International Studies Journal (ISJ)
Vol. 18, No. 3 (71), Winter 2022
Received Date: 2021/6/3
Accept Date: 2021/12/29
PP: 53-66

The Evidentiary Value of Social Media in International Criminal Proceedings

Geert-Jan Alexander Knoops Ph.D[1]

Abstract

As smartphones are currently used by an estimated 3.8 billion users, social media evidence, in conjunction with smartphone technology, is increasingly used to report on events taking place in international crime theatres. Especially since international crime theatres are not easily accessible to investigators, social media content appears to be increasingly favoured by fact-finders to create the evidentiary record of an international criminal trial. By examining the vulnerabilities and pitfalls of social media evidence, this article addresses the issue of the evidentiary value of social media content in international criminal proceedings. Section 2 describes the major risks associated with the reliance on social media evidence in international criminal trials, such as authentication or contextualisation. Section 3 analyses international criminal case law on the admissibility of social media evidence. Section 4 discusses broader concerns about social media evidence in international criminal trials, pertaining to interpretation bias and the principle of equality of arms.

Keywords

Social media evidence, Facebook evidence, International criminal court, Admissibility of social media evidence in criminal cases, Challenging social media evidence

[1] Attorney at law at Knoops' International Lawyers, Professor of Politics of International Law University of Amsterdam (UvA), Visiting Professor International Criminal Law Shandong University (China), Vice President of the European Innocence Network, Lead Counsel for the Defence at the International Criminal Court.

The author is indebted to the important assistance in preparing this article to: Ms Elsa Bohne, Intern for the defence team at the International Criminal Court and Ms Aline Petersen, Legal assistant at Knoops' Lawyers, who contributed to the composition and text of the article.

I. Introduction

In her article entitled "New technologies in international criminal investigations", dr. Rebecca J. Hamilton of the American University of Washington, College of Law, observes that "User-generated evidence[1] is changing the landscape of international criminal investigations and opening it up to new actors in ways that may ultimately be beneficial, but nonetheless involve a significant degree of risk". Especially in international crimes theatres which investigators cannot access, used-generated evidence could assist the prosecution in criminal cases in its fact-finding missions to obtain evidence and investigate crime scenes[2].

Bearing in mind that smartphones are currently used by an estimated 3.8 billion users,[3] social media evidence can appear anywhere in the world. Therefore, social media, in conjunction with smartphone technology is increasingly used to report on events taking place in international crimes theatres. For instance, the ongoing Syrian conflict – "characterized by the extensive use of online social media platforms by all sides involved"[4] – as well as the establishment of the International, Impartial and Independent Mechanism (IIM), reveal how social media content can be used to create the evidentiary

[1] Hamilton, Rebecca. (2018). *New Technologies in International Criminal Investigations*. Proceedings of the ASIL Annual Meeting. Page 131. Doi:10.1017/amp.2019.18. User-generated evidence is footage that an ordinary citizen records on their smartphone, in an effort to achieve accountability.

[2] *Ibid*. Page 132. Doi:10.1017/amp.2019.18.

[3] Number of smartphones Users Worldwide from 2016 to 2021 (in billions), STATISTA. Available at: http://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

[4] D. O'Callaghan, N. Prucha, D. Greene, M. Conway, J. Carthy and P. Cunningham, "Online social media in the Syria conflict: Encompassing the extremes and the in-betweens," *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (ASONAM 2014), Beijing, China, 2014, pp. 409-416, doi: 10.1109/ASONAM.2014.6921619. See also M. Kaylan, "Syria's War Viewed Almost in Real Time", *The Wall Street Journal,* (2013). Available at: http://www.wsj.com/articles/SB10001424127887324492604579083112566791956 (noting that between January 2012 and September 2013 "*over a million videos have been uploaded [on YouTube] with hundreds of views"].*

record in international criminal trials. For instance, recently, before the International Criminal Court (ICC), the Office of the Prosecutor (OTP) issued an arrest warrant against Libyan military commander Mahmoud Mustafa Busayf Al-Werfalli, based considerably on information derived from social media.[1]

Given the growing reliance on social media evidence in current or future international criminal proceedings, a number of authors have addressed the vulnerability of this new type of evidence. This raises the question as to whether this type of evidence is to be found admissible to be used as evidence to prove international crimes and if so, under which conditions. In particular, one should question, considering the serious nature of such crimes, whether more restrictive criteria for the admissibility of this type of evidence should be justified. This question is especially relevant since user-generated evidence, as pointed out by dr. Rebecca Hamilton, involves a significant degree of risk.

II. User-generated evidence and a "significant degree of risk"

Before addressing the potential risk of user-generated evidence within international criminal proceedings assessed in this article, we should first address some definitions of the relevant concepts in this respect. First, the term social media is often used to refer to new forms of media that involve interactive participation.[2] Though social media may take different forms, they can be characterized by two features. The first feature holds that social media are interactive Internet-based applications and the second that social media are fueled by the so called user-generated content. Users can access social media platforms via web-based apps on their laptop, or – more and more often – on

[1] *Prosecutor v. Mahmoud Mustafa Busayf Al-Werfalli*, Warrant of Arrest, ICC-01/11-01/17-2, 17 August 2017

[2] J. Manning, "Social media, definition and classes of", *Encyclopedia of social media and politics,* Thousand Oaks, CA: Sage (2014), page 1158.

their smartphones. The second concept that is relevant in this regard is Mobile Instant Messaging (MIM). These apps allow mobile users to "send real-time text messages, voice messages, picture messages, video messages, or files to individuals or groups of friends."[1] Facebook Messenger and WhatsApp are the most popular MIM apps, with up to 2 billion active users per month.[2] Finally, the term "deep fake" circulates within the area of user-generated evidence to refer to digital manipulation of sound, images, or video to impersonate someone or make it appear that a person did something – and to do so in a manner that is increasingly realistic, to the point that the unaided observer cannot detect the fake.[3] After having assessed the relevant terminology, this paragraph now addresses the major vulnerabilities of social media evidence in international criminal proceedings.

The first evidentiary pitfall of social media evidence in international criminal proceedings pertains to probative value. Indeed, the difficulty lies in reliability, and more specifically in authentication. In this respect, the issue of deep fake is a key challenge for social media evidence since "technology [now] enables even those with minimal technical skills to create forgeries that are undetectable to the lay eye."[4] The growing number of deep fakes reveals that social media evidence can easily be manipulated and tempered with. Recent law practice involving social media evidence reveals these risks of manipulation and the importance of the element of social media authenticity. In several United States cases the issue of authenticity became of main concern

[1] Y. Choi, *Mobile Instant Messaging Evidence in Criminal Trials,* 26 Cath. U. J. L. & Tech 1 (2017), page 1. Available at: http://scholarship.law.edu/jlt/vol26/iss1/3

[2] Number of Mobile Messenger Apps Users as of October 2020 (in Billions), STATISTA. Available at: http://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/

[3] R. Chesney & D. Citron, Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy?, LAWFARE (2018). Available at: https://lawfareblog.com/deep-fakes-looming-crisis-national-security-democracy-and-privacy

[4] Hamilton, Rebecca. (2018). *New Technologies in International Criminal Investigations*. Proceedings of the ASIL Annual Meeting. Page 131. Doi:10.1017/amp.2019.18.

and did impact the overall assessment by the judiciary. For instance, in *United States v. Vayner*, the defendant was convicted of a single count of unlawful transfer of false identification documents. Upon Appeal, a Ukrainian, resident in Brooklyn testified that the defendant gave him a forged birth certificate showing that this Ukrainian citizen was the father of a made-up infant daughter. The prosecution provided a printout of a webpage through a special agent of the State department's diplomatic security service, while the prosecution asserted that this was the defendant's profile page on VK.com, which is the Russian version of Facebook. On the basis of Federal Rule of Evidence 901, the defence objected against the admissibility of this VK.com page, arguing that it had not been properly authenticated. The Court rejected the defence's objection and admitted the VK.com page as evidence. However, the US Court of Appeals for the Second Circuit came to a different conclusion. It held that the government had failed to provide a sufficient basis to justify the conclusion that the printout belonged to the defendant. The Appeals Court accepted that the VK.com page reflected information about the defendant but was not convinced that the defendant had created the page himself. More specifically, the Court ruled that "the mere fact that a page with [the defendant's] name and photograph happened to exist on the Internet at the time of [the investigator's] testimony does not permit a reasonable conclusion that this page was created by the defendant or on his behalf."[1] However, the judgement of the Court of Appeals, did not define the required threshold for a proper authentication of a social media page. It observed though, that the mentioned Federal Rule 901 "required that there be some basis beyond [the defendant's] own testimony on which a reasonable juror could conclude that the page in question was not just any Internet page, but in fact Defendant's profile."[2]

[1] *United States v. Vayner*, 2014 WL 4942227 (2014 2d Cir.)
[2] *Ibid.*

The second element which might create evidentiary vulnerability of social media evidence in international criminal cases relates to contextualization. European national case law related to foreign terrorist fighters (FTFs) posing in front of dead bodies can be enlightening in various aspects. In these cases, the charges were based on Facebook posts where FTFs could be seen posing triumphantly next to enemies' dead bodies. In the end, the charged FTFs were convicted for war crimes of outrages upon personal dignity (art. 8(2)(c)(ii) Rome Statute).[1] The conclusion to draw from such case law is that these photos were deemed to be a sufficient ground for conviction for war crime of outrages upon personal dignity. However, these same photos in themselves were not to be seen as sufficient for conviction for war crime of murder, since they do not clarify anything about the element of perpetration. As a matter of fact, the key factor is contextualization. Indeed, the photos could tell that the charged FTF was inflicting degrading treatments to dead bodies; however the photos could not tell that the charged FTF had himself killed those persons whose dead bodies were exposed.[2]

The third potential risk associated with social media evidence in international criminal cases is that it implies the involvement of new actors in the investigatory landscape. Due to domestic regulations, social media platforms' hosts are more and more under an obligation to remove crime-

[1] District Court of Pirkanmaa, 18 March 2016, Judgment no. 16/112431; District court of Kanta-Häme, 22 March 2016, Judgment no. 16/112863; Scania and Blekinge Court of Appeal, 11 April 2017, Case B 3187-16; German Federal Court of Justice, 27 July 2017, judgment no. StR 57/17.

[2] Note that there is one case however where the charged FTF was convicted under Swedish law for crime against international law of extra-judicial execution. See *Prosecutor v. Haisam Omar Sakhanh*, Svea Court of Appeal, 31 May 2017, Case B 3787-16 (In addition to the Facebook photo where the defendant was posing triumphantly next to enemies' dead bodies, there was a YouTube video depicting the defendant as he was taking an active part in the extra-judicial execution of the persons whose dead bodies were exposed. In such a case, there was sufficient ground for conviction for crime against international law of extra-judicial execution, since the YouTube video allowed contextualization of the Facebook photo.)

related content. The issue is that this type of regulations may lead to "over-removal of content,"[1] and thus might create an obstacle to international criminal investigations.

As for MIM evidence more specifically, the latter two elements of content and context seem less relevant. MIM apps messages are easier to contextualize than photos and videos: included in a conversation thread, a message can be contextualized by scrolling the conversation up, and analyzing the content of previous messages. Likewise, messages are not exposed to over-removal by platforms' hosts since MIM conversations are most generally encrypted private conversations. However, MIM evidence still has to face the issue of authentication, since it is not self-authenticating. MIM evidence bears inherent risks. Some of these risks have been identified by the Court of Criminal Appeals of Texas in *Tienda v. State*, where the Court held: "computers can be hacked […] and cell phones can be purloined."[2]

With regards to hacking, a quick search on Google reveals that MIM accounts can be hacked rather easily (**see the screenshot below**). It appears that the easiest way of hacking someone's MIM account is by using a cell phone spy app, which can be bought online.

In addition, as the Court of Criminal Appeals of Texas' highlighted it, "cell phones can be purloined". Especially because "the applications are usually always turned on at all times while the mobile device is turned on", an illegitimate possessor of the mobile device may make use of a MIM app without having to log in with identifying credentials.

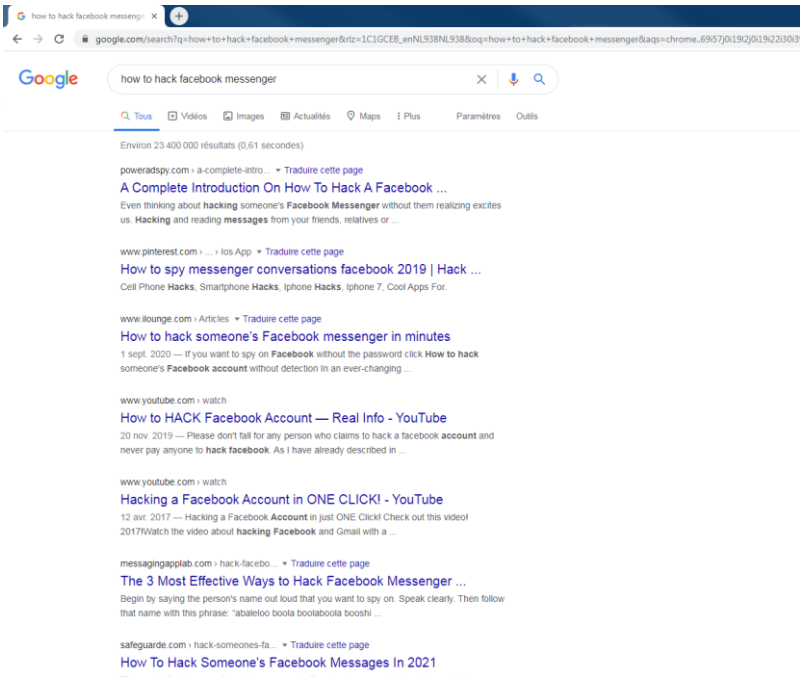
Beyond hacking and theft, the MIM app can merely be used by someone whom the owner of the MIM account allowed access using his/her credentials.

[1] Hamilton, Rebecca. (2018). *New Technologies in International Criminal Investigations*. Proceedings of the ASIL Annual Meeting. Doi:10.1017/amp.2019.18.

[2] *Tienda v. State*, 479 S.W.3d 863 (Tex. App. 2015)

In this respect, the Court in the case of *State v. Smith* rejected the Government's MIM evidence holding that "there [was] no evidence of whether, […], [the defendant] allowed others access using his password or any unique qualities regarding the messages themselves from which one may assert [that it was the defendant who] sent the messages."[1]

Thus, regarding the inherent risks associated to it, MIM evidence must be "independently authenticated" by corroborating elements, such as "the testimony of a knowledgeable witness or other accepted means."



For that reason, the Court of Criminal Appeals of Texas in the *Tienda* case has held that the circumstance that "the respondent in an internet chat room dialogue purports to identify himself" cannot be "without more" regarded as sufficient to support a finding of authenticity. Likewise, the US Court of

[1] *The State v. Smith*, 192 So.3d 836, at 842 (4h Cir. 2016)

Appeals for the Third Circuit[1] has rejected the Government's theory of self-authentication because the Government could only show that "the communications took place as alleged between the named Facebook accounts", but could not demonstrate that these accounts were actually used by their legitimate owner. On the same note, the Court in *State v. Smith* held that "[N]o evidence or testimony was offered as to whether [the defendant] created the account and/or profile on the social media platform or whether he had ever accessed the platform."[2] Therefore, it appears that "user identification is the key to authenticating MIM evidence."

With regards to MIM apps, user identification is a real challenge since "parties identify themselves by their user names". If usernames may consist of a unique name, they may also consist of a "set of characters", in other words, a pseudonym. The risk of anonymity associated with the use of pseudonyms on social media makes user identification more complicated than for traditional communication methods. For instance, in a phone conversation, "identifying the voices of a conversation typically authenticates the evidence."[3] Likewise, in an email communication, "the parties who exchange emails know each other personally or the server administrators know their identities"[4]; therefore, the parties to the conversation may identify each other, or the server administrator may produce the logs and IP addresses of the parties.

In order to easily identify its users, most MIM apps require a valid email address or a valid phone number to sign up for its service. Nevertheless, MIM users can circumvent those restrictions and anonymise themselves by resorting to disposable email addresses or phone numbers.

---

[1] *Tienda v. State*, 479 S.W.3d 863 (Tex. App. 2015)

[2] *The State v. Smith*, 192 So.3d 836, at 842 (4h Cir. 2016)

[3] Y. Choi, *Mobile Instant Messaging Evidence in Criminal Trials*, 26 Cath. U. J. L. & Tech 1 (2017), page 5

[4] *Ibid.*

Regarding all of these considerations, the following question may arise: should the standards for admissibility of social media evidence for international crimes be subjected to additional evidentiary safeguards? This question is most relevant since international criminal trials relate to the most serious charges within the world order. In addition, this question deserves special attention in light of article 69(4) of the Rome Statute, which reads "The Court may rule on the relevance or admissibility of any evidence, taking into account, *inter alia*, the probative value of the evidence and any prejudice that such evidence may cause to a fair trial or to a fair evaluation of the testimony of a witness, in accordance with the Rules of Procedure and Evidence''.[1] For instance, in the case of *The Prosecutor vs. Patrice Ngaïssona*, in February 2021 the ICC Prosecutor announced to rely in its case theory on "contemporaneous Facebook communications" by third parties, to prove certain key aspects of its case, such as for Mr Ngaïssona, the element of alleged order to certain elements or the alleged provision of equipment of ammunition. In this regard, the main vulnerability would be authentication; and thus the authenticity of the evidence should be subjected to a thorough admissibility test. The next paragraph will look into the international criminal case law regarding the admissibility of social media evidence.

III. International Criminal Case Law on Admissibility of Social Media Evidence

There has not yet been a case where the International Criminal Court (ICC) has fundamentally addressed the issue of the admissibility of social media evidence in international criminal cases. Yet, this type of evidence has been used and/or relied upon in several cases, such as in the case of *The Prosecutor v. Ahmad Al Faqi Al Mahdi ("Al Mahdi")*[2]. In this case the defendant was the

[1] *The Prosecutor v. Alfred Yekatom and Patrice-Edouard Ngaïssona* ICC-01/14-01/18, at 16.
[2] *The Prosecutor v. al Madhi*, ICC-01/12-01/15.

first ICC defendant to be charged with the destruction of cultural heritage as a war crime and the first to plead guilty before the ICC. The defendant was identified through videos which were posted on social media as promotion of Ansar Dine's attack on cultural heritage sites in Timbuktu.[1] The prosecution's evidence included YouTube video's and satellite images from before and after the destruction as well as archive photographs, audio and video recordings which showed the destruction at the time of the attack.[2] Trial Chamber VIII in its judgement, relied on a video to find that the charged attacks on mosques and mausoleums was an affront to the values of the Constitution of the UNESCO.[3] In the video, Al Mahdi had allegedly said: "It's probably the oldest mosque here in town, and is considered a heritage site […] a World Heritage Site. There are so many rumours relating to these shrines […]. Those UNESCO jackasses – this […] they think that this is heritage. Does 'heritage' include worshipping cows and trees?"[4] While this case would have been an opportunity for the Trial Chamber VIII to address the issue of admissibility of social media evidence in international criminal cases, the Chamber was not called upon to do so specifically, because the evidence was agreed upon by the parties and therefore the defence did not challenge the authenticity of the YouTube videos. Nonetheless, the Chamber did rely on social media evidence so as to conclude that the defendant was guilty in that case.

Also, in the case of *The Prosecutor v. Jean-Pierre Bemba Gombo, Aimé Kilolo Musamba, Jean-Jacques Mangenda Kabongo, Fidèle Babala Wandu and Narcisse Arido ("Bemba et al")*,[5] which is the first ICC case involving

[1] Freeman, Lindsay. (2018). Digital Evidence and War Crimes Prosecutions: The Impact of Digital Technologies on International Criminal Investigations and Trials. *Fordham International Law Journal, 41*(2), page 335.

[2] ICC-01/12-01/15-T-4-Red-ENG, pages 28- 29.

[3] *The Prosecutor v. al Madhi*, ICC-01/12-01/15.

[4] *Ibid.*

[5] *The Prosecutor v. Bemba*, ICC-01/05-01/13.

digital financial transactions as evidence, the Prosecution introduced into evidence photographs taken from a Facebook account in order to link individuals and corroborate other evidence. The defence challenged the Facebook evidence introduced by the Prosecution, namely by arguing that the photographs were not "prima facie authentic or reliable".[1] Despite the defence challenging the admissibility of Facebook evidence, Trial Chamber VII did not fundamentally address that issue. This might be explained by the fact that the Facebook photographs which were submitted to prove the defendant's guilt were already evidenced through witness testimony. Thus, as well as in the *Al Mahdi* case, the Chamber was not called upon to address the principle issue of the admissibility of social media evidence, since this type of evidence in these cases featured only in the context of a secondary and corroborating role in proving the elements of the crimes. These examples illustrate that social media evidence is already been used and argued in international criminal proceedings, in particular within the context of the prosecutor's case. It is therefore to be expected that this type of evidence will be invoked more increasingly in the future of the ICC and other international tribunals. It is therefore of importance, that the issue of admissibility of social media evidence should be thoroughly addressed by the ICC soon in order to obtain guidelines as to the boundaries of such admissibility.

IV. Other Pitfalls of Social Media Evidence

Apart from the mentioned admissibility arguments, pertaining to social media evidence in international criminal cases, two other potential risks arise which are aptly described by Dr Hamilton.[2]

[1] *The Prosecutor v. Jean-Pierre Bemba Gombo, Aimé Kilolo Musamba, Jean-Jacques Mangenda Kabongo, Fidèle Babala Wandu and Narcisse Arido*, ICC-01/05-01/13-1245-Red, at 83.

[2] Hamilton, Rebecca. (2018). *New Technologies in International Criminal Investigations*. Proceedings of the ASIL Annual Meeting. Doi:10.1017/amp.2019.18.

The first additional problem of user-generated evidence is the danger of confirmation bias. In itself, user-generated evidence is mostly presented in the context of circumstantial evidence whereby the authorship of the material has to be established. Human nature tends to interpret the information contained in user-generated evidence in conformity with the theory of the case the prosecution or defence has in mind. Visual material may well lead to visual biases that may distort the reality[1]. Accordingly, cognitive errors can easily occur when one interprets user-generated evidence. A second additional potential pitfall of this type of evidence relates to the principle of equality of arms in international criminal cases. The accessibility of this type of evidence is more problematic for a defence counsel as opposed to the prosecution. As Rebecca Hamilton observes: "While the developers of these apps believe that their software would be as useful to the defence as the prosecution, there are more users interested in holding perpetrators to account than there are users concerned with the fair trial rights of defendants. As a result, the emergence of user-generated evidence risks exacerbating already significant problems regarding equality of arms between prosecution and defence in international criminal trials."[2]

Equality of arms implies that both parties have equal access to the same material throughout the proceedings. It is a fact that for the defendant in criminal trials the obtainance of corporation by social media hosts such as Facebook is far more difficult compared to the corporation the prosecution can receive from these social media providers.[3]

---

[1] Hamilton, Rebecca. (2018). *New Technologies in International Criminal Investigations*. Proceedings of the ASIL Annual Meeting. Page 133. Doi:10.1017/amp.2019.18.

[2] *Ibid*. See also Jalloh, Charles, & DiBella Amy, "Equality of arms in international criminal law: Continuing Challenges", in *The Ashgate Research companion to international criminal law-critical perspectives* 251, 251-87 (William Schabas, Yvonne McDermott & Niamh Hayes eds. 2013).

[3] Hamilton, Rebecca. (2018). *New Technologies in International Criminal Investigations*. Proceedings of the ASIL Annual Meeting. Page 132. Doi:10.1017/amp.2019.18.

Finally, there is also a legitimacy argument. Although user-generated evidence might seem attractive as an alternative to prosecute suspects of international crimes in situations where other evidence is difficult to find, the invocation of this type of evidence – considering the abovementioned vulnerabilities and risks – questions the legitimacy of serious charges such as genocide, crimes against humanity and war crimes. The prosecution, resorting to social media evidence to prove these types of charges, might therefore undermine the acceptance of judgements which rely on social media evidence. It is therefore important to observe that in most cases where social media evidence was found admissible, this type of evidence in itself was not deemed sufficient to prove a criminal charge, requiring additional evidence to establish authenticity and prove that someone was the perpetrator.[1]

[1] *Renee vs State of Texas* 49SO.3D.248. (2010). http://caselaw.findlaw.com/tx-court-ofappeals/1608799.html).