Expert Analysis

How To Authenticate Social Media Evidence In Court

By Clifford Histed, Desiree Moore and DC Wolf

By Clifford Histed, Desiree Moore and DC Wolf June 27, 2019, 2:44 PM EDT

Law360 (June 27, 2019, 2:44 PM EDT) -- Prosecutors offer Facebook posts to show that a gang leader "green lighted" the hatchet killing of a homeless man for "snitching" on him.[1] A plaintiff in an internet stalking case offers the hundreds of abusive emails she received from anonymous senders after spurning the defendant's advances.[2] The government secures a conviction for illegal firearm possession by offering Facebook photos of the defendant with a .45 caliber pistol — but no physical evidence.[3]

Clifford Histed

Desiree Moore

DC Wolf

These cases illustrate how social media evidence has become an important feature of modern trial practice, just as it is of how we shop, work, eat, vote, watch TV and interact with one another. We can summon and use social media virtually instantly with smartphones — devices the U.S. Supreme Court recently called "almost a feature of human anatomy."[4] Given social media's pervasiveness in our culture, and the frequency with which people use it compared to other forms of communication, social media evidence is a broader and deeper trove of courtroom evidence than has ever been available before. At the same time, however, social media evidence is uniquely vulnerable to alteration or forgery, particularly as advances in technology allow so-called "bot" accounts to create social media content autonomously.[5]

A New Frontier Brings New Challenges

Offering instant messages, tweets and social media posts of all types at trial is now commonplace. Such evidence can be useful, for example, to prove a party's mental state or to prove that someone was in a given place at a given time — like on a ski slope days after an alleged injury.[6] Even before trial, social media may provide strategic value — for instance, if a plaintiff's statements on product review forums contradict the allegations in a consumer class action complaint, that could potentially help a defendant secure pretrial dismissal.

But while social media has improved our ability to tell the jury "what really happened," it also creates new challenges for how that story can be told. The jury cannot see evidence unless it is authenticated and admitted. Federal Rule of Evidence 901(a) (and numerous state analogs) requires the proponent of evidence to "produce evidence sufficient to support a finding that the item is what the proponent claims it is."

This standard imposes a relatively low bar, requiring "[o]nly a prima facie showing of genuineness ... ; the task of deciding the evidence's true authenticity and probative value is left to the jury."[7] Compared to a voicemail, a letter or even an email, however, authenticating social media evidence can be challenging due to "the ease with which a social media account may be falsified or a legitimate account may be accessed by an imposter."[8] Thus, lawyers must lay a foundation that addresses the "concern that someone other than the alleged author may

have accessed the account and posted the message in question."[9]

Courts sometimes disagree on what must be shown to satisfy this concern. Some impose a relatively high bar, requiring the proponent to all but eliminate the possibility of phony authorship.[10] Others hold that social media evidence is just like any other type of evidence,[11] requiring only the introduction of facts from which a reasonable juror could find that the evidence was created by the purported author.

We submit that the permissive approach aligns better with the text of Rule 901 and is thus correct.[12] Rule 901(a) requires only a preliminary showing that the evidence is what the proponent claims; this "does not require ... rul[ing] out all possibilities inconsistent with authenticity."[13] Evidence that an imposter created the content might be a basis for admitting the evidence conditionally under Rule 104(b) or for excluding it under Rule 403, but it should not affect whether Rule 901's threshold for authentication can be met.[14] Once the proponent presents enough evidence for a reasonable juror to find that the author was who the proponent asserts, evidence suggesting otherwise may affect the weight the jury gives the evidence but should not impact its admissibility.[15]

Even so, some courts continue to apply the more stringent approach.[16] For example, in United States v. Vayner, the U.S. Second Circuit Court of Appeals reversed a district court's decision to admit screenshots from a social media profile that contained the defendant's name, photo and work history.[17] Vayner holds that merely presenting evidence proving that a post came from a particular user's account is insufficient to authenticate the post as actually coming from that user.[18] Regardless of which approach is correct, lawyers cannot take for granted that courts will rule in their favor on evidentiary issues — particularly those involving complex technology and novel evidence in the heat of trial, amid numerous other evidentiary motions and objections.

Authenticating Social Media Evidence at Trial

Lawyers offering social media evidence at trial should be prepared to "over-authenticate" their evidence by laying a foundation that, if possible, substantially eliminates the possibility that an imposter created the content. If a witness will admit to authoring a post or owning a social media profile, and can lay a foundation supporting that admission, then the proponent's work should be done.[19] But in criminal cases (and even some civil ones), the Fifth Amendment may make this type of testimony unavailable if the witness believes that providing such testimony could be self-incriminating. And regardless, adverse witnesses often will simply be unwilling to admit they created a post or that they can remember doing so.

Authentication of social media evidence should thus rely on foundational testimony about three topics: (1) circumstantial evidence of authorship or account creation, (2) how the evidence was identified and verified (i.e., "chain of custody"), and (3) how the social media platform itself provides the evidence with indicia of reliability. Below we suggest three ways a proponent can provide this authentication.

Circumstantial Evidence of Authenticity

Witnesses can testify from personal knowledge about "contextual clues in the communication tending to reveal the identity of the sender."[20] This is the type of testimony that Rule 901(b) contemplates for circumstantially authenticating any type of communication. Consider the following lines of questioning.

Does the evidence contain information — photos, friends, locations, etc. — that is consistent with a witness's testimony about the asserted author or of how that person writes, speaks or behaves? For instance, in Allen v. Zonis, an internet-stalking in which one of the authors of this article was appellate counsel, the plaintiff testified that the writing style in abusive emails she received from anonymous senders matched that from messages the defendant had sent her previously.[21] And in Burgess v. State, a MySpace account bearing the name "Oops" was properly authenticated through an officer's testimony that he had confirmed with the defendant's sister that the defendant's nickname was "Oops."[22]

Have witnesses previously communicated with the asserted author using this profile? In Allen, the plaintiff's

authenticating testimony included the fact that she received the anonymous, threatening messages at an email account that only the defendant had ever used to communicate with her. This illustrates how linking a previously used communication channel with the purported author can be an effective means of establishing genuine authorship.

Does the post include a screen name or handle that is consistent with posts on other platforms that are more readily linked to the asserted author? For example, even if a Facebook page contains no photos or uses a false name, witnesses' testimony that the same name appears on other social media platforms containing visual depictions of the purported author can be sufficient to authenticate the Facebook page.[23]

Have the asserted author's offline activities ever corresponded to events or experiences described over social media? This can be a particularly persuasive way to authenticate social media evidence. Even a single instance where, for example, the purported author met with someone after arranging the encounter through social media can be enough to authenticate not only the messages arranging the encounter, but all messages coming from the account in question.[24]

Do timestamps or geolocation data associated with the post help connect it to particular people or events? Social media posts often contain information indicating the date, time and location of the post's creation.[25] Witness testimony that the purported author was in that location on that date can thus help authenticate the evidence. This type of data is not always accurate, however,[26] and attorneys should be prepared to offer testimony explaining any discrepancies.[27]

"Chain of Custody" Evidence

Offering testimony from investigators, electronic discovery specialists or expert witnesses can help authenticate social media evidence by establishing the evidence's "chain of custody," that is, how the proponent's investigation identified the information, verified it and led to its inclusion in the exhibit offered at trial. In particular:

How was the evidence identified and then copied, reproduced or transcribed into the exhibit being offered in court? This testimony should include a description from the witness of how the evidence was accessed and turned into an exhibit. For instance, an investigator could testify to accessing a particular website or app, taking a "screen shot" of the device's monitor and printing out the screen shot. A percipient witness can then testify as to whether the printout fairly and accurately reflects the social media evidence that the witness initially saw.

Do IP addresses or social media subscriber records link the evidence to a particular person? Social media companies may be compelled to disclose certain records in response to a subpoena, including subscriber information, which contains phone numbers or emails linked to a social media account, and IP address logs. Social media companies will generally also provide a certification from an authorized records custodian to establish a self-authenticating business record under Fed. R. Civ. P. 902(11).[28] Note, however, that this certification establishes only "that the depicted communications took place between certain Facebook accounts, on particular dates, or at particular times," which is not sufficient in isolation to authenticate the content of a social media post in relation to a particular author.[29]

What steps were taken to rule out other accounts with the same or similar usernames? Commonwealth v. Mangel affirmed the trial court's denial of the prosecution's motion in limine to admit Facebook communications where, among other things, a search on Facebook for the defendant's name yielded five profiles under that name, contradicting a detective's testimony that only one such account appeared during her search.[30] This illustrates the importance of using multiple avenues to authenticate evidence; an investigator's testimony about chain of custody may be insufficient in isolation if multiple profiles use the same name.

Did the proponent obtain the account's username and password to verify the source of the evidence? The trial court in Mangel faulted the prosecution for not obtaining the username or password for the Facebook account at issue to confirm its authenticity. To the extent available, obtaining login credentials for a social media account

— which, in theory, only the account's true owner should possess — is a reliable means of authenticating the social media account. However, given the intimacy and breadth of personal information often contained in social media accounts, courts may be wary about compelling parties to produce their login credentials, particularly in civil cases.[31]

Were social media apps on devices in the asserted author's possession logged in to accounts associated with the evidence at issue? In United States v. Lewisbey, the court held that incriminating Facebook posts were properly authenticated because (among many other circumstantial links between the defendant and the Facebook account) the Facebook app on a mobile phone confiscated from the defendant was linked to from which the incriminating statements were posted.[32] Likewise, in an internet child pornography case tried by one of the authors of this article, a computer in the defendant's bedroom was logged into AOL Instant Messenger at the time of his arrest under a screenname involved in chat logs discussing child pornography.[33] As mentioned above, in theory, only the true owner of a social media account has the means to access that account. So the fact that an account is accessible on a device in the asserted author's possession is a particularly strong indicator of genuine authorship.

Technological Safeguards of Authenticity

Background information about a social media platform's operation can explain how the platform, by design, seeks to guard against phony content. This might require testimony from an expert or from a representative of the social media company. For example:

- Do the platform's terms of service prohibit using false or invented profiles?
- Does the platform require users to create accounts using unique login credentials?
- Must users verify their accounts using email confirmation, two-factor authentication or other additional layers of security?[34]
- In the witness's training or experience, how often has evidence of this type proven to be fraudulent, and what would one expect to see if that were the case?

Eliciting testimony on these issues in isolation likely will not be sufficient to authenticate the substance of a social media communication. But covering all three of the areas discussed above — circumstantial evidence of authorship, chain of custody and the operation of the platform — will help ensure that social media evidence is properly authenticated. Authentication is supposed to be a lenient standard. Once the proponent meets the low bar of authentication, arguments to the contrary should go to the weight to be given the evidence rather than to its admissibility, and it should ultimately be up to the trier of fact to accept or reject such evidence.

The Next Frontier: Even More Challenges

Two years ago, researchers used many hours of video from Barack Obama's weekly address to teach an artificial intelligence program to map spoken-word audio onto video of mouth shapes. Researchers then used the program to create a photorealistic video of Obama appearing to speak the words from an audio clip of the researchers' choosing.[35] These techniques can be used to make convincing videos, known as "deepfakes," of people appearing to say just about anything.[36] Similar technology is being used to create photorealistic images of people who do not exist,[37] and to paint public figures such as Facebook CEO Mark Zuckerberg or House Speaker Nancy Pelosi in an unflattering light.[38]

While this technology is not yet widespread, other types of digital deception already are,[39] with one study estimating that between 9% and 15% of all Twitter users were not people but "bots," software-controlled accounts "algorithmically generating content and establishing interactions."[40] The capacity to create convincing forgeries of social media content likely will continue to increase.

While authentication under the rules of evidence is a lenient standard, it must be scrupulously applied as the pervasiveness of digital fakery increases. Lawyers must be creative and thorough in authenticating social media evidence, presenting information not only linking evidence to an asserted author, but also tending to rule out links to potential imposters. Likewise, lawyers opposing the admission of evidence should require the proponent to demonstrate that evidence is not fabricated. For example:

- Is there any reason to think someone other than the asserted author would have the desire, means and opportunity to falsely create the evidence?

- Do the social media company's records indicate that the account in question was affected by a data breach, and if so, has the account's password been changed since then?

- Are there any identifiable instances in which someone other than the asserted author posted to the account in question? Were any necessary remedial steps taken, and were those steps documented?

- Does the platform actively review content in an effort to identify and remove false or misleading posts? [41] How, how often, and are such efforts documented?

- Is there any forensic evidence that indicates the evidence has been tampered with?[42]

As technology with the potential for deceptive applications becomes cheaper and more widespread, addressing courts' concerns about authenticity will be essential. Presenting testimony from experts who understand digital fakes and are adept at identifying them may become an informal requirement. These concerns will be particularly important in criminal cases to ensure that the government does not knowingly or unknowingly use adulterated evidence to prove criminal culpability.

Clifford C. Histed and Desiree F. Moore are partners and Daniel-Charles Wolf is an associate at K&L Gates LLP.

Disclosure: DC Wolf was appellate counsel in Allen v. Zonis.

[1] See Colorado v. Glover, 363 P.3d 736 (Colo. Ct. App. 2015).

[2] See Allen v. Zonis, No. 76768-2-I, 2018 WL 6787925, at *11 (Wash. Ct. App. Dec. 24, 2018) (unpublished).

[3] See United States v. Farrad, 895 F.3d 859 (6th Cir. 2018).

[4] United States v. Carpenter, 585 U.S.—, 138 S. Ct. 2206 (2018).

[5] See, e.g., Onur Varol et al., Online Human-Bot Interactions: Detection, Estimation, and Characterization, arXiv:1703.03107v2 [cs.SI] (Mar. 27, 2017), https://arxiv.org/abs/1703.03107v2 (not peer-reviewed).

[6] See generally Hon. Paul W. Grimm et al., Authentication of Social Media Evidence, 36 Am. J. Trial Adv. 433, 437–38 (2013).

[7] United States v. Fluker, 698 F.3d 988, 999 (7th Cir. 2012); see also United States v. Jones, 107 F.3d 1147, 1155 n.1 (6th Cir. 1997) ("The [authentication] rule requires only that the court admit evidence if sufficient proof has been introduced so that a reasonable juror could find in favor of authenticity or identification. The rest is up to the jury.") (quoting 5 Jack B. Weinstein et al., Weinstein's Evidence ¶901(a), at 901–19 (1996)).

[8] United States v. Browne, 834 F.3d 403, 412 (3d Cir. 2016).

[9] Griffin v. State, 19 A.3d 415, 423 (Md. 2011).

[10] See, e.g., id.

[11] See, e.g., Tienda v. State, 358 S.W. 3d 633, 634–35 (Tex. Crim. App. 2012).

[12] See Grimm, supra note 6, at 455–56 (describing these two approaches and concluding that the latter — which lets the party opposing authentication rebut the proponent's showing and lets the court admit the evidence conditionally if a reasonable jury could find either way — is superior).

[13] United States v. Blanchard, 867 F.3d 1, 6 (1st Cir. 2017) (district court did not err in admitting prostitution adds from Backpage.com; despite discrepancies between image metadata and authenticating witness testimony, a witness testified to creating and posting the ad, and other testimony corroborated the timing and location of the posting).

[14] Fed. R. Evid. 104(b) ("Relevance That Depends on a Fact.When the relevance of evidence depends on whether a fact exists, proof must be introduced sufficient to support a finding that the fact does exist. The court may admit the proposed evidence on the condition that the proof be introduced later."); Fed. R. Evid. 403 ("The court may exclude relevant evidence if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence.").

[15] See, e.g., Allen, 2018 WL 6787925, at *11 (". . . Zonis's argument that Allen also wrote e-mails in a similar manner, specifically using all caps, is contrary evidence that goes to weight, but not authentication or admissibility.").

[16] See, e.g., Richardson v. State, 79 N.E.3d 958, 962–64 (Ind. Ct. App. 2017) (affirming exclusion of evidence even though messages came from Facebook Messenger app on password-protected phone recovered from victim's body, state's authenticating witness admitted not knowing who wrote the message and that Facebook messages could be sent through another device logged into the same account).

[17] United States v. Vayner, 769 F.3d 125, 131 (2d Cir. 2014). But see Burgess v. State, 742 S.E. 2d 464, 467 (Ga. 2015) (reaching a different result under similar facts by holding that screenshots from a Myspace profile were properly authenticated where officer testified he had confirmed with defendant's sister that defendant went by the nickname shown on the profile, and where photos on Myspace were consistent with other photos of

defendant).

[18] Commonwealth v. Mangel, 181 A.3d 1154, 1162 (Pa. Super Ct. 2018) (citing Vayner, 769 F.3d at 131).

[19] See, e.g., Stout v. Jefferson Cty. Bd. of Ed., 882 F.3d 988, 1008 (11h Cir. 2018) (Facebook posts were properly authenticated when alleged authors admitted to creating them).

[20] Mangel, 181 A.3d at 1162.

[21] See Allen, 2018 WL 6787925, at *10–12.

[22] Burgess, 742 S.E. at 467.

[23] See Cotton v. State, 773 S.E. 2d 242, 245 (Ga. 2015) (incriminating Facebook messages were properly authenticated where witness testified, inter alia, that she had seen videos depicting the defendant on YouTubeunder the same screen name associated with the messages and saw that the defendant's friends and family were Facebook friends with an account under the same alias).

[24] See Commonwealth v. Foster F., 20 N.E. 3d 967, 971 (Mass. App. Ct. 2014) (where the juvenile defendant "appeared on January 28 to play a dating game with the victim . . . exactly as the person sending messages from the Juvenile's Facebook account had proposed," Facebook messages sent after the July 28 sexual assault, which contained incriminating admissions, were properly authenticated).

[25] See Blanchard, 867 F.3d at 6.

[26] For instance, virtual private network, or "VPN" services can be used to make it appear as though one's computer is located somewhere other than its true location. Likewise, sending emails across different time zones can sometimes affect the accuracy of the times listed in the email chain.

[27] See id.

[28] See Farrad, 895 F.3d at 865–66.

[29] Id. (citing Browne, 834 F.3d at 410–11).

[30] See Mangel, 181 A.3d at 1163.

[31] See John G. Browning, With "Friends" Like These, Who Needs Enemies? Passwords, Privacy, and the Discovery of Social Media Content, 36 Am. J. Trial Advoc. 505 (2013) (discussing courts' differing approaches in addressing motions to compel social media login credentials).

[32] United States v. Lewisbey, 843 F.3d 653, 658 (7th Cir 2016).

[33] See Brief of Appellee at 4–5, United States v. Allen, 605 F.3d 461 (7th Cir. 2010) (No. 09-2539).

[34] See, e.g., May Elliott, Two-factor Authentication: How and Why to Use It, CNET (March 28, 2017, 3:51 PM PDT), available at https://www.cnet.com/how-to/how-and-why-to-use-two-factor-authentication/(explaining two-factor authentication).

[35] Supasorn Suwajankorn et al., Synthesizing Obama: Learning Lip Sync from Audio, ACM Trans. Graph. 36, 4, Article 95 (July 2017), available at https://doi.org/10.1145/3072959.3073640.

[36] James Vincent, Watch Jordan Peele use AI to make Barack Obama Deliver a PSA about Fake News, The Verge - TL;DR (Apr. 17, 2018 1:14 PM EDT), https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-

news-video-barack-obama-jordan-peele-buzzfeed.

[37] James Vincent, ThisPersonDoesNotExist.com uses AI to Generate Endless Fake Faces, The Verge - TL;DR (Feb. 15, 2019, 7:38 AM EST), https://www.theverge.com/tldr/2019/2/15/18226005/ai-generated-fake-people-portraits-thispersondoesnotexist-stylegan.

[38] Allyson Chiu, Facebook Wouldn't Delete an Altered Video of Nancy Pelosi. What About One of Mark Zuckerberg?, Washington Post (June 12, 2019, 6:32 AM), https://www.washingtonpost.com/nation/2019/06/12/mark-zuckerberg-deepfake-facebook-instagram-nancy-pelosi/.

[39] The doctored video of Nancy Pelosi was not an AI-created "deepfake," but merely a slowed, pitch-altered video clip in which Ms. Pelosi appeared to be drunkenly slurring her words, illustrating how misleading evidence can be created easily and distributed quickly even without sophisticated technology. See Ian Bogost, Facebook's Dystopian Definition of 'Fake', The Atlantic (May 28, 2019), https://www.theatlantic.com/technology/archive/2019/05/why-pelosi-video-isnt-fake-facebook/590335/.

[40] Onur Varol et al., Online Human-Bot Interactions: Detection, Estimation, and Characterization, arXiv:1703.03107v2 [cs.SI] (Mar. 27, 2017), https://arxiv.org/abs/1703.03107v2 (not peer-reviewed).

[41] See, e.g., Christine Fisher, Facebook Fact Checkers Will Soon Review Instagram Posts: Instagram Wants to Limit the Reach of False Posts and Misinformation, Engadget (May 6, 2019), https://www.engadget.com/2019/05/06/instagram-facebook-fact-checkers-misinformation/.

[42] See Jonathan Mraunac, The Future of Authenticating Audio and Video Evidence, Law360 (July 26, 2018, 12:57 PM EDT), https://www.law360.com/articles/1067033/the-future-of-authenticating-audio-and-video-evidence (discussing the idea that video and audio recording devices could encode uneditable encrypted digital signatures on recordings, "similar to the ballistic markings left on a bullet by the barrel of a firearm"); Jennifer Langston, Lip-Syncing Obama: New Tools turn Audio Clips into Realistic Video, UW News (July 11, 2017), https://www.washington.edu/news/2017/07/11/lip-syncing-obama-new-tools-turn-audio-clips-into-realistic-video/ (discussing Suwajankorn et al., supra, and stating that "[b]y reversing the process—feeding video into the network instead of just audio—the team could also potentially develop algorithms that could detect whether a video is real or manufactured").