Deepfakes in the Courtroom:
A New Evidentiary Challenge

Riana Pfefferkorn
TILT @ Silicon Flatirons
February 10, 2021

Image credits: Indiana State Museum (left); University of Washington (right)

Stanford University

Overview of Deepfakes

- Portmanteau of “deep learning” (a ML subfield) and “fake”
- Neural network trained on photos/video/audio of real people
  - Neural network: computer learns to perform a task by analyzing a training set of examples
  - More input → higher verisimilitude of output
- Growing prevalence of Generative Adversarial Networks
- 2 parts of a GAN:
  - *Generator* creates samples of video/images/audio
  - *Discriminator* tries to tell generated samples from real ones
  - Discriminator & generator “teach” each other iteratively

Why Worry about Deepfakes?

- Cat-and-mouse game
- State of the art is improving rapidly
- Freely available, easy to use software
- Quality, usability, accessibility will keep going up

→ Telling real from fake will be ever harder for humans & AI

- Will detection be a lost cause or stalemate in the long run?
- Alternative: authentication, not detection
  - Tools to authenticate footage & reveal any manipulation of it

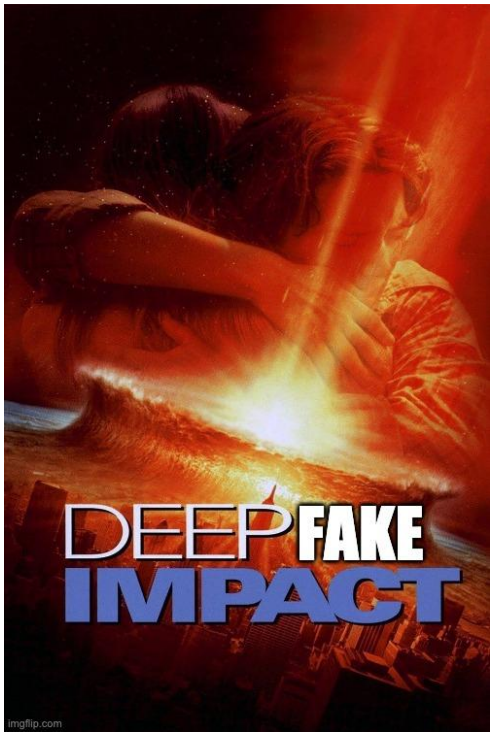

Image credit: Google Image Search

- Active role: as viewer of image
  - Can we tell what's real?
- Passive role: as subject of image
  - Depicted in malicious deepfake?
  - Photos used as training data?

Seeing Is Believing, but Deepfakes Aren't Our First Rodeo

Image credits: knowyourmeme.com (left); Wikipedia (right)

Deepfakes will touch every role in the courtroom...

- Counsel
- Judges
- Witnesses
- Juries

...and may come up in a number of ways:

- Purpose-made for the litigation
- Litigant doesn't realize video is fake
- Archival evidence (e.g. videos from a newsroom)

Authentication Prerequisite for Admissibility into Evidence

Federal Rule of Evidence 901(a):

> "To satisfy the requirement of authenticating or identifying an item of evidence, the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it is."

- State rules (e.g. CO) are similar/identical to federal
- This is a low bar!
- Admissibility → judge, weight → jury

Video Authentication

- Authenticate video evidence by showing it's an accurate representation of the scene depicted
- Witness can, but need not, have taken the video or seen the event being recorded
- Firsthand knowledge is preferable but not required
  - Witness who wasn't present can still authenticate if has prior knowledge of people & places depicted
- Low bar leaves room for mischief:
  - Witness unwittingly authenticates a deepfake
  - Custodian of records vouches for deepfake from archive

Challenging a Video's Authenticity

- Move to strike (civil) / exclude (criminal)
- Produce evidence to support contention of fakery
  - Challenge provenance, chain of custody; call expert & lay Ws
  - Cross-X proponent's W: who created it, when, with what tech
  - Lay witness: e.g. the person depicted in the fake video
    - Be mindful of W credibility issues & criminal Ds' 5th Am. rights
  - Expert witness: trained in digital video forensics
  - For now, many deepfakes are "shallowfakes"; easy to detect

Authentication Battles Will Increase Case Time & Costs

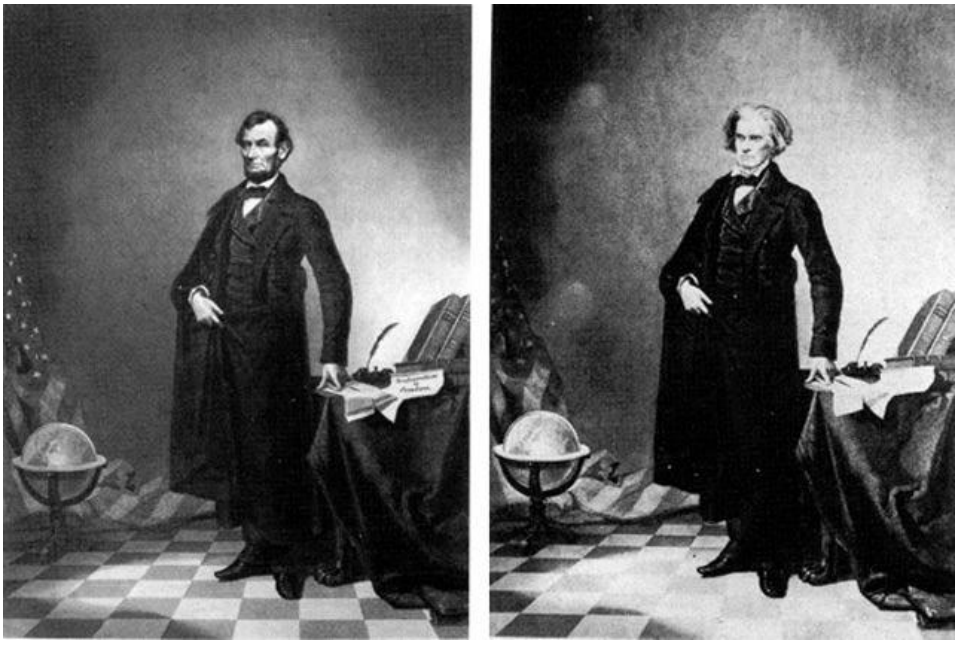

Image credit: iconicphotos.wordpress.com

Should We Raise the Bar for Authenticating Videos?

- *Gonzales v. People*, 2020 CO 71 (Sept. 2020)
  - "In the absence of evidence suggesting that a proffered [video] has been altered or fabricated, ... a proponent may authenticate a recording by presenting evidence sufficient to support a finding that it is what the proponent claims."
  - Appeals court: "the fact that the falsification of electronic recordings is always possible does not, in our view, justify restrictive rules of authentication that must be applied in every case when there is no colorable claim of alteration"

Deepfakes' Ramifications for Genuine Evidence

- "Verified at capture" technologies for photo/video/audio
- Reverse *CSI* effect
- Liar's Dividend
- Access-to-justice disparities
- Skepticism of low-tech videos replaces "seeing is believing"

Practice Pointers

- **Use existing strategies** for authenticating digital evidence against allegations of alteration or forgery.
- **Adjust your litigation budget** estimates to account for extended timelines and the cost of lay and expert witnesses.
- **Do your due diligence.** Learn the telltale signs of deepfakes, verify provenance before offering a video as evidence, and prepare how you'll respond if its authenticity is questioned. If a "smoking gun" video seems too good to be true, it probably is.
- **The courts' integrity depends on yours.** If a client pushes you to go forward with a suspected or known fake, call your state bar's ethics hotline or your firm's ethics counsel.

Thank You

Email: riana@stanford.edu
Twitter: @Riana_Crypto

Further Reading:

- Riana Pfefferkorn, *"Deepfakes" in the Courtroom*, 29 B.U. Pub. Int. L.J. 245 (2020) (available on TILT event webpage)
- Robert Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 Cal. L. Rev. 1753 (2019)
- Peter Manseau, *The Apparitionists: A Tale of Phantoms, Fraud, Photography, and the Man Who Captured Lincoln's Ghost* (Houghton Mifflin Harcourt 2017)

Postscript: Deepfakes and Professional Responsibility

- Proponent of video: offering a fake as evidence?
  - Attorney may not offer into evidence a video she knows or reasonably believes is a deepfake. MRPC 3.3(a)(3)
- Opponent of video: baselessly accuse video of being fake?
  - No making frivolous arguments. MRPC 3.1, FRCP 11(b)(2)
  - No baselessly denying factual contentions. FRCP 11(b)(4)
  - No motion practice just to harass, delay, or needlessly increase litigation costs. FRCP 11(b)(1)
- Zealous client rep vs. duty to "further the public's … confidence in the rule of law and the justice system." MRPC preamble ¶ 6