Artificial Intelligence (AI)
Interim Guidance

June 2024

from the **AI Rapid Response Team** *at the National Center for State Courts*

AI and the Courts:
Digital Evidence and Deepfakes in the Age of AI

AI advances are causing challenges in the courtroom as judges grapple with evidentiary issues related to digitally enhanced evidence as well as the emergence of deepfakes (convincing false pictures, videos, audio, and other digital information). These advances make it easier and cheaper to enhance digital evidence or create deepfakes causing evidentiary issues to arise.

Digitally Enhanced Evidence

Digitally enhanced evidence is audio, videos, or images that have been enhanced by AI software. The purpose is generally to improve the quality of audio, videos, or images. This differs from past uses, such as zooming in on an image, speeding up or slowing down a video, or separating a voice from background noise, in that AI may fill in pixels on the image with what the software thinks should be in the image, thus altering it from the original.

This technology was recently at the center of a criminal trial in Washington state when digitally enhanced video was not admitted into evidence. The court based its decision on the testimony of the expert witness who testified "the AI tool(s) utilized ... added approximately sixteen times the number of pixels, compared to the number of pixels in the original images to enhance each video frame, utilizing an algorithm and enhancement method unknown to and unreviewed by any forensic video expert." The court found that the expert "demonstrated that the AI method created false image detail and that process is not acceptable to the forensic video community because it has the effect of changing the meaning of portions of the video."

It may be necessary for courts to consider changes to the rules of evidence but until that happens, Judges may need to require expert testimony to determine the authenticity and reliability of audio, videos, and images that are challenged rather than relying on the standards for admission.

What is a Deepfake?

"Deepfake" refers to fabricated or altered but realistic audio, videos, or images made using software, for example, by embedding another person's likeness into an image or video. Deepfakes have become very sophisticated in recent years, and it is not easy for an average person to identify the audio, video, or image as fake.

Deepfakes and the Courts

The issue of deepfakes can arise in any court proceeding in which a party presents digital evidence in the form of an image, video, or audio. Fabricated evidence could be submitted as authentic evidence or authentic evidence could be challenged as fabricated evidence. When a party alleges that digital evidence has been fabricated, expert testimony may be needed to authenticate the challenged evidence. This could result in a battle between the experts and higher litigation costs for all parties and could widen the access to justice gap.[1]

[1] Delfino, Rebecca, Pay-to-play: Access to Justice in the Era of AI and Deepfakes (February 10, 2024). Loyola Law School, Los Angeles Legal Studies Research Paper No. 2024-08.

Of concern is the effect that deepfakes could have on the case's outcome because of the considerable impact that visual evidence has on fact finders. According to studies referenced in a recent law journal article, as compared to jurors who hear just oral testimony, "jurors who hear oral testimony along with video testimony are 650% more likely to retain the information."[2] Once jurors have seen video evidence, it is very hard for the impact to be undone, even with admonishments to the jury. Another study published in 2021 by the Center for Humans and Machines at the Max Planck Institute for Human Development and the University of Amsterdam School of Economics, demonstrates the difficulty of identifying deepfakes. The study found that the participants could not reliably detect deepfakes. The study found that people are biased towards identifying deepfakes as authentic (not vice versa) and overestimate their own abilities to detect deepfakes even after being instructed on how to detect deepfakes.[3] The mere existence of deepfakes combined with proliferation of online information, both real and fabricated, that people are exposed to daily may also lead to jury skepticism because people do not know what information they can trust.[4]

Current Evidentiary Rules

The existing Federal Rules of Evidence and the various state rules of evidence require that any evidence submitted must be real and that the party submitting the evidence has the obligation to authenticate it, by proving that the evidence is what it purports to be. Judicial officers already have an obligation to determine whether the probative value of the evidence submitted outweighs the possible unfair prejudice, confusion of the issues, or misleading of the jury that would result from its admission.

Are the Current Rules Sufficient?

Prior to the advent of deepfakes, the rules of evidence have been sufficient to adapt to technology changes. Laws and rules of evidence addressing deepfakes lag behind the technology. At present, tools to detect deepfakes are not as sophisticated as the tools to create deepfakes such that not all deepfakes will be identifiable. To mitigate the impact of deepfakes on litigation and jurors, judicial officers should identify related evidentiary issues and rule on those prior to trial and outside the presence of the jury, if possible.

The legal community is having ongoing discussions about the need for changes to the rules of evidence. However, it will be important for the courts to address the potential for harm to the legal process that deepfakes pose, and to evaluate whether more stringent rules should be adopted for the admission of digital evidence. In addition, for case types with high rates of self-representation, relying on the parties to challenge the authentication of evidence, which the current adversarial process requires, may be unrealistic. If deepfakes proliferate, courts may need to reconsider who is responsible for determining whether evidence is authentic, especially if reliable technology tools become available that would enable courts to determine if something is real or fake. If deepfakes become ubiquitous, the perception may shift to believing every piece of evidence is fake or has been altered; if so, this may require a more arduous authentication process routinely involving experts, costs, new technologies, elongating the length of trials. This would be a significant shift from current practices.

[2] Rebecca A. Delfino, Deepfakes on Trial: A Call To Expand the Trial Judge's Gatekeeping Role To Protect Legal Proceedings from Technological Fakery, 74 HASTINGS L.J. 293 (2023).

[3] Köbis NC, Doležalová B, Soraperra I. Fooled twice: People cannot detect deepfakes but think they can. iScience. 2021 Oct 29;24(11):103364. doi: 10.1016/j.isci.2021.103364. PMID: 34820608; PMCID: PMC8602050.

[4] Rebecca A. Delfino, Deepfakes on Trial: A Call To Expand the Trial Judge's Gatekeeping Role To Protect Legal Proceedings from Technological Fakery, 74 HASTINGS L.J. 293 (2023).