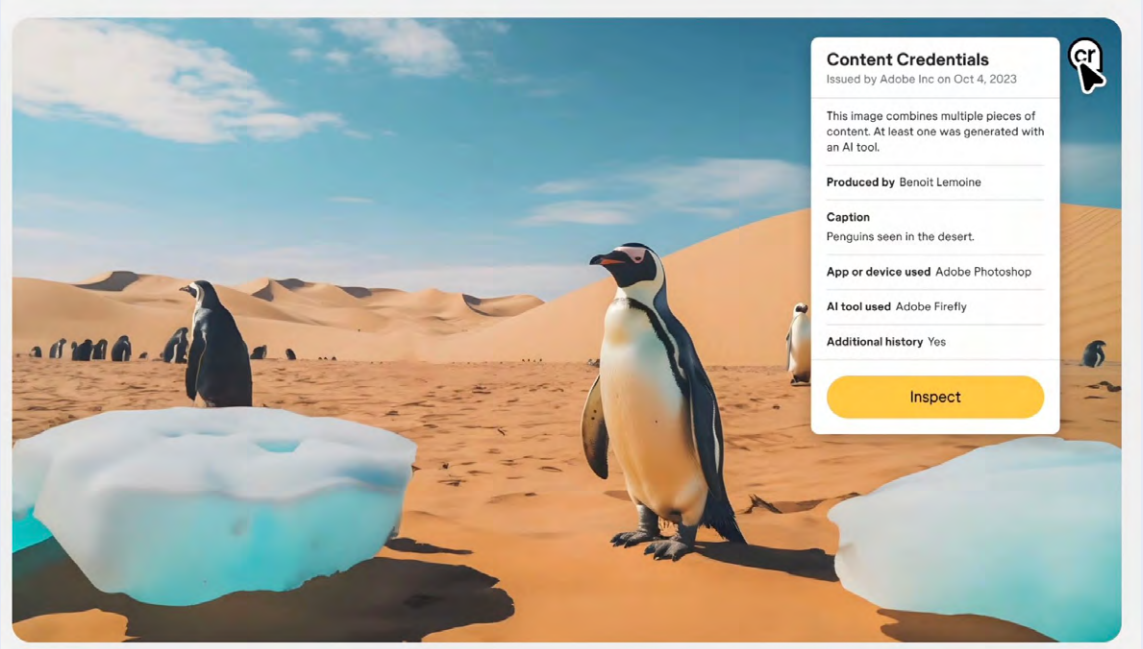

Figure B: exemple de références de contenu expliquant les méthodes utilisées pour créer une image à l'aide de l'Intelligence Artificielle. Source: https://blog.adobe.com/en/publish/2023/10/10/new-content-credentials-icon-transparency

La précision des métadonnées peut également dépendre de la configuration de l'utilisateur pour des dispositifs tels que les caméras de télévision en circuit fermé (CCTV), dans lesquelles l'heure doit être saisie manuellement et peut donc facilement être erronée.

Les métadonnées manquantes constituent un autre défi pour les chercheurs en sources libres d'accès, car les métadonnées sont souvent "retirées" du contenu lorsque celui-ci est publié sur des sites de réseaux sociaux ou envoyé par le biais d'applications de messagerie telles que WhatsApp. Étant donné qu'une grande partie du contenu pertinent pour ce guide provient des réseaux sociaux, il est probable que les métadonnées originales ne soient pas jointes aux rapports soumis par les chercheurs ou d'autres acteurs.

Métadonnées jointes lors du téléchargement ou en ligne

Alors que les métadonnées du fichier original sont souvent supprimées du contenu en ligne au cours du processus de téléchargement, des métadonnées utiles peuvent être ajoutées au contenu au moment du téléchargement et pendant la durée de son existence en ligne. Comme indiqué dans les sous-sections D et E ci-dessous, les détails relatifs à l'heure et au lieu enregistrés avec le contenu au cours du processus de téléchargement peuvent fournir aux enquêteurs des indices supplémentaires pour l'évaluation de ces informations clés. En outre, les interactions en ligne avec le contenu - telles que les commentaires, les partages et autres - peuvent fournir aux enquêteurs des informations utiles.

Ces dernières années, de nombreux outils intégrant des métadonnées dans une image ont vu le jour. Ces normes de provenance des médias permettent de suivre la fabrication d'un média, ainsi que les modifications qu'il a pu subir, d'une manière qui rend très difficile l'altération de la signature cryptographique sans laisser de preuves de la tentative. À des fins de vérification, les outils les plus utiles sont ceux qui intègrent des métadonnées hachées cryptographiquement au moment de la collecte, plutôt qu'à un stade ultérieur (car l'image aurait pu être manipulée entre-temps). Ces outils sont souvent appelés "technologie de capture contrôlée". La conception de ces logiciels peut varier et l'intégrité de leurs métadonnées ne doit pas être considérée comme acquise. De même, le fait qu'une image ne comporte pas de métadonnées chiffrées ne signifie pas que son contenu n'est pas fiable ou qu'il ne peut pas être authentifié par d'autres moyens.

Métadonnées ajoutées par l'enquêteur

Les métadonnées peuvent également être ajoutées par les enquêteurs, après avoir obtenu le contenu, dans le cadre du processus d'analyse ou de conservation. Par exemple, les enquêteurs peuvent ajouter des informations au paquet numérique qui représentent leur propre interprétation du contenu, comme de nouvelles données sur le type d'événement (par exemple, "frappe aérienne" ou "torture"). Dans le cadre du processus de conservation, les enquêteurs peuvent également ajouter des métadonnées telles qu'un horodatage (indiquant quand l'enquêteur a reçu les données ou une estimation de l'heure de l'événement décrit) ou une valeur de

hachage. Une valeur de hachage est une forme unique d'identification numérique (une chaîne alphanumérique) qui confirme, par l'utilisation de la cryptographie, que le contenu collecté n'a pas été modifié depuis le moment où la valeur de hachage a été calculée.[30] Des valeurs de hachage peuvent être attribuées à un élément pour aider à établir qu'il n'a pas été altéré entre le moment où le hachage a été appliqué et le moment où il est soumis à un tribunal ou à un autre organisme d'enquête. Si une image numérique est modifiée, même légèrement, cela se traduira par une valeur de hachage entièrement nouvelle.

Filigranes invisibles

Pour les médias synthétiques, les filigranes invisibles sont intégrés au niveau des pixels du contenu visuel ou encodés dans la fréquence audio. Ils sont imperceptibles à l'œil ou à l'oreille humaine, mais ils peuvent être détectés par des logiciels formés à cet effet. Leur modification ou leur suppression nécessite un savoir-faire technique. Un analyste d'images judiciaires peut être en mesure de confirmer si une vidéo ou une image a été créée à l'aide d'une intelligence artificielle.

Principaux enseignements: Les métadonnées doivent être considérées comme faisant partie d'une image corroborante globale. Elles peuvent être utiles pour formuler une hypothèse sur l'heure ou le lieu d'une photo ou d'une vidéo, ou pour déterminer si le contenu a été modifié, mais elles doivent être évaluées. Des métadonnées incorrectes ou manquantes ne signifient pas nécessairement que le contenu n'est pas fiable. Pour les médias synthétiques, des filigranes invisibles peuvent aider des outils de détection bien formés à fournir plus de contexte sur une image.

30 Protocole de Berkeley, § 155(h).

C. Informations sur la source

Traditionnellement, les témoins indiquent la source d'une photo ou d'une vidéo. Cependant, étant donné la nature distincte de l'environnement numérique, il se peut que cela ne soit pas possible pour l'imagerie numérique provenant de sources libres d'accès en ligne. Par exemple, l'ampleur du contenu numérique peut rendre impossible la présence de témoins pour chaque photo ou vidéo, ou bien le téléchargeur ou le partageur du contenu peut être anonyme ou décédé. Il est important de noter que la plupart des techniques d'enquête décrites dans cette section ne permettent pas de confirmer l'identité de la source d'une photo ou d'une vidéo. En revanche, elles sont utiles pour déterminer les principales caractéristiques de la source (telles que des affiliations politiques potentielles, une localisation physique apparente ou un lien régulier avec un événement).

Il existe de nombreux acteurs dont les rôles doivent être pris en compte lors de l'évaluation de la source d'une image numérique, notamment le créateur (qui a enregistré le contenu original), le téléchargeur (qui a publié le contenu sur l'internet), le partageur (qui a distribué le contenu en ligne ou par le biais de chats de groupes de messagerie) et le compilateur ou modificateur (qui peut avoir assemblé plusieurs vidéos en une seule ou modifié le contenu d'une manière ou d'une autre). Ces rôles peuvent se chevaucher. Par exemple, le créateur peut également être le téléchargeur.

Dans certains cas, il existe des indicateurs de l'identité du téléchargeur. Par exemple, certains comptes de réseaux sociaux peuvent être "vérifiés",[31] , ce qui suggère l'identité probable du téléchargeur, bien que le niveau de vérification associé aux différentes plateformes de réseaux sociaux puisse varier considérablement. La "vérification" se réfère ici à l'auteur du message et non au contenu lui-même. Il ne faut pas supposer que le contenu publié par un compte "vérifié" est de facto crédible ou fiable.

Compte tenu de la nature de l'environnement numérique, la source peut être anonyme ou pseudo-anonyme. Un compte pseudo-anonyme peut être un compte représentant un réseau de personnes filmant et partageant du contenu, généralement lié à une certaine cause ou à un conflit. Un exemple est "Raqqa is Being Slaughtered Silently" (Raqqa est massacrée en silence), un groupe d'activistes qui publie régulièrement des messages sur les activités d'ISIS à Raqqa, en Syrie.[32] Dans certains cas, les comptes anonymes peuvent être des comptes d'imitateurs (qui se font passer pour une personne ou une entité célèbre) ou des bots (comptes qui génèrent automatiquement des messages). Pour évaluer une source anonyme,

31 Lorsqu'un compte est vérifié sur une plateforme de médias sociaux (telle que Facebook, Twitter, Instagram ou YouTube), cela signifie que la plateforme a confirmé (selon ses propres normes) que le profil est authentique pour la personne ou l'entreprise qu'il représente. Les plateformes prennent en compte divers facteurs pour déterminer si elles doivent vérifier un compte, notamment l'identification à l'aide d'une pièce d'identité ou d'une adresse électronique officielle, la couverture médiatique, le nombre d'adeptes et l'activité du compte.

32 Voir par exemple la page Facebook "Raqqa is Being Slaughtered Silently الرقة تذبح بصمت", disponible à <https://www.facebook.com/Raqqa.SI/>.

il est utile d'analyser le comportement du compte. Par exemple, si le compte publie régulièrement des messages sur un conflit ou une cause, et si ce contenu est cohérent dans le contexte général (par exemple, s'il contient des informations sur des régions connues ou des parties à un conflit), ou si la source fait preuve d'affiliation politique ou de partialité. Un autre indicateur de fiabilité pourrait être le fait que d'autres comptes vérifiés et crédibles suivent ce compte.

Il est bien sûr possible que des sources qui ne publient pas régulièrement sur une situation puissent néanmoins publier un seul élément de contenu crédible et pertinent. C'est le cas d'une vidéo montrant l'assassinat de deux femmes et de deux enfants au Cameroun en 2018, qui a été largement diffusée dans les médias.[33] La vérification de cette vidéo a conduit à l'arrestation et à la condamnation finale des soldats impliqués dans les exécutions.[34]

Principaux enseignements: Il existe de multiples façons d'évaluer la source d'une image numérique. Cependant, étant donné la nature de l'environnement en ligne dans lequel les séquences peuvent être partagées et re-partagées, et le fait que les comptes puissent être anonymes, il peut être difficile de déterminer la source d'une image ou d'une vidéo numérique provenant de sources libres d'accès avec une certitude absolue. Lorsque la source ne peut être identifiée, cela ne signifie pas que le contenu n'est pas fiable ou qu'il n'a pas de valeur probante.

D. Informations sur l'emplacement

Les enquêteurs utilisent diverses techniques pour déterminer où une photo ou une vidéo a été prise. La géolocalisation désigne "l'identification ou l'estimation de l'emplacement d'un objet, d'une activité ou de l'endroit à partir duquel un élément a été généré".[35] Le processus de géolocalisation vise à déterminer où le contenu a été créé. Un rapport de géolocalisation doit inclure une présentation transparente des éléments d'information granulaires qui ont été utilisés pour identifier l'emplacement géographique. La précision obtenue dépend d'un certain nombre de variables.

Les métadonnées d'une photo ou d'une vidéo peuvent inclure un géotag (coordonnées GPS), qui peut être utilisé comme point de départ pour évaluer le lieu où la séquence a été prise. Toutefois, les métadonnées originales peuvent être manquantes (en particulier si une photo ou une vidéo a été publiée sur les réseaux sociaux, ce qui supprime les métadonnées) ou avoir été modifiées, et ne doivent donc être utilisées que comme élément de corroboration.

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 juillet 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 septembre 2020) <https://www.bbc.co.uk/news/world-africa-54238170>.

35 Protocole de Berkeley, § 190.

Dans sa forme la plus simple, la géolocalisation consiste à faire correspondre des caractéristiques géographiques visibles dans le contenu d'une image (structures naturelles ou construites par l'homme) à un lieu réel, à l'aide d'images satellite ou d'autres documents de référence connus, tels que Google Street View. En règle générale, plus il y a de caractéristiques uniques présentes et pouvant être mises en correspondance avec des données de référence, plus le degré de confiance dans le fait que la photo ou la vidéo a été prise à cet endroit précis est élevé.

Recherche inversée d'images ou de vidéos

Une recherche inversée consiste à télécharger une image ou des images fixes d'une vidéo vers un moteur de recherche, afin que l'algorithme de recherche puisse identifier d'autres copies de la même image ou d'images similaires sur internet. Une recherche inversée d'images peut révéler si une image ou une vidéo a été mise en ligne avant la date de sa création présumée. Cela peut démontrer qu'une image ou une vidéo n'est pas ce que son auteur prétend qu'elle est. Toutefois, l'absence de correspondance ne prouve pas de manière irréfutable que l'image est crédible. Par exemple, des images antérieures peuvent avoir été mises hors ligne ou n'avoir jamais été téléchargées.

La limite d'une recherche d'image inversée est qu'elle ne scanne que la base de données d'un moteur de recherche, qui ne comprend qu'un petit pourcentage du contenu d'internet. Elle n'inclut pas, par exemple, le contenu du deep web (qui n'est pas indexé par les moteurs de recherche) ou du dark web (la partie d'internet à laquelle on ne peut accéder qu'à l'aide d'un logiciel spécialisé, comme le navigateur Tor). Il est possible qu'une recherche d'images inversées au moment de l'enquête ne donne aucun résultat, mais que le même processus de recherche exécuté à une date ultérieure - par exemple au moment d'une procédure judiciaire - donne des résultats. Les bases de données des moteurs de recherche ne cessent de s'enrichir, et le contenu indexé peut varier considérablement d'un moteur de recherche à l'autre. L'augmentation du volume de médias synthétiques en ligne peut affecter les archives audiovisuelles et fausser les résultats. Inversement, les médias générés ou modifiés par l'IA peuvent toujours donner un résultat de recherche d'image inversée.

Dans certains cas, la recherche inversée d'une photo ou d'une image clé d'une vidéo peut permettre d'établir une correspondance directe avec un lieu particulier. Cela se produit principalement lorsque l'image comprend une rue connue, un point de repère ou une autre structure facilement identifiable.

En fin de compte, une recherche d'image inversée d'un élément découvert sur l'internet peut être utilisée pour déterminer si la photo ou la vidéo analysée par l'enquêteur est la première mise en ligne connue de l'image. Les recherches d'images inversées peuvent également être utilisées pour exclure certains lieux si l'image est déjà apparue en ligne dans une base de données de recherche et qu'il a déjà été confirmé qu'il s'agissait d'un lieu différent de celui faisant l'objet de l'enquête.

Figure C: Photo du Taj Mahal et capture d'écran de la recherche inversée d'images sur Google. La photo a été téléchargée pour effectuer une recherche à l'aide de la fonction "recherche par image" de Google et a donné le résultat ci-dessus, qui indique le nom du monument.

Utiliser les indices de l'image

Pour déterminer où une image numérique a été prise, il existe parfois des indices dans l'image qui peuvent être utilisés comme informations principales. Les détails visibles liés à l'emplacement, tels que les entreprises et les noms de rue, peuvent faire l'objet d'une recherche. D'autres informations visibles peuvent également indiquer un lieu: les plaques d'immatriculation des voitures sont généralement spécifiques à un lieu, tout comme les uniformes de la police ou de l'armée, tandis que les lampadaires peuvent également différer de manière significative d'un lieu à l'autre. Ce type de caractéristiques dans la photo ou la vidéo peut aider à réduire les possibilités raisonnables quant à l'endroit où le contenu a été créé.

Un enquêteur annotera souvent une capture d'écran de la photo ou de la vidéo avec des cases colorées afin de mettre en évidence certaines caractéristiques, telles que des bâtiments distincts, des arbres ou des chaînes de montagnes visibles, qui sont utilisées dans le cadre de l'analyse. Une fois ces caractéristiques identifiées, les enquêteurs peuvent les associer à un lieu en utilisant une ou plusieurs des techniques suivantes:

Vue de la rue ou cartes en 3D

La vue des rues ou d'autres cartes en 3D peuvent être utilisées pour faire correspondre les caractéristiques lorsqu'elles sont disponibles et peuvent être utiles lorsque des bâtiments ou des points de repère connus sont déjà étiquetés.

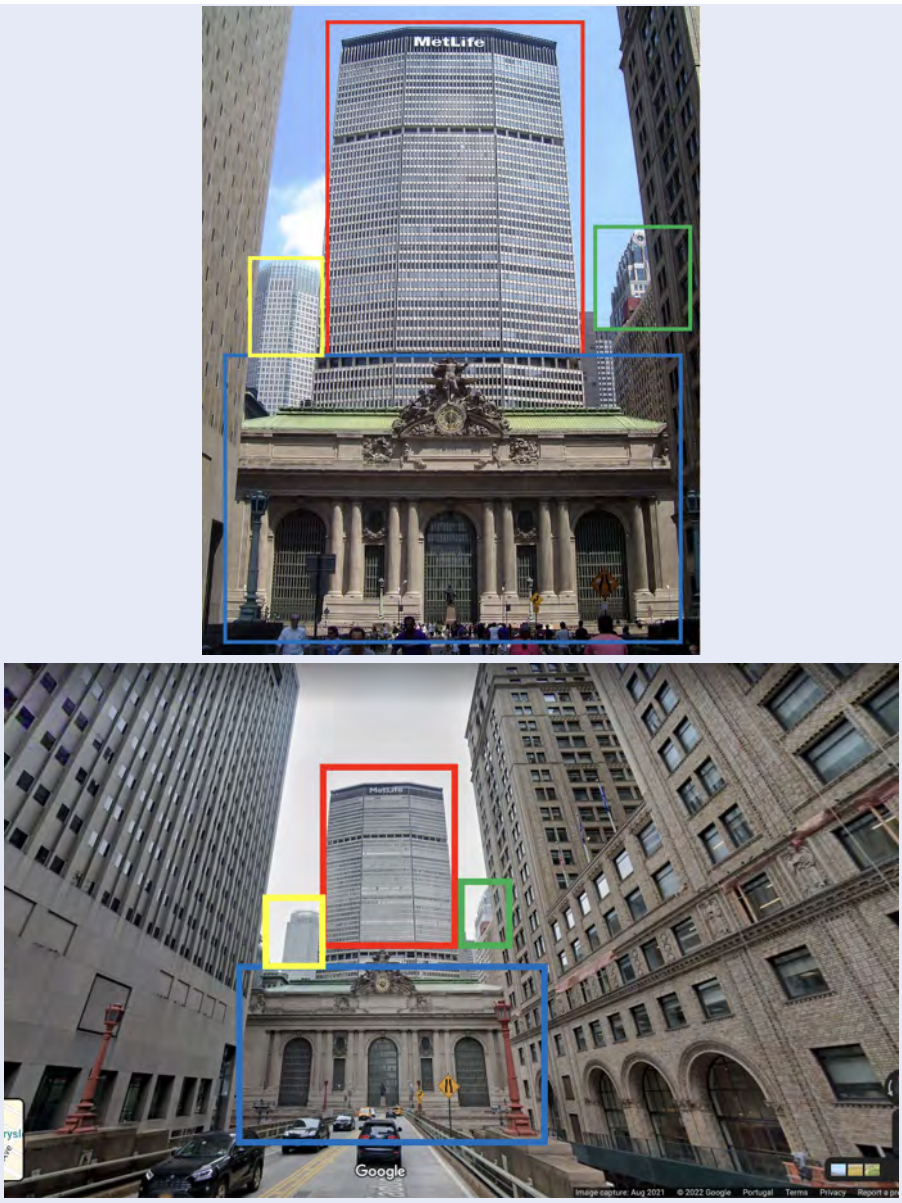

Figure D: Exemple d'utilisation de la fonction Google Street View pour faire correspondre une photographie Google Images de la gare Grand Central à New York (image du haut) avec des images de la vue de la rue sur Google Maps (image du bas).

Imagerie satellitaire

L'imagerie satellitaire peut être utilisée pour faire correspondre les caractéristiques d'une vue à vol d'oiseau. Pour ce faire, on utilise généralement des marqueurs, tels que des boîtes colorées, pour montrer quelles structures d'une photo ou d'une vidéo correspondent à quels points de l'imagerie satellitaire.

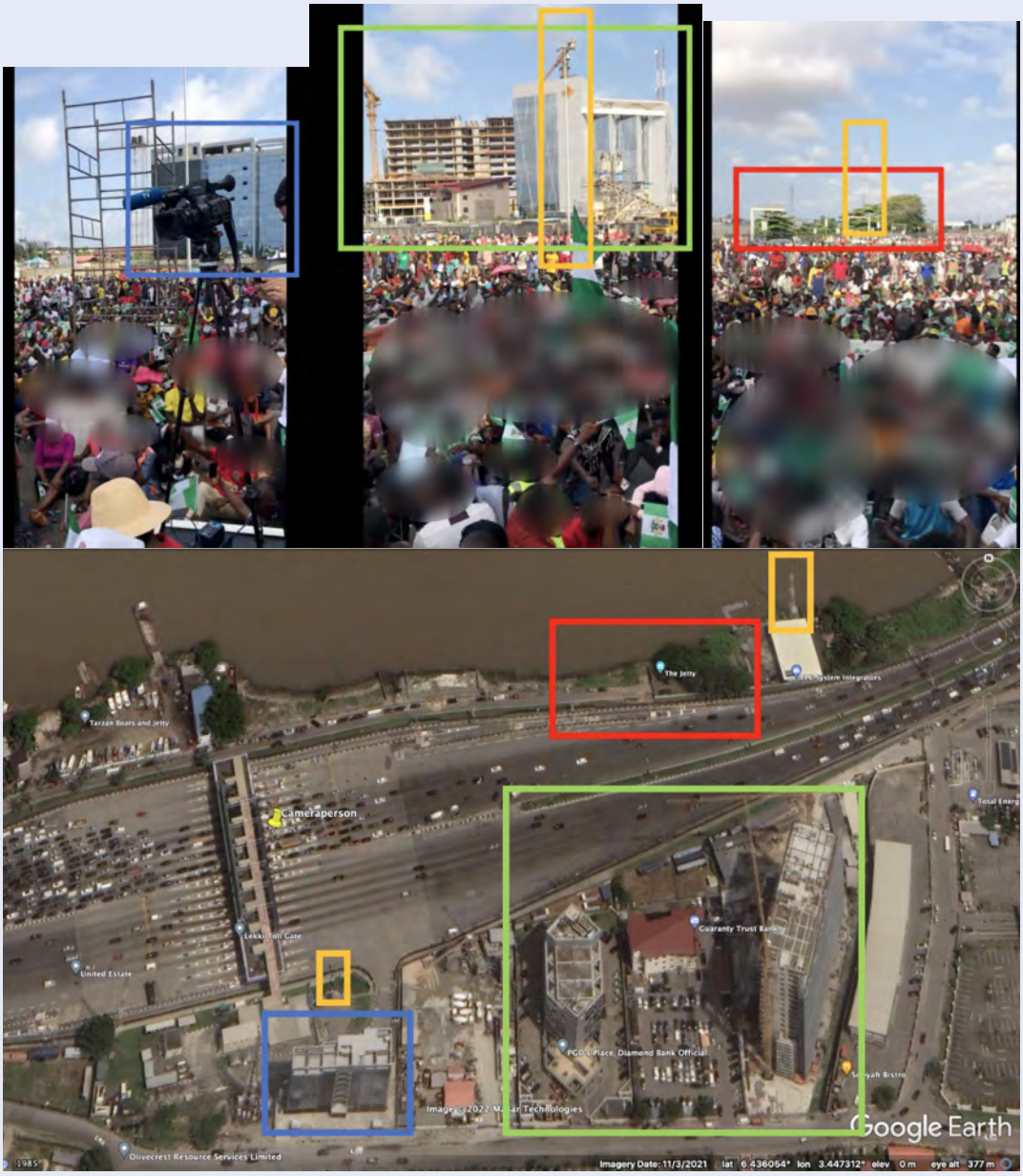

Figure E: Géolocalisation d'une vidéo d'un incident survenu au péage de Lekki au Nigeria, réalisée par l'unité de vérification numérique de l'université d'Essex. Les enquêteurs ont annoté les images avec des cases colorées pour indiquer où les structures vues dans la vidéo de source ouverte semblent correspondre aux structures visibles sur l'imagerie satellite de Google Earth Pro.

Cartographie du terrain

La cartographie du terrain consiste à rechercher des caractéristiques topographiques telles que des chaînes de montagnes et à les faire correspondre à l'imagerie satellitaire, ce qui peut être utile s'il n'y a pas de vue de rue ou peu de structures humaines dans les images capturées, ou si l'on ne dispose que d'une imagerie de qualité médiocre.

Figure F: Exemple de cartographie du terrain tiré de la géolocalisation par le Citizen Evidence Lab d'Amnesty International d'une vidéo d'un incident à Mahbere Dego en Éthiopie. Les montagnes identifiées par des lignes rouges en arrière-plan de la vidéo ont été comparées à des images satellite de Google Earth Pro.

Les défis de la géolocalisation

La géolocalisation prend souvent du temps et peut être très difficile. Il est donc important d'en comprendre les limites. Le rapport d'enquête accompagnant la géolocalisation doit clairement exposer la méthodologie adoptée et expliquer les éventuelles limites de l'analyse.

Par exemple, la géolocalisation implique généralement de comparer le contenu à l'imagerie satellitaire, dont la disponibilité et la qualité peuvent varier. En fonction du point de la Terre, l'imagerie satellitaire peut être de moindre qualité en raison de restrictions (par exemple, lorsque les gouvernements empêchent les sociétés d'imagerie satellitaire de mettre à disposition des images à haute résolution pour certaines zones, ce qui s'est déjà produit à Gaza).[36] L'imagerie satellitaire peut également être obstruée par la couverture nuageuse ou contenir des zones délibérément occultées par les gouvernements (par exemple, au moment de la rédaction de ce document, la Chine a occulté des zones de la région du Xinjiang sur Baidu Maps).[37] Si la séquence a été filmée en intérieur, la géolocalisation par comparaison avec l'imagerie satellite peut s'avérer impossible; d'autres méthodes telles que la recherche d'images inversées et la recherche de d'autres pistes doivent être utilisées à la place. En raison d'un ou de plusieurs de ces facteurs, ou simplement

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

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

de l'absence de toute caractéristique déterminante dans la séquence (par exemple, une vidéo filmée en mer sans aucun point de repère visible), la géolocalisation peut s'avérer très difficile, voire impossible. Dans les cas où la géolocalisation n'est pas possible, il est important que l'enquêteur explique pourquoi il n'a pas pu géolocaliser le contenu.

Principaux enseignements: Les enquêteurs peuvent utiliser de nombreuses méthodes pour déterminer où une photo ou une vidéo a été prise, notamment: l'analyse des métadonnées, la recherche inversée d'images et de vidéos, l'utilisation d'indices à partir de l'image, la comparaison de caractéristiques sur l'imagerie satellite et la cartographie du terrain. Toutes ces méthodes ont leurs limites. Les évaluations fiables de la localisation doivent prendre en compte et utiliseront plusieurs méthodes d'analyse.

E. Informations temporelles

Il est généralement plus difficile de déterminer quand une photo ou une vidéo a été prise que l'endroit où elle a été prise, bien que le processus d'enquête fasse appel à des méthodes similaires. Pour déterminer la date de création d'une photo ou d'une vidéo, l'enquêteur examine d'abord les métadonnées afin de déterminer si elles contiennent un horodatage de la date de création. Cependant, dans de nombreux cas, les métadonnées sont inexactes ou manquantes, et les enquêteurs doivent utiliser d'autres techniques pour chronolocaliser une photo ou une vidéo. La chronolocalisation est la "corroboration des dates et heures des événements décrits dans un élément d'information, généralement des images visuelles".[38] Voici quelques techniques de chronolocalisation courantes.

Horodatage du téléchargement

Les horodatages des messages en ligne peuvent également être utiles pour la chronolocalisation, qui sera toujours indiquée avec un message sur les sites de réseaux sociaux. Il est important de noter que les horodateurs indiquent le moment où le compte en question a téléchargé le contenu, et non le moment où le contenu a été créé. En tant que tels, les horodateurs peuvent être utilisés comme indicateurs du moment où un événement a eu lieu et peuvent constituer un point de départ utile pour la chronolocalisation, mais ils ne peuvent pas être utilisés pour indiquer l'heure exacte à laquelle un événement s'est produit.

En outre, l'horodatage peut varier en fonction de la plateforme sur laquelle le contenu a été publié, car les sites web des réseaux sociaux fonctionnent dans des fuseaux horaires différents ou suivent des protocoles d'horodatage différents. Par exemple, certaines plateformes peuvent afficher les messages dans le fuseau horaire local du spectateur ou dans le fuseau horaire où se trouve l'entreprise, et non dans le fuseau horaire de l'endroit où la photo ou la vidéo a été prise à l'origine.

38 Protocole de Berkeley, § 191.

Identification des pistes au sein d'une image

Les images contiennent parfois des indices qui permettent de déterminer la date ou l'heure à laquelle une photo ou une vidéo a été prise. Cela a toujours été le cas, mais il est désormais plus facile d'accéder à ces indices. Par exemple, s'il y a des détails temporaires tels que des affiches à l'arrière-plan d'une image, l'imagerie historique facilement disponible sur street view peut aider à déterminer quand ces affiches ont été posées. C'est ce qui s'est produit lorsque l'ancien collaborateur de Trump, George Papadopoulos, a été accusé d'avoir quitté le pays alors qu'il faisait l'objet d'une enquête du FBI, après qu'une photo de lui à Londres a été publiée.[39] Les journalistes ont pu contester que la photo avait été prise à la date alléguée en raison d'indices dans l'arrière-plan de l'image, notamment une affiche sur un lampadaire; il a finalement été constaté que la photo datait de quatre ans et qu'elle n'était pas contemporaine.[40] D'autres caractéristiques de la photo ou de la vidéo, telles que la présence de décorations de Noël, pourraient également aider à préciser la saison à laquelle le contenu a été créé. Bien entendu, ces détails peuvent avoir été modifiés dans ou hors du contenu.

Analyse de l'ombre

L'analyse des ombres reconnaît que les ombres projetées par les objets et les individus peuvent indiquer la position du soleil au moment de la prise de vue. Si le lieu et la date sont connus, la longueur et la direction des ombres créées par le soleil et représentées sur les images peuvent indiquer l'heure approximative à laquelle la photo ou la vidéo a été prise. Les calculateurs d'ombres calculent indépendamment la longueur et la direction approximatives de l'ombre pour des structures de différentes tailles en fonction de la position du soleil à l'endroit estimé, de la date

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

40 Ibid.

et de l'heure saisis par l'enquêteur.[41] Cependant, l'analyse du temps à l'aide de ces outils est inexacte. Lorsqu'elle est possible, l'analyse des ombres fournit la meilleure estimation de l'enquêteur par son examen visuel pour une fenêtre de temps au cours de laquelle une photo ou une vidéo a été prise, mais il ne s'agit pas d'un calcul exact.

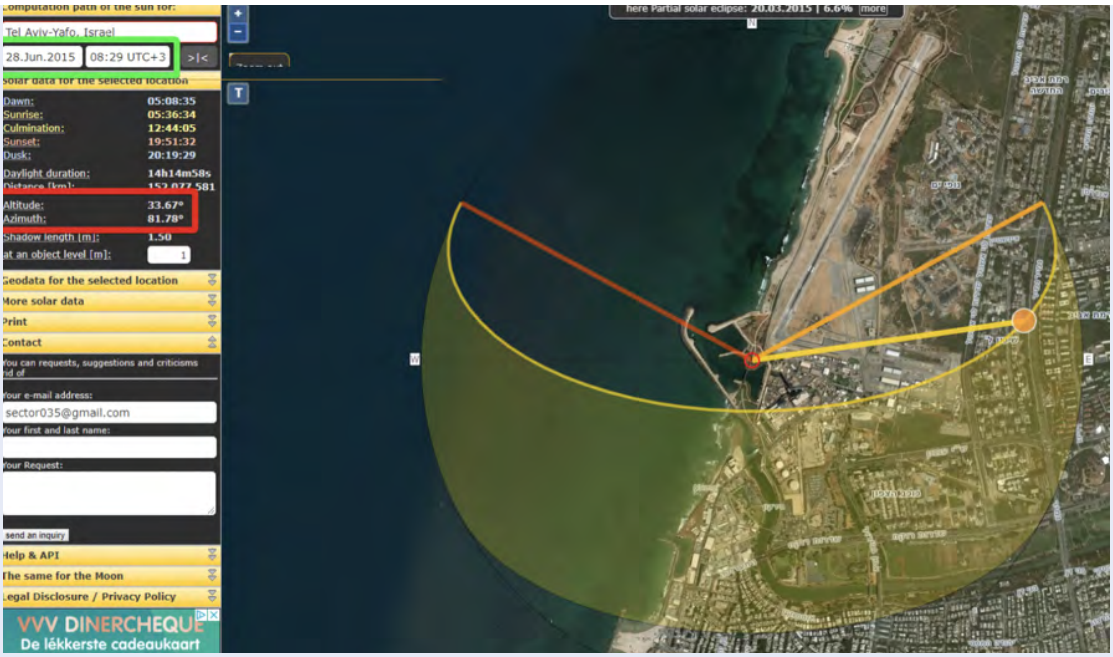

Figure G: Chronolocalisation par Sector035 d'une photo prise sur l'Israel National Trail à Tel Aviv. Sur la première photo, l'ombre a été identifiée, puis sa longueur a été étendue en fonction des bâtiments à l'arrière-plan (deuxième et troisième photos). SunCalc a ensuite été utilisé pour estimer le temps correspondant. (https://medium.com/quiztime/lining-up-shadows-2351ae106cec)

Analyse historique des conditions météorologiques

Les rapports météorologiques historiques peuvent être utilisés pour corroborer davantage l'heure à laquelle une photo ou une vidéo a été prise, généralement en montrant que les conditions météorologiques décrites dans les images coïncident avec les conditions météorologiques signalées à la date présumée. Cependant, les analyses météorologiques historiques peuvent être peu fiables et les conditions météorologiques des photos ou des vidéos peuvent désormais être modifiées grâce à la technologie algorithmique et à la technologie des "deepfakes" (simulacres profonds). [42]

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

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

Recherche inversée d'images et de vidéos

La recherche d'images inversées peut également être utilisée pour la chronolocalisation. La recherche d'images inversées peut être utile pour tester des hypothèses sur la date de création d'une photo ou d'une vidéo. Par exemple, si la légende d'une photo indique qu'elle a été prise à une date précise (ex. décembre 2017), mais qu'une recherche d'image inversée montre que l'image existait en ligne auparavant (ex. février 2014), cela indiquerait que la photo n'a pas été prise à la date alléguée. Toutefois, si la recherche inversée d'images ne donne aucun résultat, cela ne signifie pas que le contenu n'existait pas auparavant, car les bases de données de recherche ne représentent qu'un faible pourcentage de l'information en ligne. En général, les recherches d'images inversées peuvent fournir des informations utiles pour tester les hypothèses des enquêteurs lorsqu'elles donnent des résultats, mais si elles ne donnent aucun résultat, il ne faut pas en déduire que la photo ou la vidéo n'existait pas auparavant.

Comparaison de l'imagerie satellitaire

Les comparaisons d'images satellitaires peuvent être utilisées pour la chronolocalisation lorsque des changements dans une zone au fil du temps sont visibles sur l'imagerie satellitaire. Par exemple, l'imagerie satellitaire historique peut être utilisée pour préciser la période de construction ou de destruction des certains bâtiments.[43] En visualisant des images de différentes périodes, les analystes peuvent voir quand de nouveaux détails apparaissent ou disparaissent. Par exemple, l'imagerie satellitaire peut indiquer quand des bâtiments ont été incendiés ou des monuments culturels profanés,[44] ou peut montrer un sol perturbé, indiquant potentiellement la présence d'une fosse commune. Cependant, il est important de noter que l'imagerie satellitaire (en particulier celle qui est accessible au public, comme celle de Google Earth Pro) ne contient généralement pas d'images historiques claires pour chaque date spécifique et ne permet généralement qu'une analyse temporelle réduite à quelques mois.

43 Sam Dubberley et Joe Freeman, "Killings, corruption, land grabs: human rights violations against the Rohingya today" (*Amnesty International*, 25 août 2020) <https://citizenevidence.org/2020/08/25/rohingya-verification/>.

44 Benjamin Strick, 'Géolocalisation de la destruction des infrastructures au Cameroun: A Case Study of Kumbo and Kumfutu' (*Bellingcat*, 21 novembre 2018) <https://www.bellingcat.com/resources/case-studies/2018/11/21/geolocation-infrastructure-destruction-cameroon-case-study-kumbo-kumfutu/>; SITU Research, 'ICC Digital Platform: Tombouctou, Mali " <https://situ.nyc/research/projects/icc-digital-platform-timbuktu-mali>.

Figure H: Capture d'écran de la plateforme numérique développée par SITU Research pour la Cour pénale internationale décrivant la destruction de sites du patrimoine culturel au Mali. Les images montrent le site du mausolée El Kounti avant et après la destruction. (http://icc-mali.situplatform.com/).

Principaux enseignements: Les enquêteurs peuvent utiliser de nombreuses méthodes pour estimer la date à laquelle une image numérique a été prise, notamment l'analyse des métadonnées, l'analyse des ombres, les pistes d'images, l'analyse des conditions météorologiques historiques, l'analyse de l'heure et de la date de téléchargement, les recherches inversées d'images et de vidéos et les comparaisons d'images satellitaires. Toutes ces méthodes ont des limites. Les évaluations fiables du moment où une image numérique a été prise doivent prendre en compte et utiliser plusieurs méthodes d'analyse.

5.

Conclusion

L'imagerie numérique provenant de sources libres d'accès est de plus en plus utilisée par les tribunaux, les organes créés en vertu des traités relatifs aux droits humains et d'autres organismes d'établissement des faits. Elles peuvent constituer des éléments très probants pour l'évaluation médico-légale et judiciaire des violations présumées du droit international des droits humains, du droit international humanitaire et du droit pénal international, et peuvent être présentées par l'accusation et la défense dans les procès pénaux pour étayer leurs arguments. Cependant, la prévalence croissante des photos et vidéos provenant de sources libres d'accès en tant qu'éléments de preuve implique des risques d'interprétation erronée ou de confiance mal placée, soit dans les documents provenant de sources libres d'accès, soit dans l'analyse de l'enquêteur. Les techniques et les descriptions contenues dans ce guide sont conçues pour aider les juges et les enquêteurs à évaluer les documents numériques provenant de sources libres d'accès. En même temps, ce type de matériel fait généralement partie d'un ensemble plus large de preuves présentées au tribunal ou à l'organisme d'enquête. Les informations numériques provenant de sources libres d'accès doivent en fin de compte être évaluées selon les mêmes règles de preuve que celles généralement appliquées par l'institution ou le tribunal concerné, et sous réserve des charges et des normes de preuve établies par l'institution ou le tribunal.

Glossaire

Chronolocalisation: corroboration des dates et heures des événements décrits dans une information, généralement visuelle. Par exemple, il peut être possible de déterminer l'heure à laquelle une photographie a été prise en examinant la longueur des ombres créées par la lumière du soleil, ainsi que d'autres indicateurs.

Dark web: la partie d'internet qui n'est accessible qu'au moyen de logiciels spécialisés et qui permet aux utilisateurs et aux exploitants de sites web de rester anonymes et intraçables.

Deep web: la partie d'internet qui n'est pas indexée et qui n'est donc pas accessible par les moteurs de recherche.

Géolocalisation: identification ou estimation de l'emplacement d'un objet ou d'une activité, ou de l'emplacement à partir duquel un élément a été généré. Par exemple, il peut être possible de déterminer l'emplacement à partir duquel une vidéo ou une photographie téléchargée sur l'internet a été prise à l'aide de techniques de géolocalisation. Ces techniques peuvent comprendre, par exemple, l'identification de caractéristiques géographiques uniques dans une photographie avec leur emplacement réel sur une carte.

Information numérique provenant de sources libres d'accès: information publiquement disponible en format numérique, généralement acquise sur Internet.

Information provenant de sources libres d'accès: information que tout membre du public peut observer, acheter ou demander, sans avoir besoin d'un statut juridique particulier ou d'un accès non autorisé.

Intelligence artificielle (IA): branche de l'informatique consacrée au développement de programmes permettant aux machines d'apprendre à réagir à des variables inconnues et à s'adapter à de nouveaux environnements.

Médias synthétiques: également appelé médias génératifs, est défini comme un contenu visuel, auditif ou multimodal qui a été généré ou modifié par un algorithme (généralement par le biais de l'intelligence artificielle). Ces produits sont souvent très réalistes, ne seraient pas identifiables comme synthétiques par le commun des mortels et peuvent simuler des artefacts, des personnes ou des événements.

Métadonnées: ce sont des données sur les données. Elles contiennent des informations sur un fichier électronique qui sont soit intégrées, soit associées au fichier. Les métadonnées comprennent souvent les caractéristiques et l'historique d'un

fichier, comme son nom, sa taille et ses dates de création et de modification. Les métadonnées peuvent décrire comment, quand et par qui ou quoi un fichier numérique a été collecté, créé, consulté, modifié et formaté.

Pseudonymisation: le traitement de données à caractère personnel de telle sorte que les informations ne puissent plus être attribuées à une personne concernée spécifique sans l'utilisation d'informations supplémentaires.

Recherche inversée d'images/vidéos: Une recherche inversée consiste à télécharger une image ou une vidéo vers un moteur de recherche afin que l'algorithme de recherche puisse identifier d'autres copies de la même image ou d'images similaires sur l'internet. La limite d'une recherche inversée d'images est qu'elle ne scanne que la base de données d'un moteur de recherche, qui ne comprend qu'un petit pourcentage du contenu actuellement présent sur l'internet. Elle n'inclut pas, par exemple, le contenu du deep web (qui n'est pas indexé par les moteurs de recherche tels que Google) ou du dark web (la partie de l'internet à laquelle on ne peut accéder qu'au moyen d'un logiciel spécialisé, tel que le navigateur Tor).

Valeur de hachage cryptographique: calculs qui peuvent être effectués sur n'importe quel type de fichier numérique pour générer une chaîne alphanumérique de longueur fixe qui peut être utilisée comme preuve qu'un fichier numérique n'a pas été modifié depuis que son contenu a été haché. Cette chaîne reste la même à chaque fois que le calcul est effectué, tant que le fichier n'est pas modifié.

Vérification: se réfère au processus d'établissement de l'exactitude ou de la validité des informations qui ont été collectées en ligne. La source, le contenu et l'élément ou le fichier numérique doivent être considérés collectivement et comparés pour en vérifier la cohérence.

Remerciements

La traduction et la production ont été financées par l'ESRC Impact Acceleration Account de l'université de Swansea et par le projet TRUE de l'université de Swansea, financé par l'UKRI Frontier Research Grant EP/X016021/1. Ce travail a également été soutenu par l'Institute for Humanities and Social Sciences (IHSS) de l'université Queen Mary de Londres.

Nous tenons à remercier les personnes suivantes, qui ont toutes fait des commentaires judicieux sur divers aspects du texte à différents stades du processus de rédaction: **Dato' Shyamala Alagendra,** avocate **pénaliste international; Hadi Al Khatib,** directeur général, Mnemonic; **Siobhán Allen**, avocate principal, Global Legal Action Network (GLAN); **shirin anlen**, technologue des médias, WITNESS; **Pavlo Bogachenko,** associé principal, DLA Piper; **Son Excellence la juge Solomy Bossa**, juge, Chambre d'appel, Cour pénale internationale; **Jacobo Castellanos,** coordinateur, WITNESS; **Camille Chabot,** chercheuse, UC Berkeley School of Law Human Rights Center et Université de Pékin; **Son Excellence la juge Margaret De Guzman,** juge, Mécanisme international appelé à exercer les fonctions résiduelles des Tribunaux pénaux, **Dr Jeff Deutch,** chercheur principal, Mnemonic; **Sam Dubberley**; **Michael Elsanadi,** enquêteur de sources libres d'accès, Mnemonic; **Jessica Gavron,** directrice juridique, European Human Rights Advocacy Centre; **Dr Matthew Gillett**, maître de conférences, École de droit et des droits humains, Université d'Essex; **Jonathan Hak KC; Anne Hausknecht,** doctorante, projet TRUE, Université de Swansea; **Peter Haynes KC; Professeur Laurence R. Helfer**, professeur de droit, Duke University, membre du Comité des droits de l'homme des Nations unies; **Gabriele Juodkaite-Granskiene,** juge, Cour suprême de Lituanie; **Koen Kluissen,** inspecteur-détective et enquêteur de sources libres d'accès (crimes internationaux), ministère public néerlandais; **Son Excellence la juge Joanna Korner,** juge, Cour pénale internationale; **Philip Leach,** professeur de droit des droits humains, Middlesex University London; **Kateryna Latysh,** MSCA4Ukraine boursière postdoctorale, Vilnius University et professeure agrégée, Yaroslav Mudryi National Law University; **Nema Milaninia,** conseiller spécial de l'Ambassadeur itinérant des États-Unis pour la justice pénale mondiale; **Dearbhla Minogue,** avocate principale, Global Legal Action Network (GLAN); **Judy Mionki,** avocate de la défense, Cour pénale internationale; Région Afrique officier de liaison, Comité des droits de l'homme, Association internationale du barreau; **Yvonne Ng,** responsable du programme des archives, WITNESS; **Raluca Racusan; Son Honneur le juge Keith Raynor,** juge, Angleterre et Pays de Galles; **Yaroslavna Sychenkova,** consultante indépendante; **Professeur Yuval Shany,** Chaire Hersch Lauterpacht en droit international, Université hébraïque de Jérusalem; **Konstantina Stavrou,** doctorante, Université de Vienne; **Benjamin Strick,** Centre for Information Resilience; **Hryhorii Zhurakivskyi,** Mission d'assistance de l'Union européenne en Ukraine.

Merci également à tous les juges et anciens juges qui ont fait part de leurs commentaires mais qui ont souhaité rester anonymes.

Évaluation de l'imagerie numérique provenant de sources libres d'accès: *Un guide pour les juges et les enquêteurs*