Deepfakes and the Legal Profession

Here are tips to navigate a world where images, video, and audio can easily deceive anyone, including crime victims, witnesses, lawyers, jurors, and judges.

BY KRISTOPHER TURNER

The internet lit up in late March 2023 when an image of Pope Francis wearing a puffy jacket was posted across social media. The image went viral and many people found it delightful. Soon, however, it was revealed that the "puffy pope" was an AI-generated image created by the tool Midjourney.[1] This incident illustrates how easy it is to be fooled by a "deepfake" (artificially generated) photo.

There have been countless instances of deepfakes in the months since the puffy pope jacket debuted, ranging from audacious scams with a deepfake video of Elon Musk promoting a new (and fake) business opportunity[2] to a deepfaked phone call from a person identified in the call as Joe Biden to primary voters in New Hampshire telling them not to vote.[3]

As technology evolves and improves, truth and fiction may become even more difficult to distinguish. Legal professionals must be prepared to deal with evidence that may have been artificially generated, speak to jurors who are dubious of evidence presented in court, and help prepare courtroom experts to educate others about how to tell deepfake media from legitimate media.

Deepfakes: What They Are and Why They Are a Problem

Manipulated media is not a new phenomenon, but the technology behind the newest wave is more advanced than ever. Deepfakes – images, video, or audio that has been created with advanced deep-learning machine models – are the result of the generative AI (GenAI) evolution that has changed how content is created and work is completed. With the development and release of GenAI such as ChatGPT and image-generation tools such as DALL-E, the ability to easily (and cheaply) generate entirely new content has become democratized. Anyone with $20 for a monthly subscription can now access high-quality image generators that can, if used with a small amount of training and skill, confuse or mislead incautious viewers and listeners. Combined with decreased trust in news organizations over the past decade, deepfakes can further undermine that trust by creating more confusion.

Effect on the Legal Profession

Possible effects of deepfakes on lawyers' practices and the judicial system are unknown but could be ominous. Deepfakes can confuse a casual news reader or social media scroller; they might have a similar effect in a courtroom or during other stages of litigation. Evidence, both

Kristopher Turner, U.W. 2020, is Associate Director of Public Services at the University of Wisconsin Law Library, Madison. Access the digital article at www.wisbar.org/wl.

kris.turner@wisc.edu

how it is processed and how it is interpreted, can become more complicated. As lawyers begin to process photos or other media that could be entered into evidence, it likely will become more difficult to tell whether an image or other piece of evidence is legitimate or has been modified in some way. Attorneys must continue to consider the source and chain of evidence (law enforcement, witness phone video, social media) and be prepared to defend evidence against allegations that it was artificially generated.

Many legal scholars, judges, and practicing attorneys have suggested potential updates to the rules of evidence to deal with the risks posed by deepfakes,[4] but none have been adopted yet. Forensic analysts can help trace how deepfakes were developed, but for the most part, these experts have been used by law enforcement agencies to track down the perpetrators who created deepfakes. Lawyers might need to employ these types of experts to prove that evidence belongs in, or should be excluded from, a case. These experts could be used both to vet evidence before the case begins and during the case itself to settle any questions for jurors about the evidence's legitimacy.

Judges can exert some agency in combatting deepfakes by requiring authentication of evidence, but this runs the risk of slowing down the legal process and potentially contributing to the degradation of public trust in fact-finding institutions by shifting some of the power from jurors to judges. However, requiring authentication would diminish the potential for the misuse of GenAI technology, for example, by attorneys who see the potential for sowing doubt with jurors by suggesting that a key piece of evidence is not real or who bring frivolous litigation based on nonexistent events "documented" by deepfakes.

The fundamental concern of loss of trust in the courts because facts are no longer relied upon might seem far-fetched, but deepfakes are already creating that same distrust in news media.[5] Meanwhile, in an effort to rein in the worst of the potential deepfake abuses, President Biden issued an executive order in 2023[6] focused on the safe and trustworthy development of AI. One thing is certain – just as with other emerging technology that has entered the legal arena, ignoring GenAI and deepfakes is not the answer.

Identifying and Working with Deepfakes

The best way to maintain current awareness with deepfakes is to continue to learn about the newest GenAI tools and their capabilities. The Wisconsin Rules of Professional

Conduct[7] require that attorneys stay abreast of the benefits and risks of relevant technology, and GenAI fits into this category. Even if lawyers won't be using deepfake technology directly, the products of this technology likely will soon play a role in their professional (and personal) lives.

Legal professionals must be prepared to deal with evidence that may have been artificially generated, speak to jurors who are dubious of evidence presented in court, and help prepare courtroom experts to educate others about how to tell deepfake media from legitimate media.

Simply being aware of the possibility of deepfake technologies is an important first step; the equally vital second step is testing yourself to better understand how advanced the technology is the equally vital second step. For example, you can take quizzes to see if you can tell the difference between real images and AI-generated ones. The *New York Times* created one in January 2024,[8] but other freely accessible sites are dedicated to this type of training as well.[9]

Tips to Detect Deepfakes

Here are some ways to better detect deepfakes:

- For videos, pay special attention to people's faces. Often, videos are focused on facial transformations (adding someone's face to another body or modifying their mouths to fit a new speech pattern). Check for odd movements, too much blinking, glasses with no reflections, and other small tells.
- For audio, listen for typical human pauses and placeholder words like "uh," "um," and "ah." Typically, AI audio will add too many or none at all. AI is especially poor at adding pauses when a human speaker would be considering their next word; AI has already generated the text that will be spoken.
- For images, the devil is in the details. Watch for small items, like a missing watch chain, button, or belt buckle, an additional digit on appendages, background that isn't centered correctly, or a subtle eye-color or eye-shape imperfection.
- For all deepfakes, consider if they seem "too perfect" — some AI-generated images have an electric sheen or shiny outlines on the objects they are trying to represent. This is more common in less advanced tools but is a reliable giveaway.
- Possibly most important, think about the context of the image. Take a moment and approach the media with skepticism. Are you seeing what you want to see? Is the image plausible in a vacuum? Seek a second source to confirm or debunk the media and consider the bigger picture from there.

Conclusion

Deepfakes represent both a technological leap and a challenge for factfinders and courts. As with all technology, deepfakes and GenAI generally can be used for good or ill. Education and training are necessary to ensure that legal professionals are not caught by surprise in courtrooms or in their personal lives.

One piece of advice for these emerging tools is to keep in mind that today is the worst day for AI. Every day going forward, it will continue to evolve. Until stronger regulations are in place, skepticism of media that is too good (or perhaps too amusing) to be true should be the guideline. Lawyers should continue to educate themselves and others and maintain awareness of the newest advancements of technology. If they do, deepfakes will remain a mere amusement and nothing more. **WL**

ENDNOTES

[1] Simon Ellery, *Fake Photos of Pope Francis in a Puffer Jacket Go Viral, Highlighting the Power and Peril of AI*, CBS News (March 28, 2023), https://www.cbsnews.com/news/pope-francis-puffer-jacket-fake-photos-deepfake-power-peril-of-ai/.

[2] Stuart A. Thompson, *How 'Deepfake Elon Musk' Became the Internet's Biggest Scammer*, N.Y. Times (Aug. 14, 2024), https://www.nytimes.com/interactive/2024/08/14/technology/elon-musk-ai-deepfake-scam.html (behind paywall for some readers).

[3] Shannon Bond, *A Political Consultant Faces Charges and Fines for Biden Deepfake Robocalls*, NPR (May 23, 2024), https://www.npr.org/2024/05/23/nx-s1-4977582/fcc-ai-deepfake-robocall-biden-new-hampshire-political-operative.

[4] Judge Herbert B. Dixon Jr., *The "Deepfake Defense": An Evidentiary Conundrum*, Judges' J. (June 11, 2024), https://www.americanbar.org/groups/judicial/publications/judges_journal/2024/spring/deepfake-defense-evidentiary-conundrum/.

[5] Sam Reardon, *How Deepfakes Are Impacting Public Trust in Media*, Pindrop (Oct. 17, 2024), https://www.pindrop.com/blog/deepfakes-impacting-trust-media.

[6] Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[7] *See* SCR 1.1 cmt. 8. The rules are available at https://www.wicourts.gov/sc/rules/chap20a.pdf.

[8] Stuart A. Thompson, *Test Yourself: Which Faces Were Made by A.I.?*, N.Y. Times (Jan. 19, 2024), https://www.nytimes.com/interactive/2024/01/19/technology/artificial-intelligence-image-generators-faces-quiz.html.

[9] *See, e.g.*, Northwestern Kellogg, *Detect Fakes: AI-Generated or Real?*, https://detectfakes.kellogg.northwestern.edu/ (last visited Dec. 6, 2024); *Real or AI? The Game*, https://real-or-fake-the-ai-game.onrender.com/ (last visited Dec. 6, 2024). **WL**