# Interactive Discussion:
# *AI for Real or Fake???*

ACEDS MIDWEST AI SERIES

**Artificial Intelligence in Legal Practice**

Tools, Workflows, and Real-World Results


## Series Schedule

| Date | Session | Format |
|---|---|---|
| March 25, 2026 | Foundations in AI for Legal Professionals: ...Buckle in | Educational Session |
| April 22, 2026 | AI in Action: Lessons from Legal Practitioners | Practitioner Panel |
| May 27, 2026 | Interactive Discussion: AI for Real or Fake??? | Expert Panel |
| June 24, 2026 | Vendor Demonstration: AI Tools Across the Legal Technology Landscape | Vendor Showcase |
| July 22, 2026 | Judicial Perspectives on Artificial Intelligence | Judicial Panel |
| August 19, 2026 | Final Interactive Workshop: Implementing AI in Legal Practice | Workshop |

https://ediscovery.aceds.org/aceds-midwest-ai-series

# Brett Burney

## Principal

## **Burney Consultants LLC**

Always anxious to encourage legal professionals to take advantage of technology, Brett primarily helps law firms and corporate legal departments successfully navigate their eDiscovery challenges.

Brett co-authored the “eDiscovery for the Rest of Us” book with Nextpoint and Tom O’Connor and regularly speaks to lawyers and legal groups on a wide variety of technology-related topics. Brett served as the Co-Chair of the 2026 ABA TECHSHOW Planning Board and serves as a Trustee for the EDRM 2.0 Project.

You can contact him at burney@burneyconsultants.com

Burney consultants

# Sarah Thompson

## Chief Product Officer

## BlueStar Case Solutions

- 20+ years of experience in legal technology, eDiscovery, and digital forensics
- Recognized expert and thought leader in the legal applications of artificial intelligence and legal technology
- Creator of BlueStar's Next-Gen AI Solutions to streamline review, analysis and of complex ESI
- Developed Siemly, the first platform to detect and investigate litigation events in Microsoft 365 without data collection
- Frequent CLE instructor on AI, legal tech strategy, and defensible workflows
- Trusted advisor to law firms and corporate legal departments on AI readiness, risk mitigation, and workflow automation

BLUESTAR

# Robert B. Fried

## EVP of Forensics & Chief Investigative Officer

Page One, Inc

- 24 years performing digital forensic collections & investigations
- Vice President, Society of Professional Investigators
- Adjunct Assistant Professor, Hofstra University
- Author, Forensic Data Collections 2.0 book series
- Content Creator, www.forensicsbyfried.com
- Developer, https://forensics.expert

PAGEONE

## What will we learn today?

- What is it?
- How is it identified?
- Who do you call?
- How is it used?
- Why should you care?
- Do's and Don't's / Considerations

# What are deepfakes?

- A multimedia based (video, photo, or audio recording) that seems real but has been manipulated with AI.
- The underlying technology can replace faces, manipulate facial expressions, synthesize faces, and synthesize speech.
- Deepfakes can depict someone appearing to say or do something that they in fact never said or did.

# Ways AI-Generated evidence might come to court

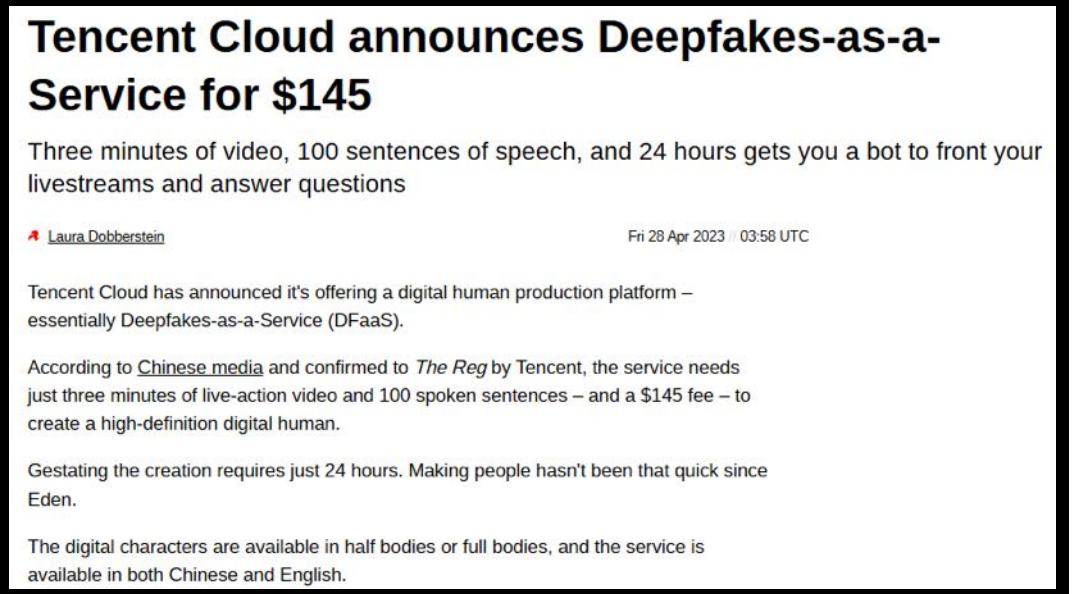
- ***ACKNOWLEDGED*** **AI-GENERATED EVIDENCE**
  - The parties agree the evidence is the product of an AI system; the dispute concerns the **validity, reliability, or bias of the AI system**
  - A party's experts reveals they used an AI system to assess the evidence; the dispute concerns the **validity, reliability, or bias of the expert's use of the AI system**

# Ways AI-Generated evidence might come to court (cont'd)

- ***UNACKNOWLEDGED*** **AI-GENERATED EVIDENCE**
  - One party claims the evidence is genuine and the other claims it has been manipulated or is a deepfake; the dispute concerns the **authenticity of the evidence**

# Deepfake-as-a-Service (DFaaS)

- On-demand, realistic synthetic media
- Anyone can commission deepfakes — no expertise required
- Tools are used to fabricate alibis, fake evidence, or impersonate officials
- Undermines trust in digital evidence used in investigations and courtrooms
- Raises new challenges in authentication, attribution, and victim protection



## Tencent Cloud announces Deepfakes-as-a-Service for $145

Three minutes of video, 100 sentences of speech, and 24 hours gets you a bot to front your livestreams and answer questions

Laura Dobberstein — Fri 28 Apr 2023 // 03:58 UTC

Tencent Cloud has announced it's offering a digital human production platform – essentially Deepfakes-as-a-Service (DFaaS).

According to Chinese media and confirmed to *The Reg* by Tencent, the service needs just three minutes of live-action video and 100 spoken sentences – and a $145 fee – to create a high-definition digital human.

Gestating the creation requires just 24 hours. Making people hasn't been that quick since Eden.

The digital characters are available in half bodies or full bodies, and the service is available in both Chinese and English.

# DEEPFAKE TECHNOLOGY

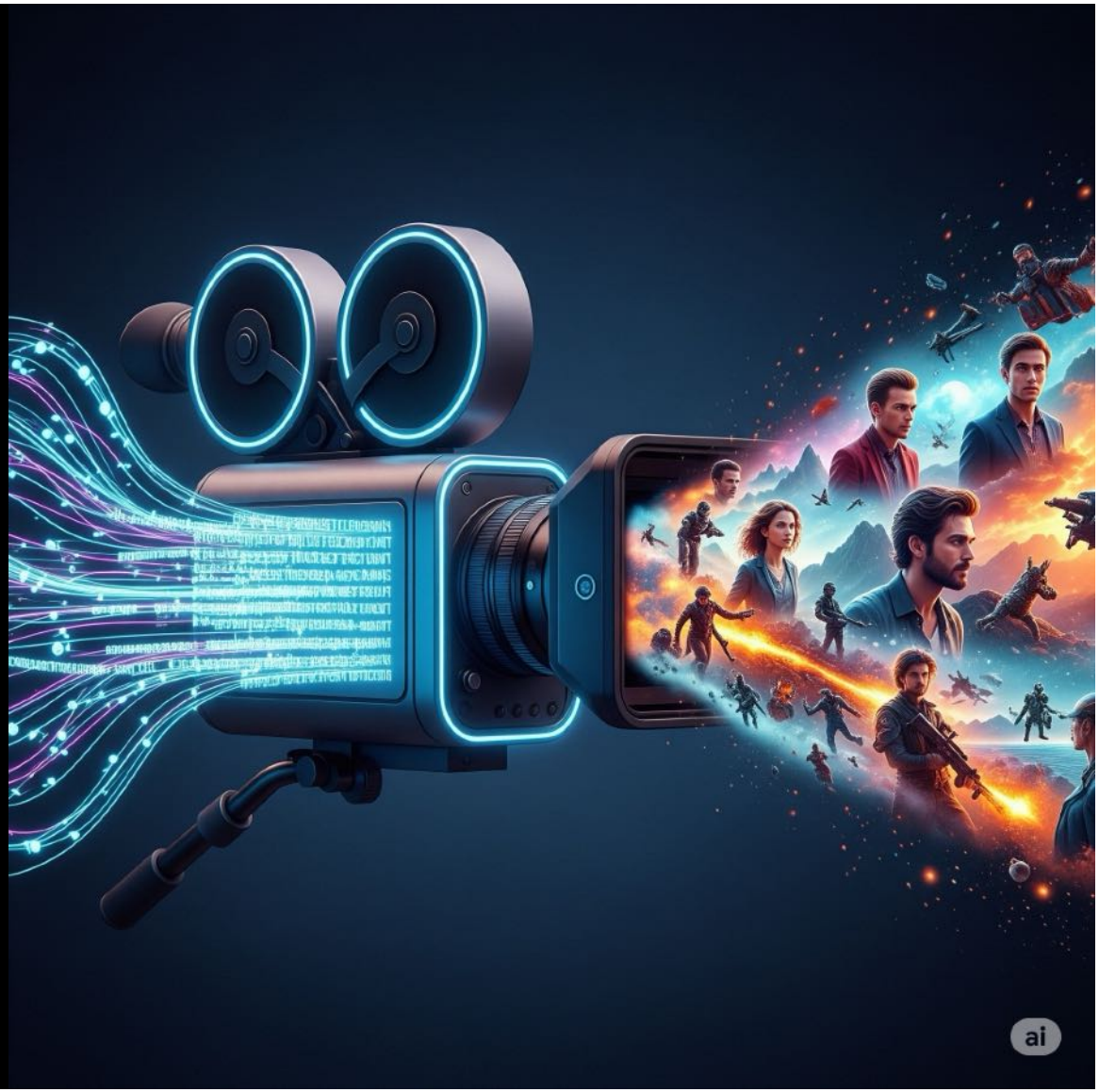
WIDELY AVAILABLE

EASY TO USE

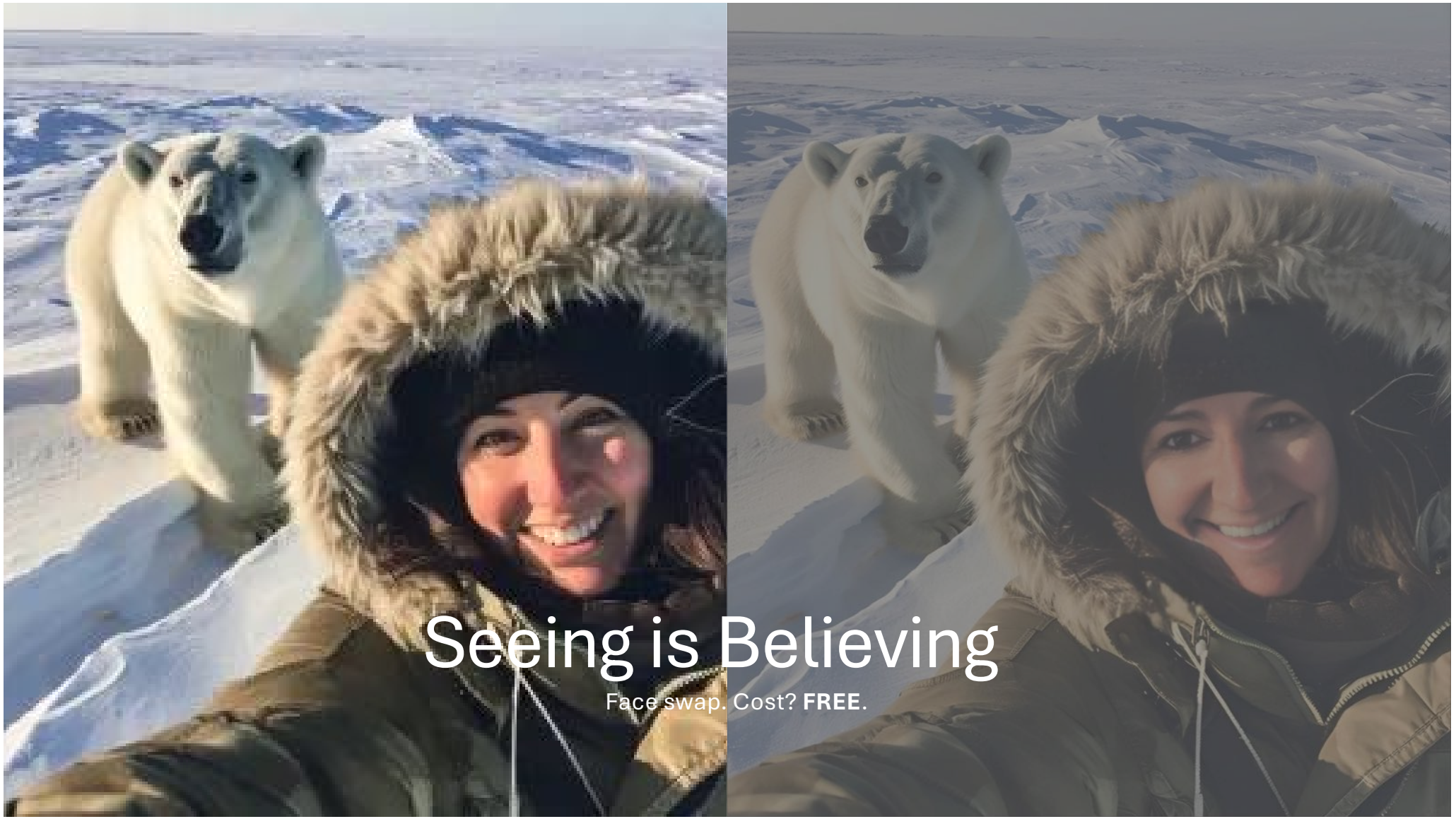
UNREGULATED

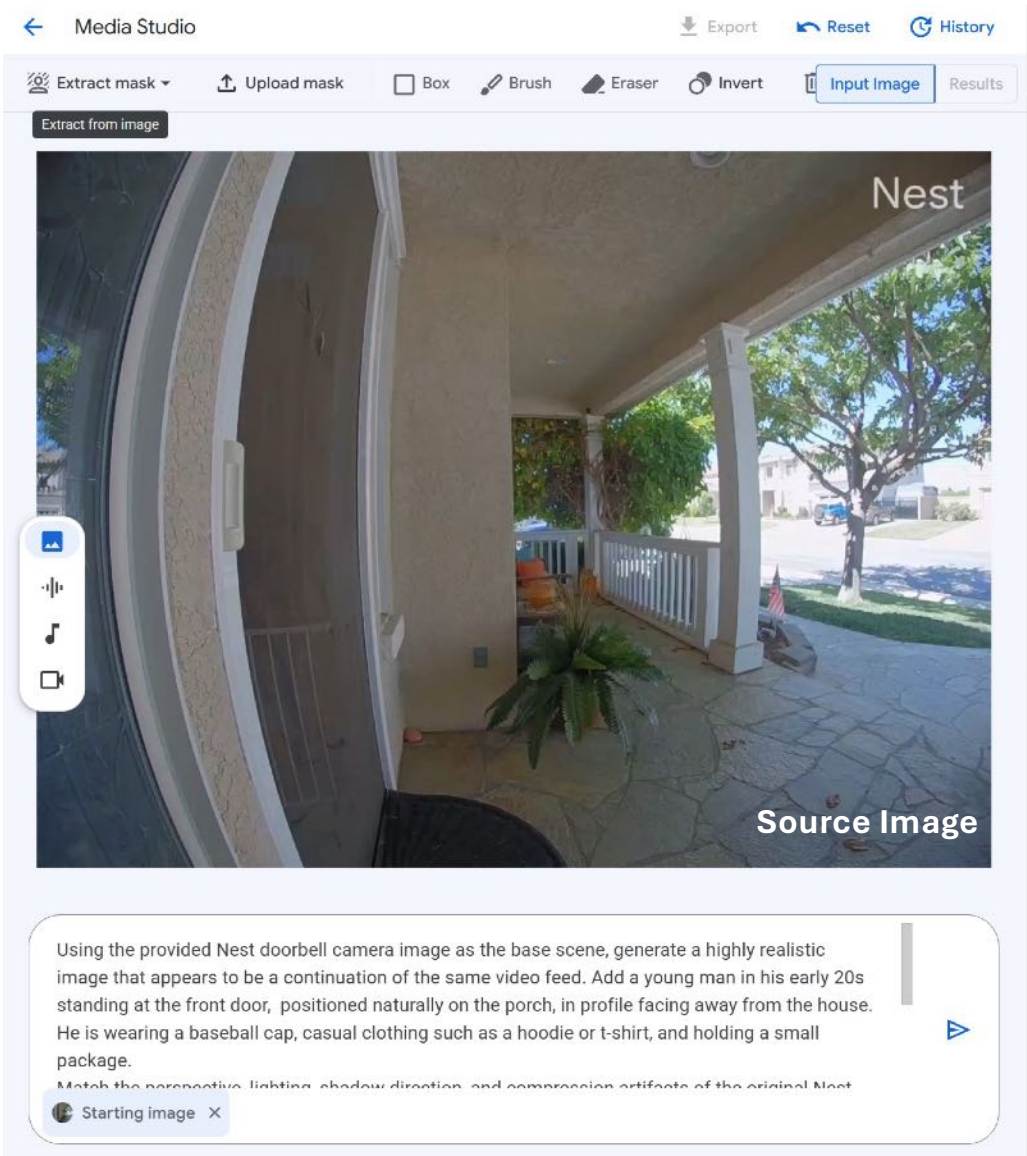
CHEAP OR FREE

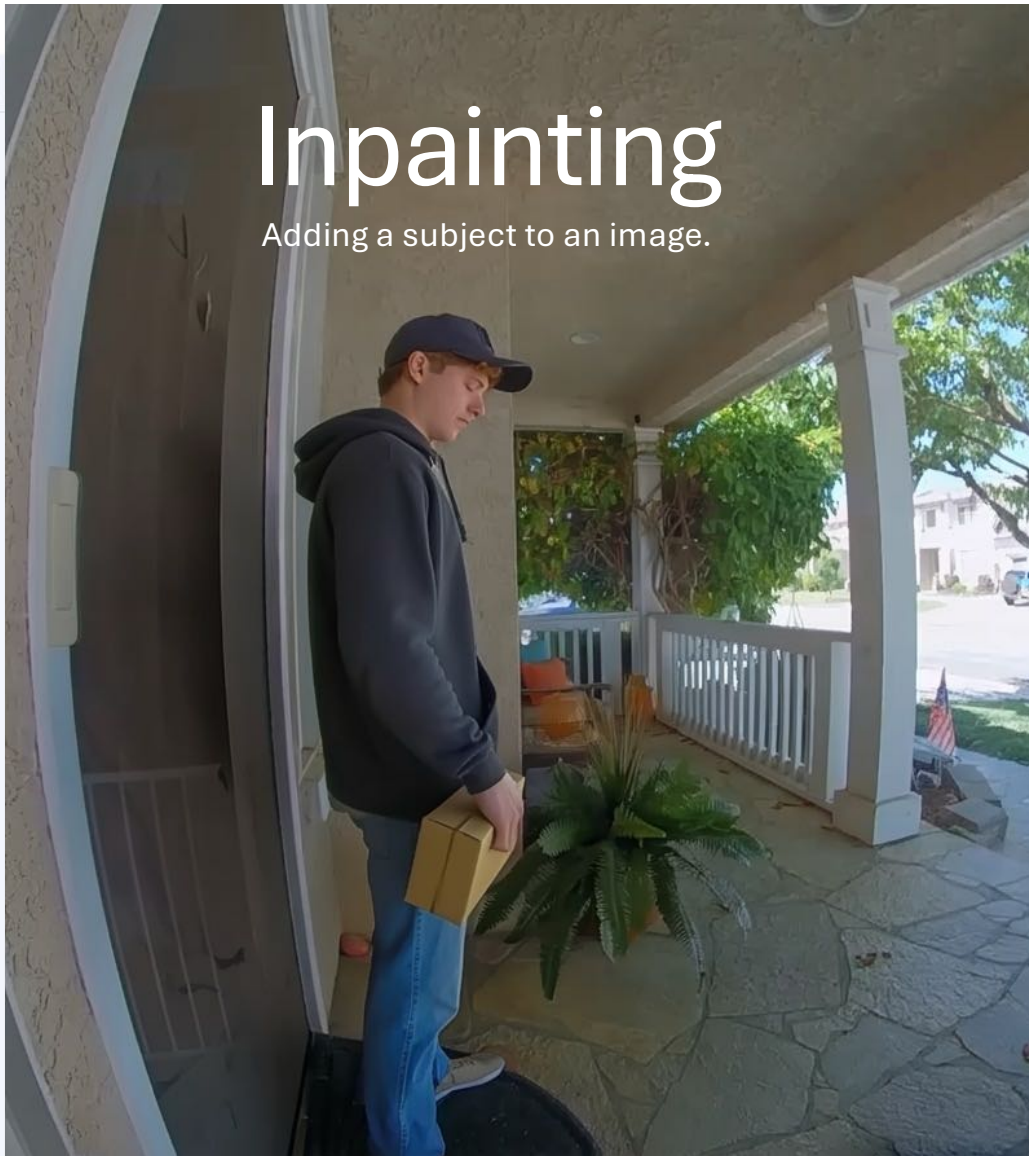
<u>AWESOME.</u>




Deepfake Examples


Seeing is Believing
Face swap. Cost? **FREE**.


Media Studio
Export
Reset
History
Extract mask
Upload mask
Box
Brush
Eraser
Invert
Input Image
Results
Extract from image
Nest
Source Image
Using the provided Nest doorbell camera image as the base scene, generate a highly realistic image that appears to be a continuation of the same video feed. Add a young man in his early 20s standing at the front door, positioned naturally on the porch, in profile facing away from the house. He is wearing a baseball cap, casual clothing such as a hoodie or t-shirt, and holding a small package.
Starting image


Inpainting
Adding a subject to an image.

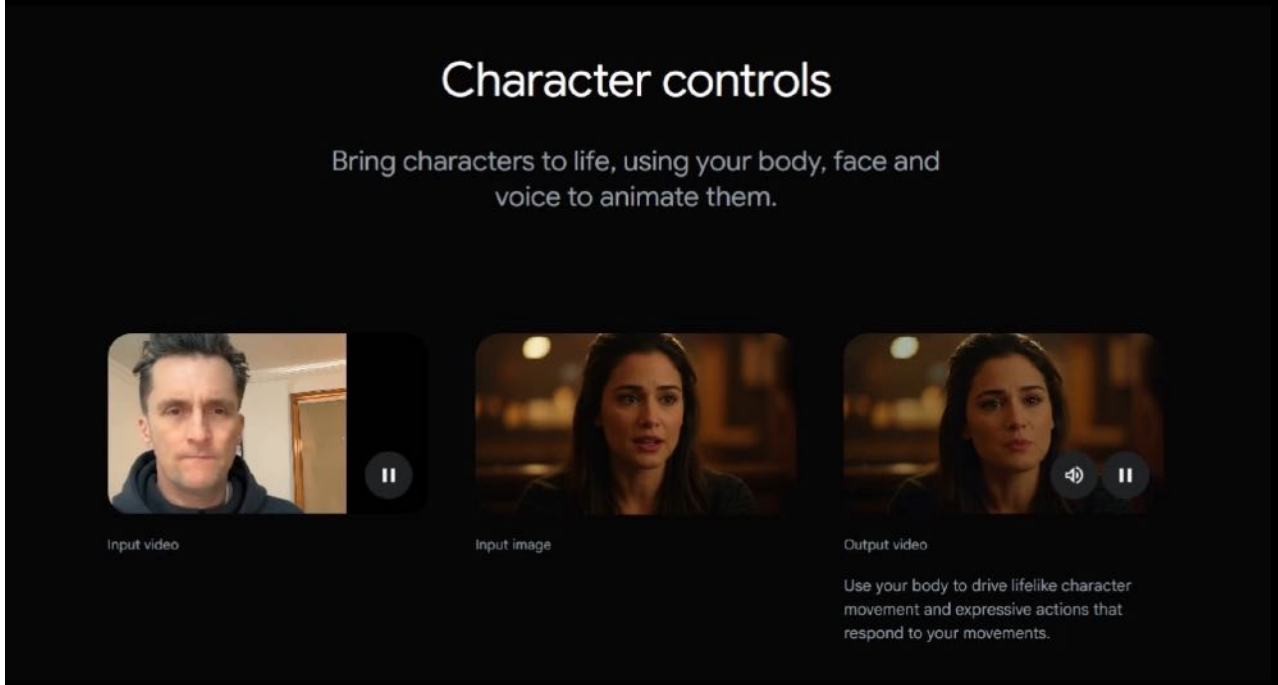

Text to Video
Hyper-realistic
Veo

# Puppet Mastering (Avatar Animation)

A recording or real-time video of an AI-generated avatar of a person, using scripted prompts.

Looks and sounds like the real person— used for impersonation and reenactment.

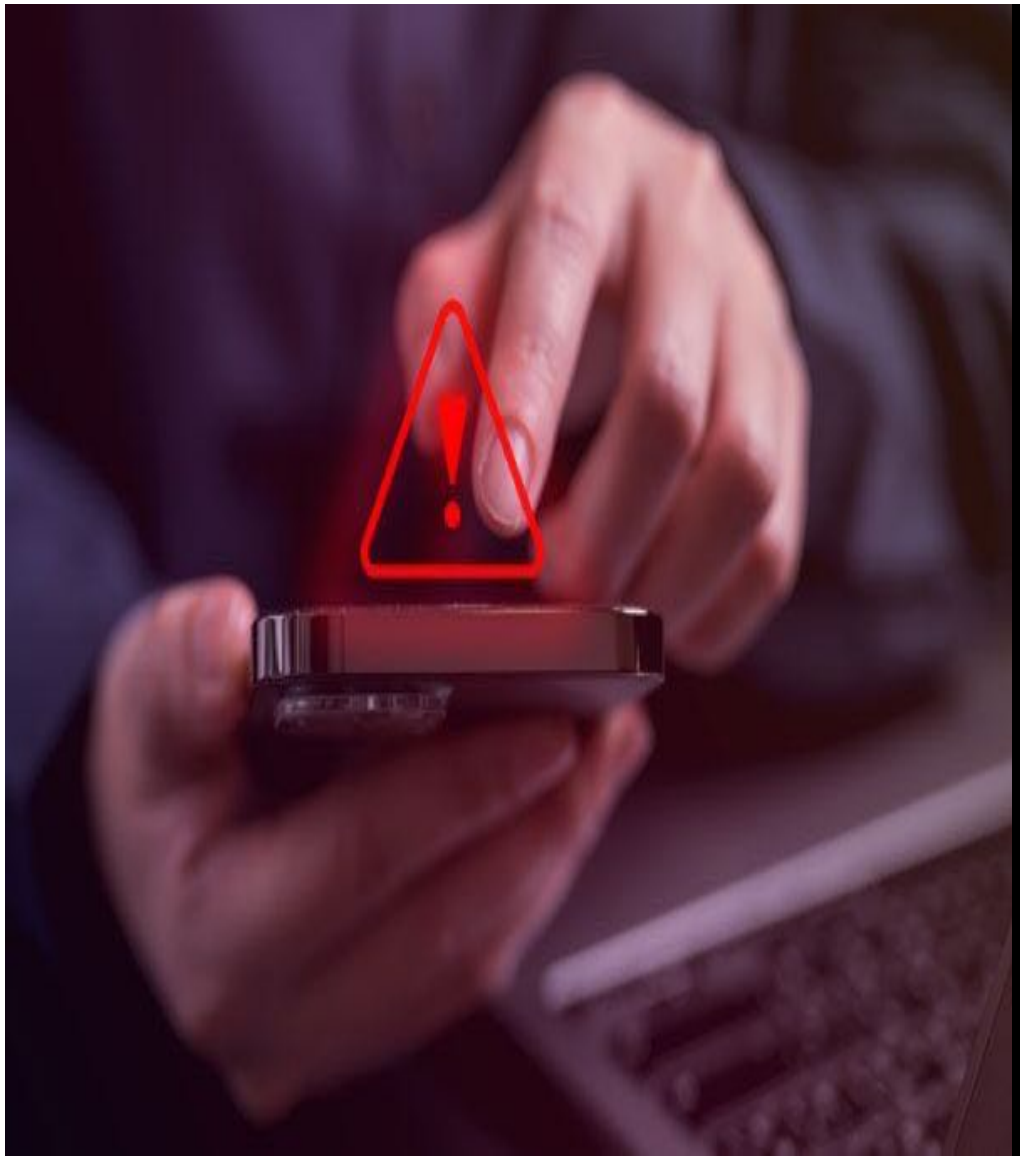
**Can be created from a single photo and 60 seconds of voice recording.**




https://deepmind.google/models/veo/

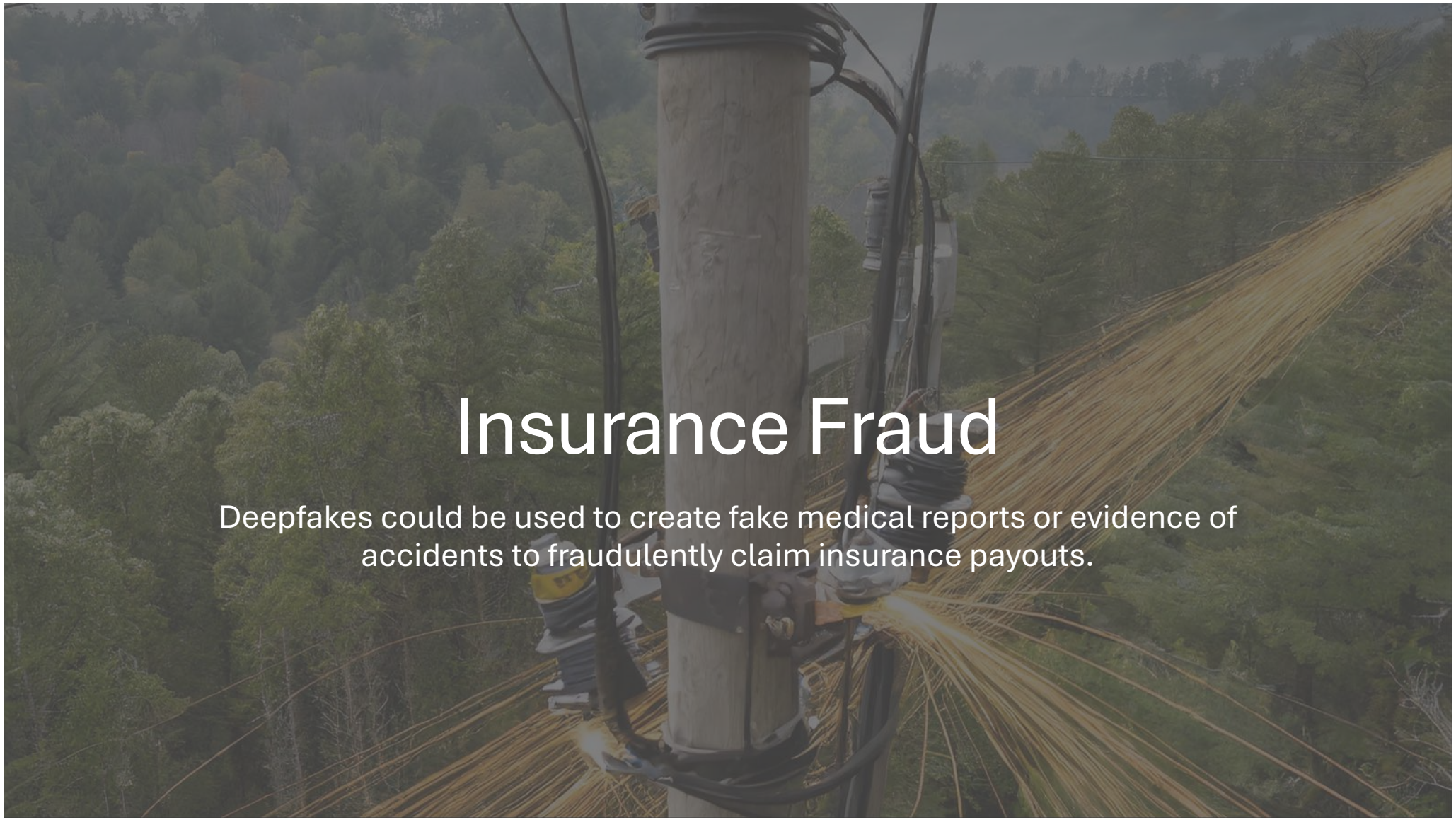
# Real-time Deepfakes

A person appears live on camera as someone else using real-time facial and voice masking — enabling fraud, misrepresentation, and real-time manipulation.



# Grandparent Scams

A call from someone posing as a loved one asking for immediate help. With deepfake technology, scammers can imitate voices to convince relatives to send money. Becoming prominent in 2020, this scam has only grown, causing up to 1.65 billion dollars in damage just last year.



# Insurance Fraud

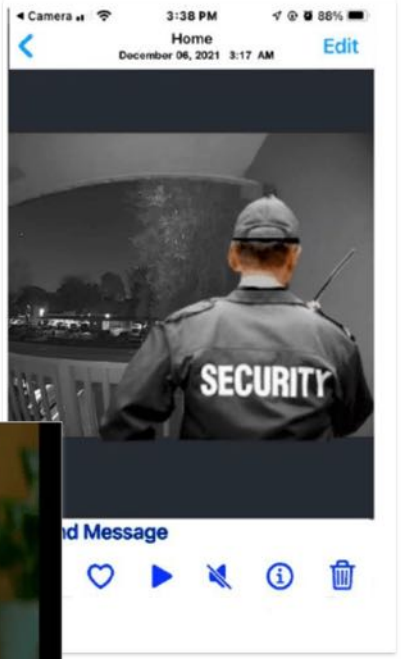
Deepfakes could be used to create fake medical reports or evidence of accidents to fraudulently claim insurance payouts.

SEPTEMBER 25, 2025 | DOUG AUSTIN

# Deepfake Videos and Images Lead to Terminating Sanctions: eDiscovery Case Law

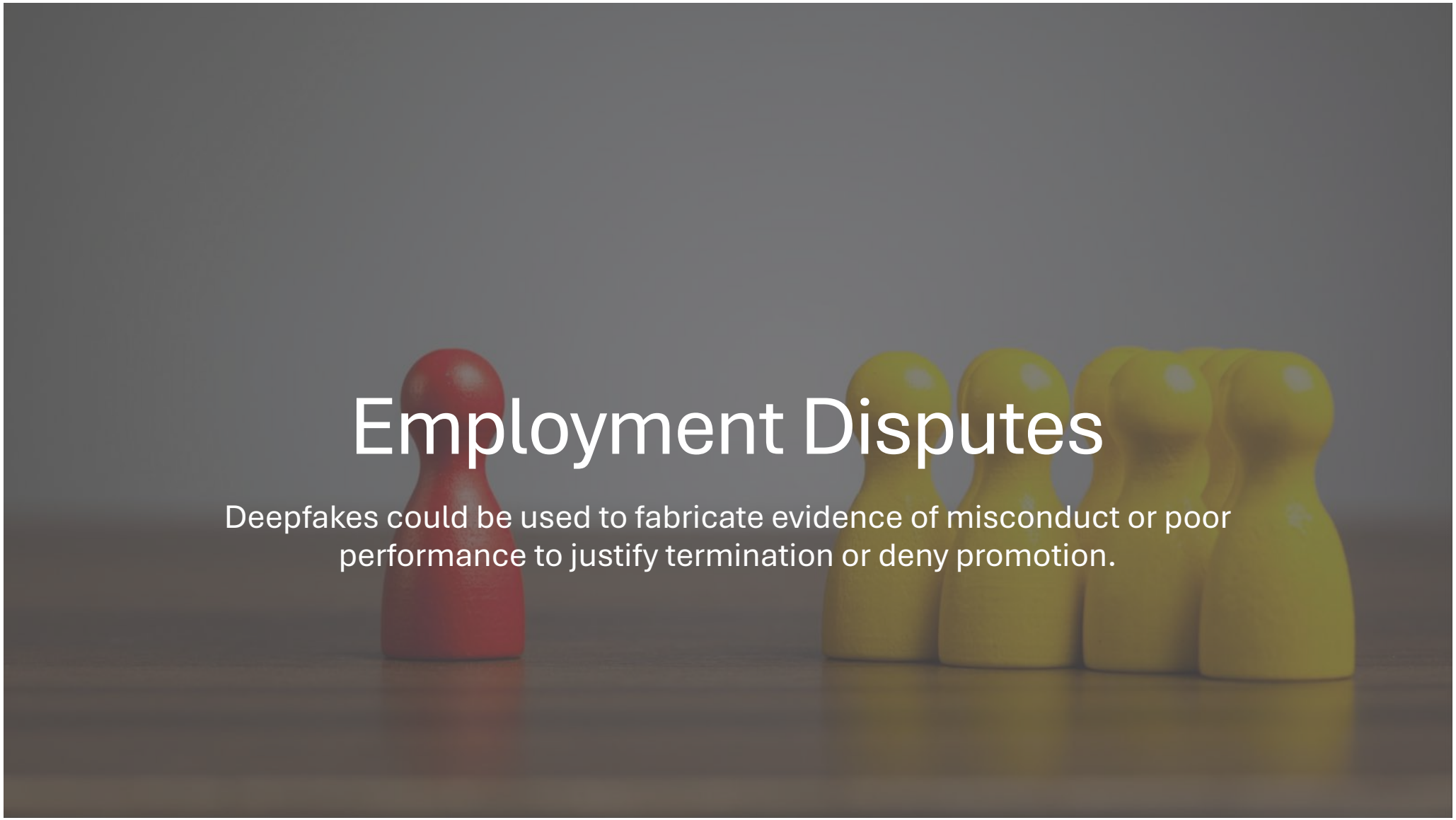
In *Mendones v. Cushman & Wakefield, Inc., No.: 23CV028772 (Cal. Super. Sept. 9, 2025)*, California State Superior Court Judge Victoria Kolakowski found "that a terminating sanction is appropriate" after the Court determined that several exhibits provided by Plaintiffs were deepfake videos and images.

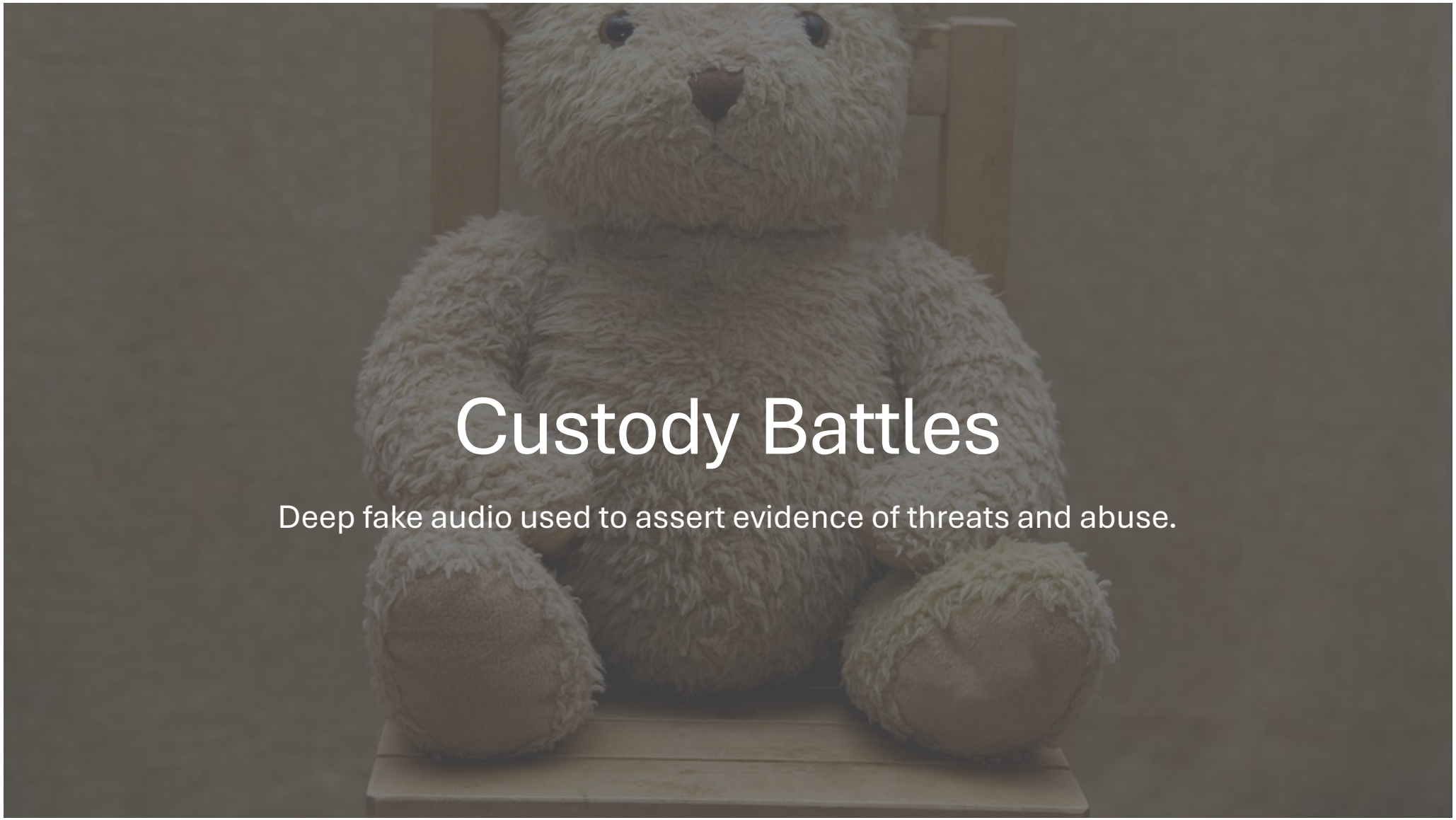
eDiscoveryToday

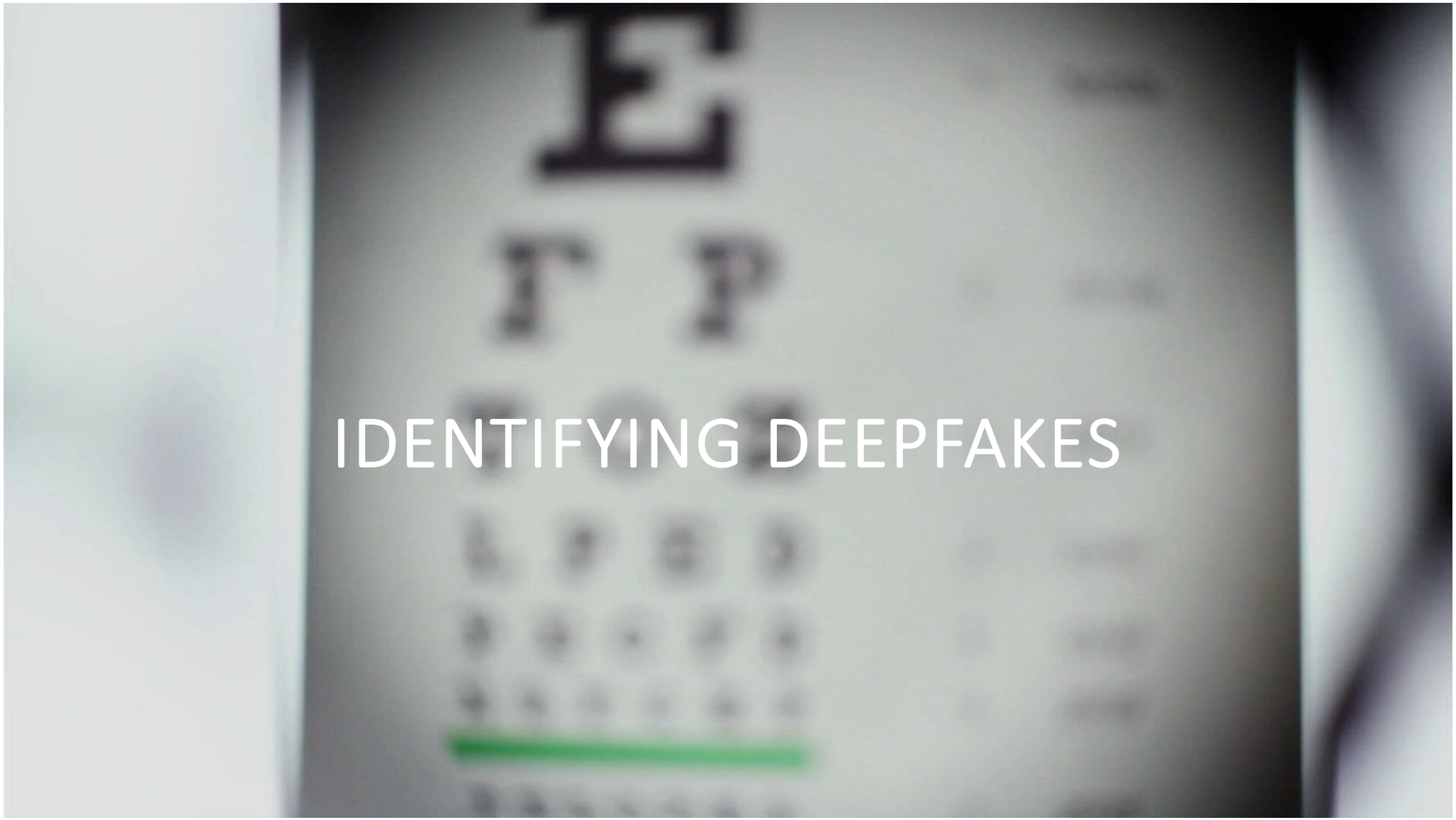


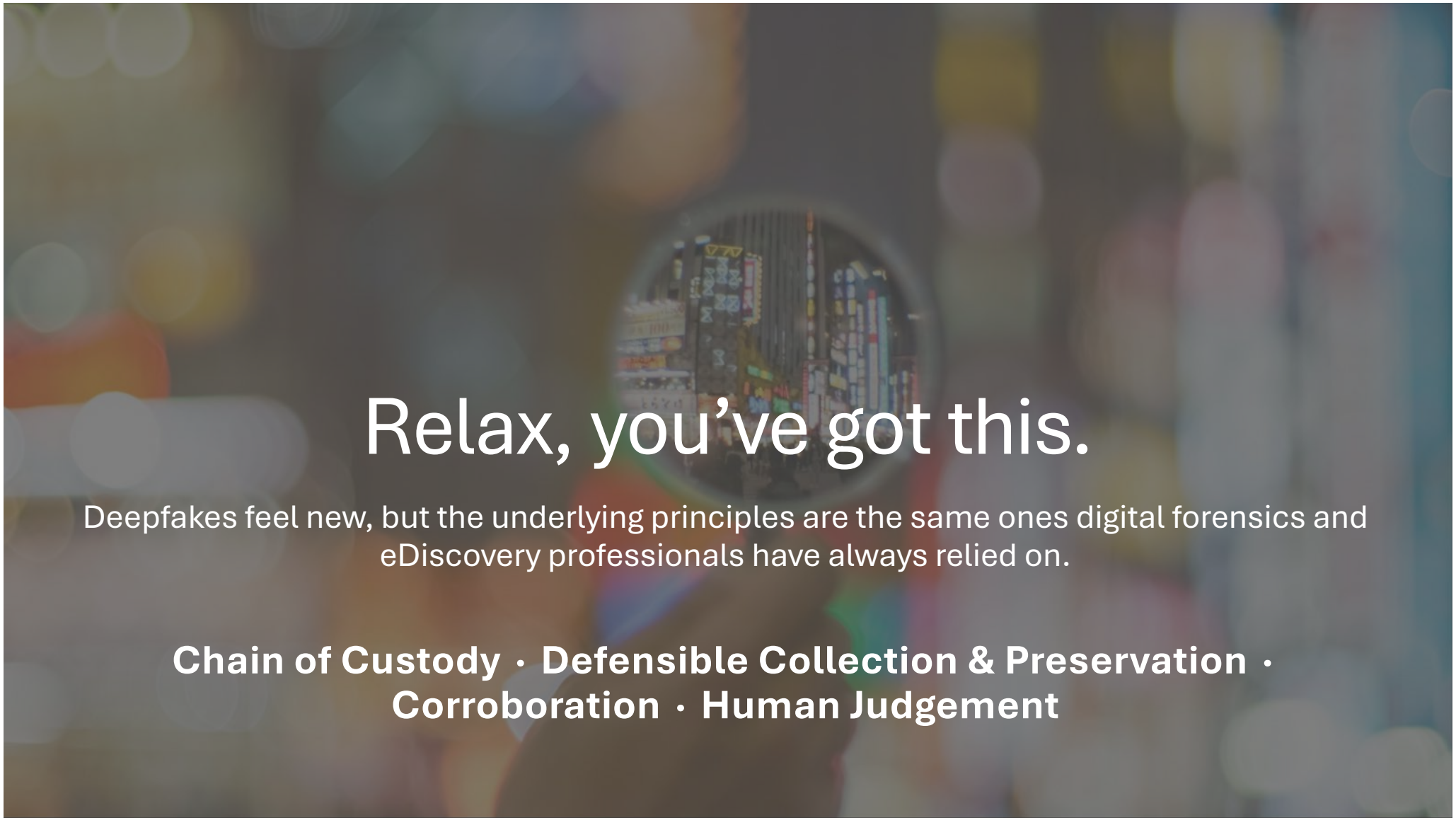


*"Deepfake Videos and Images Lead to Terminating Sanctions: eDiscovery Case Law"*


Employment Disputes
Deepfakes could be used to fabricate evidence of misconduct or poor performance to justify termination or deny promotion.


Custody Battles
Deep fake audio used to assert evidence of threats and abuse.


IDENTIFYING DEEPFAKES

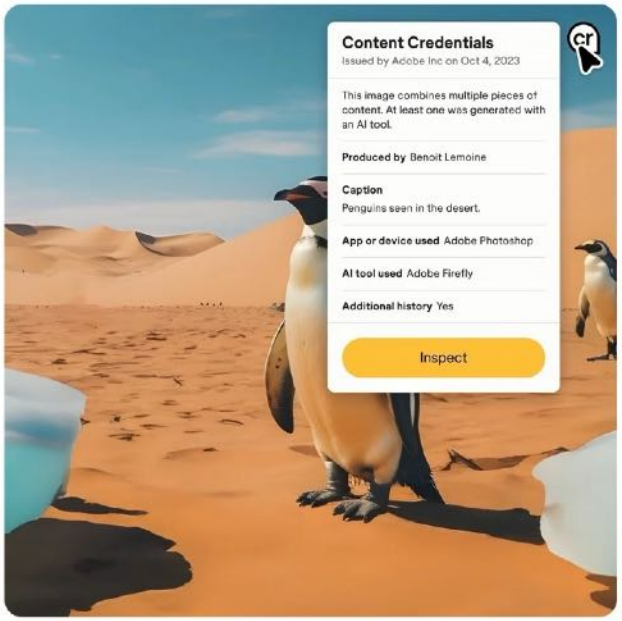

Relax, you've got this.
Deepfakes feel new, but the underlying principles are the same ones digital forensics and eDiscovery professionals have always relied on.
Chain of Custody · Defensible Collection & Preservation · Corroboration · Human Judgement

content credentials

Inspect

# Wait, where did this image come from?

Deepfakes. Voice cloning. Synthetic media. It's hard to tell what's accurate and authentic these days.

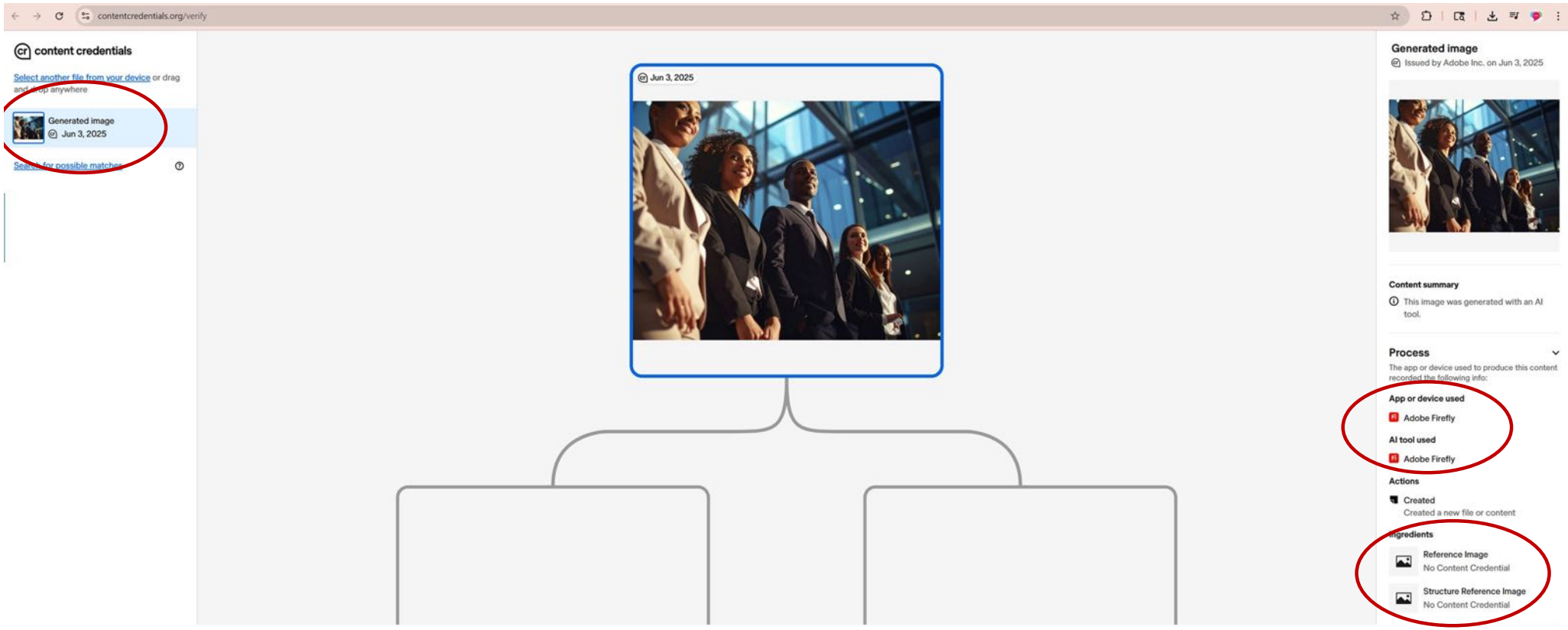



https://contentcredentials.org/verify/

# Provenance

Data that is embedded or included in the digital content's metadata, for the purpose of verifying the media's authenticity, origin, or history of modification.

**Tracking provenance is crucial in verifying the authenticity of media.**

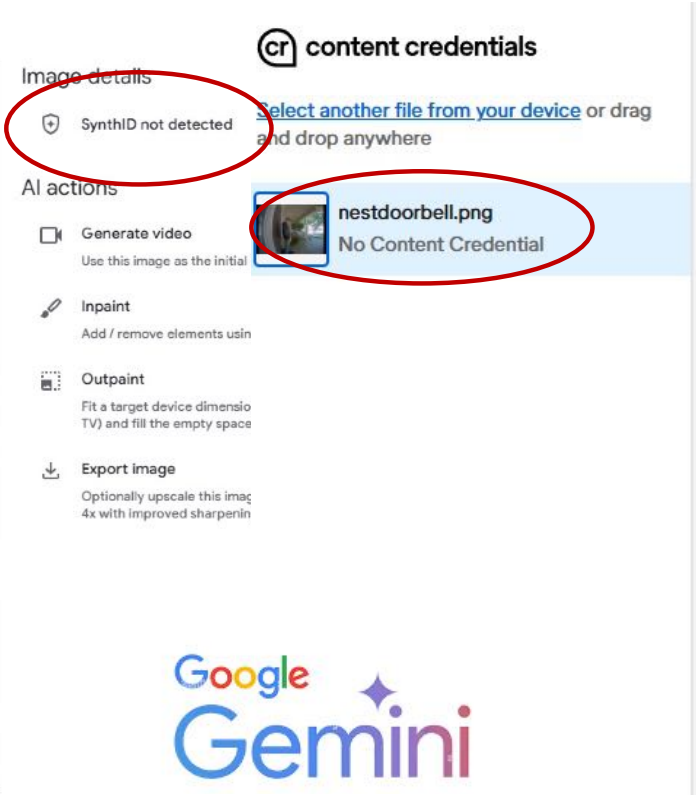



https://contentcredentials.org/verify/

# C2PA Verify

Content Credentials can capture a detailed history of changes over time. The Verify feature allows you to explore this information in depth and upload any content to see if it has Content Credentials.


Watermarks
Added by the AI platform when they generate media to identify the media as AI generated.
Invisible people, but a mathematical key easily reveals it.
Non-watermarked
Watermarked


# Provenance & Watermarks are Optional

Even the founding members of C2PA do not **require** watermarks and provenance data


# Metadata Anomalies

**Lack of metadata tells you that the media has been manipulated in some way.**


| Metadata of nestdoorbell | |
|---|---|
| Checksum | 3174a6908e353374634a05e3d9b79e26 |
| Filename | nestdoorbell.png |
| Filesize | 1316 kB |
| Filetype | PNG |
| Filetypeextension | png |
| Mimetype | image/png |
| Imagewidth | 1280 |
| Imageheight | 896 |
| Bitdepth | 8 |
| Colortype | RGB |
| Compression | Deflate/Inflate |
| Filter | Adaptive |
| Interlace | Noninterlaced |
| Iptcdigest | c8b38965b44537a918a7e75998c8e3f6 |
| Xmptoolkit | XMP Core 5.5.0 |
| Digitalsourcefiletype | http://cv.iptc.org/newscodes/digitalsourcetype/trainedAlgorithmicMedia |
| Digitalsourcetype | http://cv.iptc.org/newscodes/digitalsourcetype/trainedAlgorithmicMedia |
| Credit | Made with Google AI |
| Imagesize | 1280x896 |
| Megapixels | 1.1 |

# Metadata

**Scrutinize metadata to verify provenance.**

## Think Critically

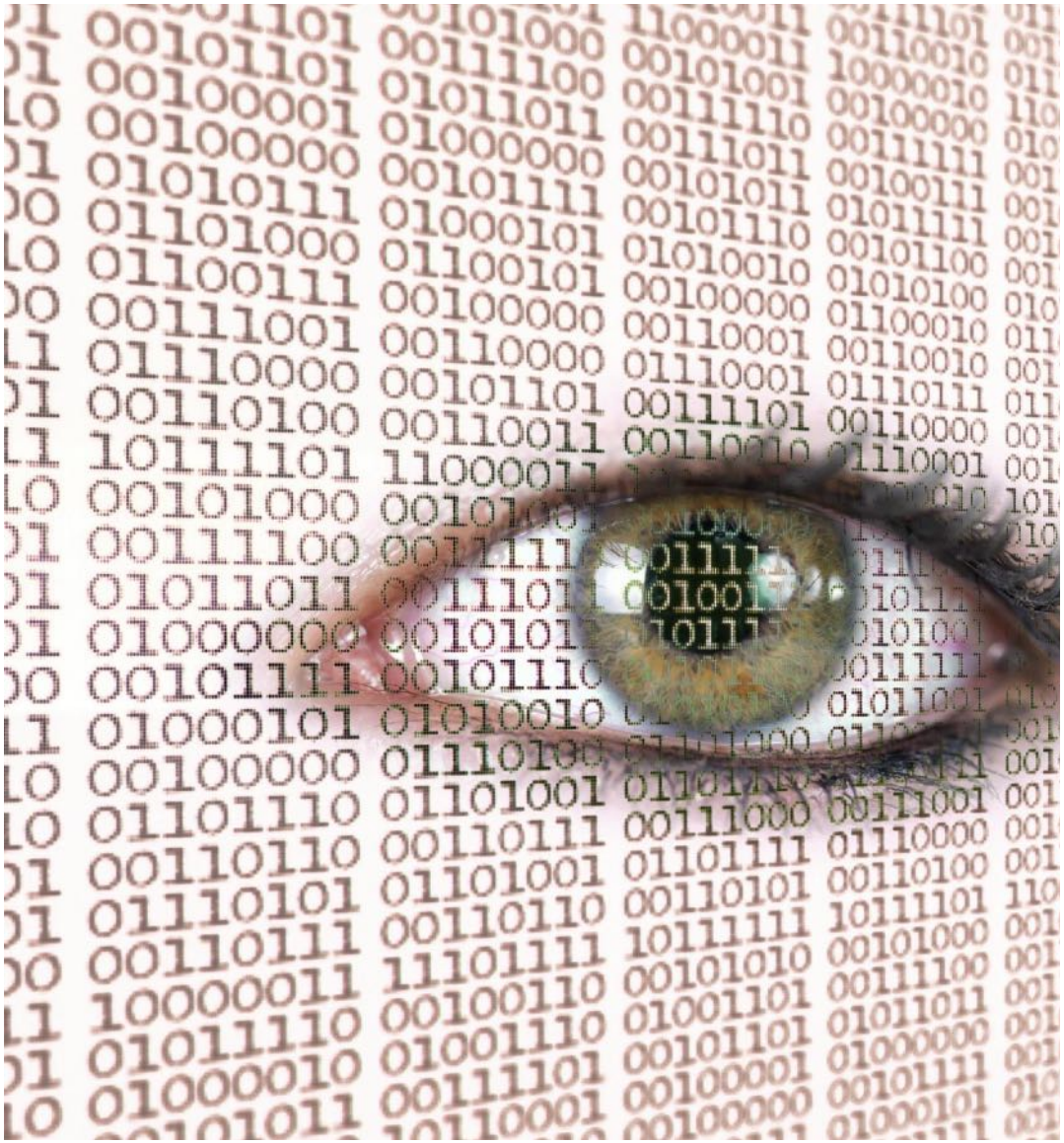
- Does it pass the “smell test”?
- Is the content of the media out of character for the source?
- Is there any external corroboration?
- Are there physical or environmental inconsistencies?

# Scrutinize the Media for Audio/Visual Anomalies

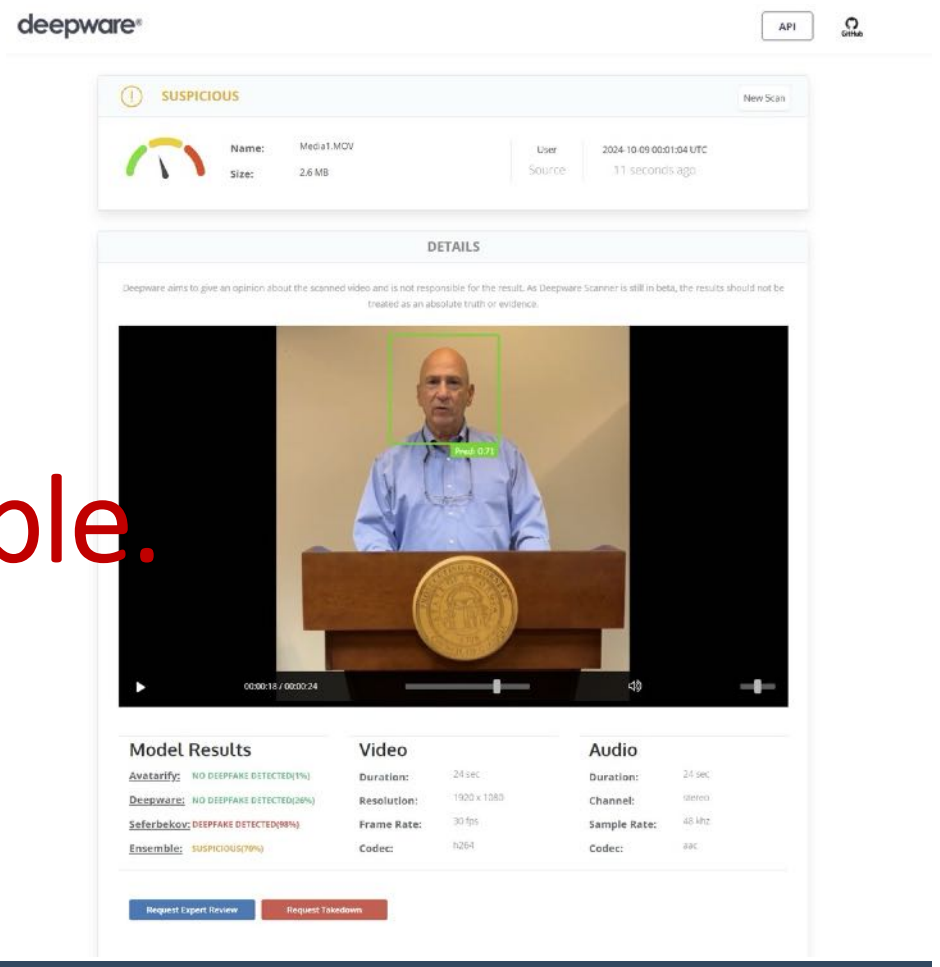
- Unnatural facial movements or expressions.
- Inconsistent lighting and shadows.
- Lip sync issues
- Background noise anomalies.
- Unnatural speech patterns.
- Inconsistencies in tone and emotion.

## Analyze the file

- Can you verify provenance?
- Is the source of the media unusual?
- Are there metadata gaps?
- Are there watermarks?
- Is the filetype or size unusual?




Unreliable.

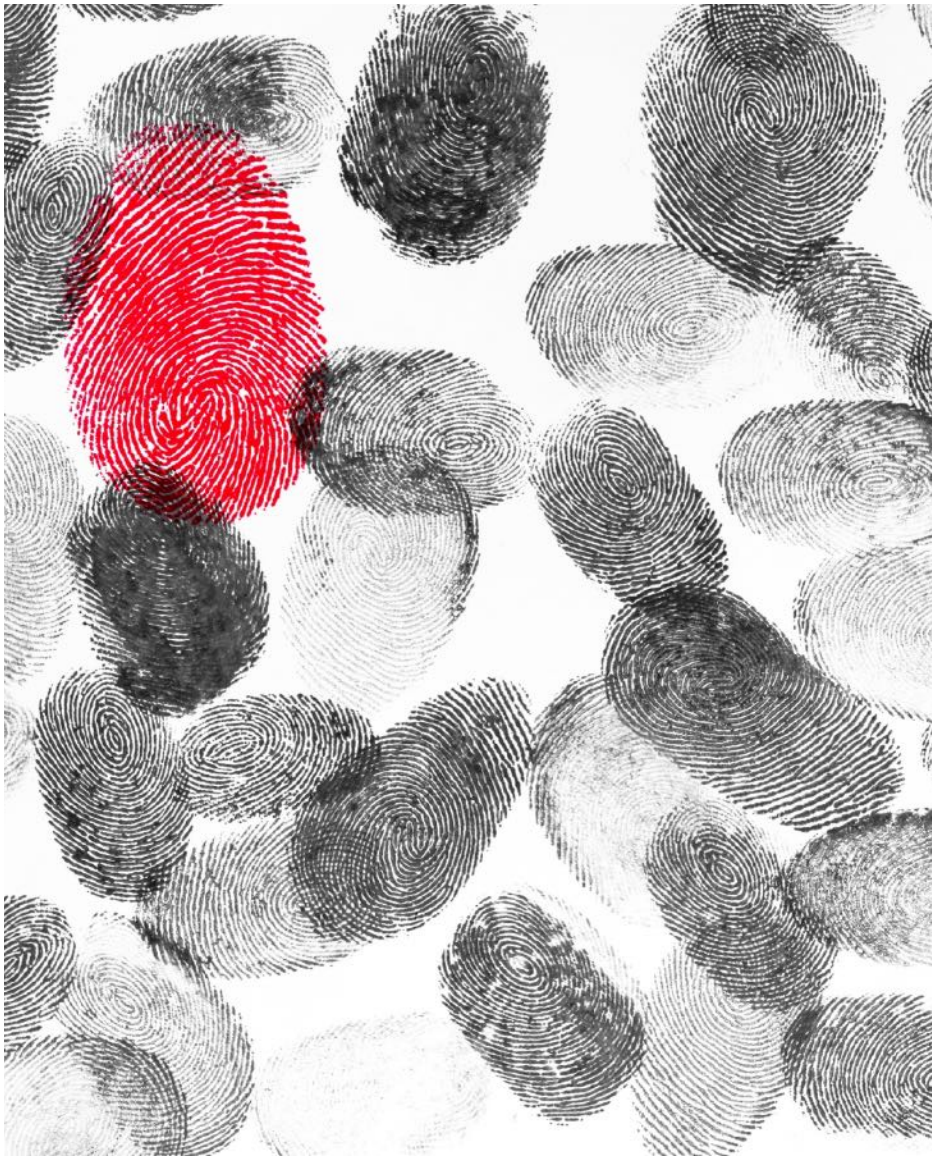


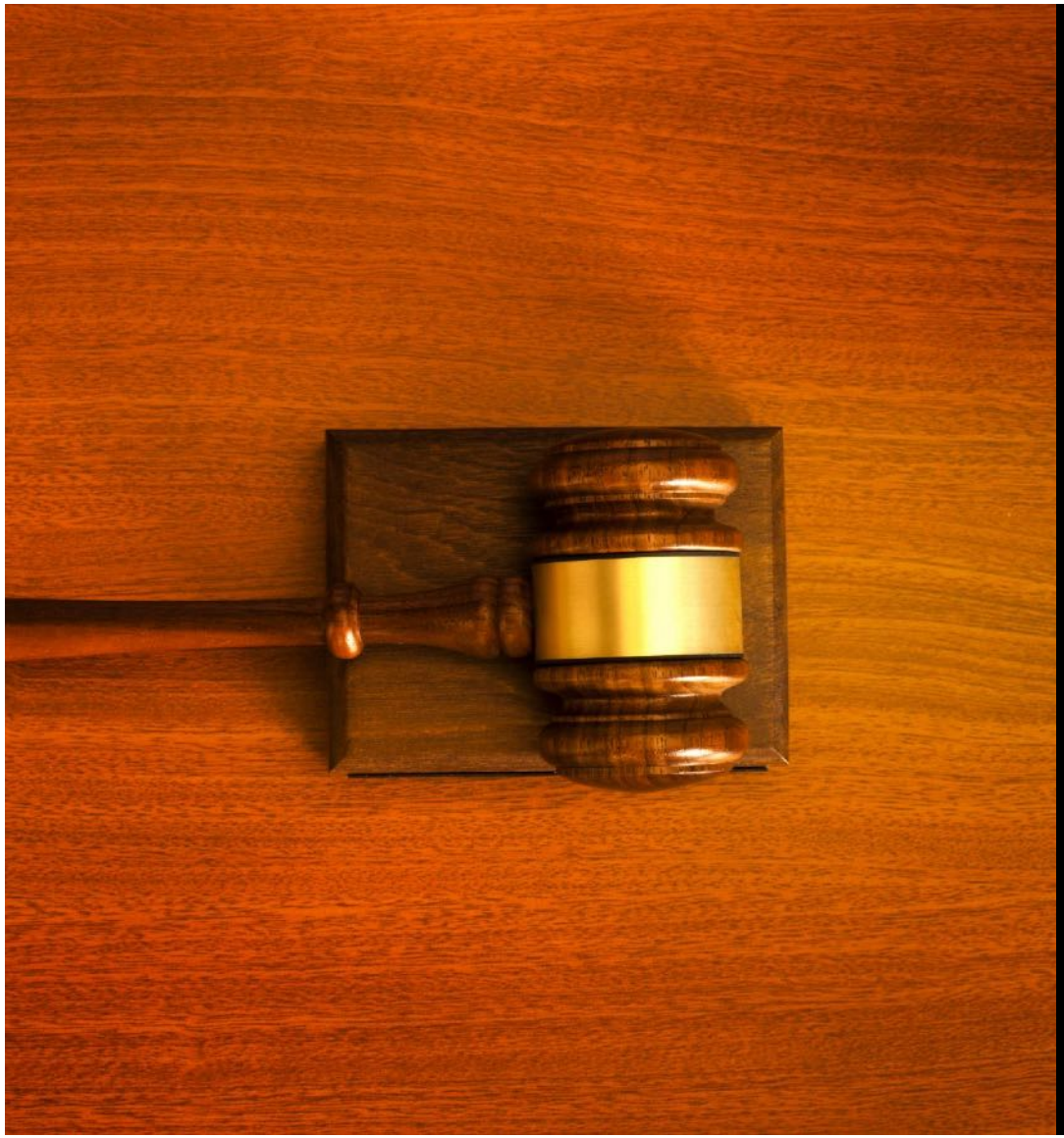

# Deep fake detection tools

# Forensic Software

**Provides legitimacy and clarity to media authentication**

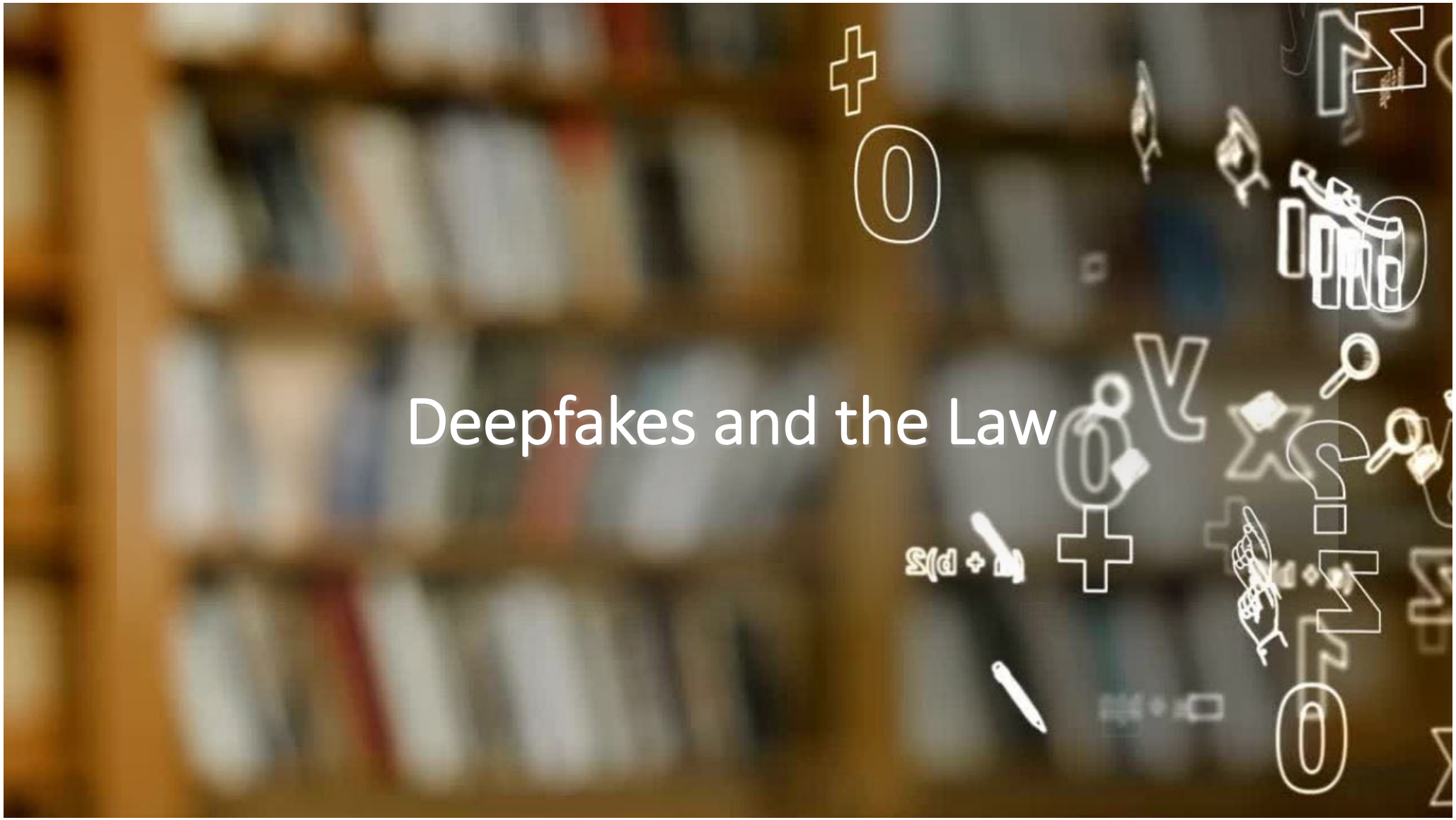
- Advanced Media authentication
  - Detects alterations, including edits and tampering.
  - Verifies file authenticity and identifies editing software.
  - Tracks the generational history of the media through messaging apps and email
  - Uses sensor noise patterns to correlate images or frames to a known device fingerprint
- Easy to understand reports for courtroom presentation
- Accelerates early detection of suspect files
- Supports findings with measurable, reproducible results





## Practical Strategies

- Use a “layered” authentication strategy
- “Front-load” Authentication
- Educate


Deepfakes and the Law

# Professional Responsibilities

## Maintain Technological Competence

Stay informed about AI-generated media and emerging forensic methods to recognize manipulated evidence.
*(ABA Model Rule 1.1 – Competence; Comment 8)*

## Verify Digital Evidence Authenticity

Attorneys must take reasonable steps to ensure digital evidence is authentic and not misleading before presenting it in litigation.
*(Rules 3.3 & 3.4; FRE 901)*

## Correct False Evidence

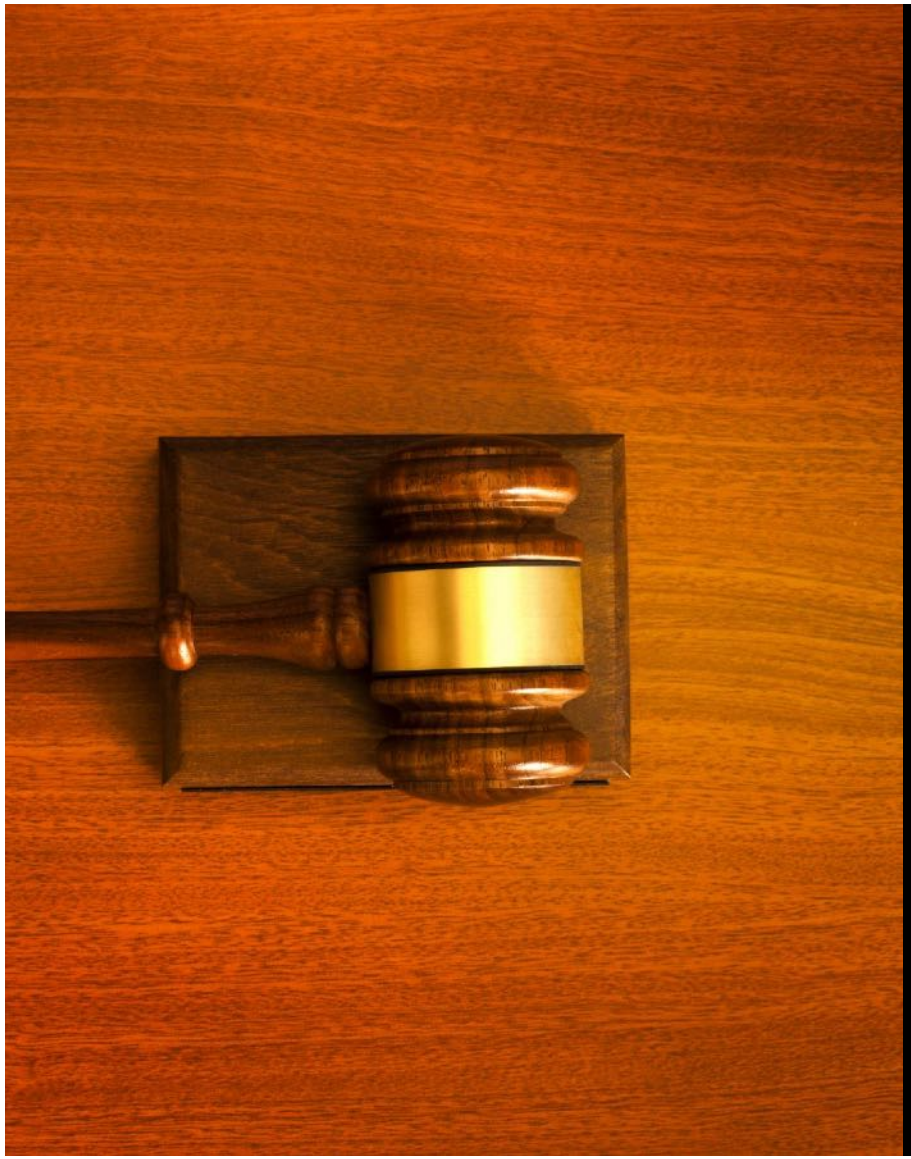
If an attorney later learns that a video, image, or recording presented to the court is manipulated or false, they must take remedial measures. *(Rule 3.3 – Candor Toward the Tribunal)*

## Use Reliable Experts and Methods

Deepfake detection testimony must rely on qualified experts and reliable methodologies. *(Rule 1.1; FRE 702)*

## Be Truthful About Digital Media

Lawyers may not make false statements or misrepresent the authenticity of digital evidence to the court or opposing parties.
*(Rules 3.3 and 4.1)*



# Key Evidence Rules

**FRE 901 — Authentication**

The proponent must present sufficient evidence showing the media **is what it claims to be.**

- Common authentication methods in deepfake disputes include:
- **FRE 901(b)(1):** Witness testimony verifying the recording or image
- **FRE 901(b)(4):** Distinctive characteristics, metadata, or contextual evidence
- **FRE 901(b)(9):** Evidence describing the process or system that produced the media

**FRE 902 — Self-Authenticating Evidence**

- Certain electronic records may be admitted without live testimony when accompanied by proper certification.

**FRE 702 — Expert Testimony**

Digital forensic experts may be required to analyze:

- metadata and file history
- editing artifacts
- device or sensor fingerprints
- indicators of AI generation or manipulation

**FRE 403 — Risk of Misleading the Jury**

- Courts may exclude evidence if its probative value is substantially outweighed by the risk of unfair prejudice or misleading the jury. Highly realistic deepfakes may raise Rule 403 concerns because synthetic media can be persuasive even if unreliable.

The Judges' Journal Technology Part 2

# The "Deepfake Defense": An Evidentiary Conundrum

Herbert B Dixon Jr

Jun 11, 2024 8 min read

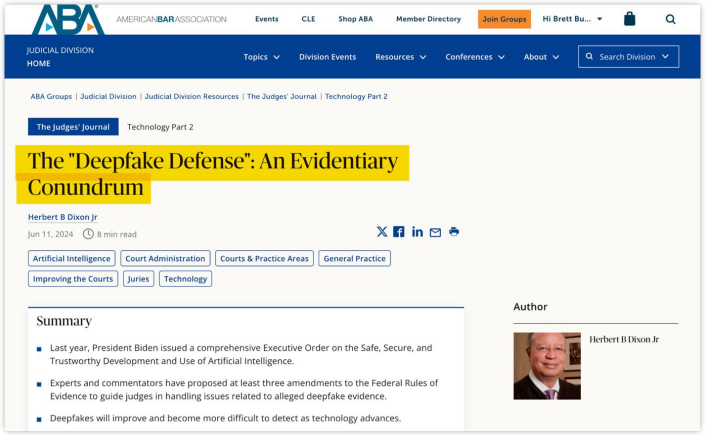
Artificial Intelligence | Court Administration | Courts & Practice Areas | General Practice | Improving the Courts | Juries | Technology

## Summary

- Last year, President Biden issued a comprehensive Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.
- Experts and commentators have proposed at least three amendments to the Federal Rules of Evidence to guide judges in handling issues related to alleged deepfake evidence.
- Deepfakes will improve and become more difficult to detect as technology advances.

Author

Herbert B Dixon Jr




# Proposed Evidence Rules Regarding Alleged Deepfakes

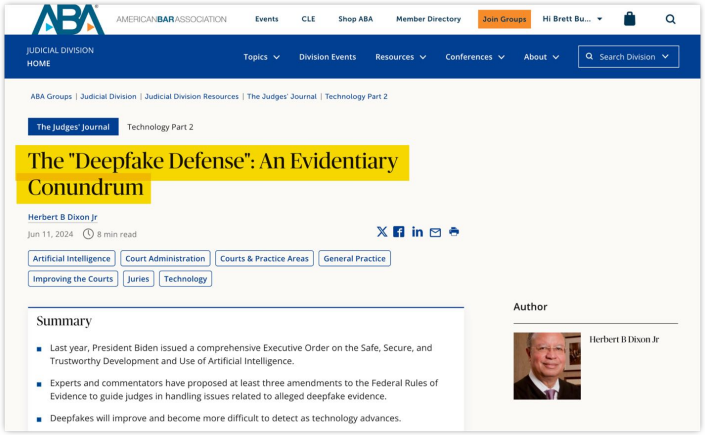
# #1

**John P. LaMonaga** - urges a higher standard to prove authenticity than merely a witness with knowledge testifying that the exhibit fairly and accurately portrays the events or scene at issue.

**Proposes new Fed. R. Evid. 901(b)(11):**

*Before a court admits photographic evidence under this rule, a party may request a hearing requiring the proponent to corroborate the source of information by additional sources.*

*A Break from Reality: Modernizing Authentication Standards for Digital Video Evidence in the Era of Deepfakes*, 69 Am. U. L. Rev. 1945, 1984 (2020), https://bit.ly/3Qt0nmW



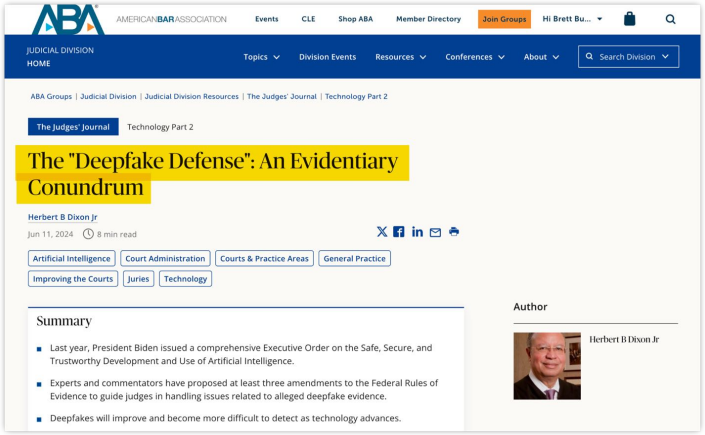

# Proposed Evidence Rules Regarding Alleged Deepfakes

# #2

**Professor Rebecca Delfino -** proposes that because of the danger of deepfakes, the judge (not the jury) should decide authenticity.

**Proposes new Fed. R. Evid. 901(c):**
*Notwithstanding subdivision (a), to satisfy the requirement of authenticating or identifying an item of audiovisual evidence, the proponent must produce evidence that the item is what the proponent claims it is in accordance with subdivision (b). The court must decide any question about whether the evidence is admissible.*

*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), https://bit.ly/4bls3Ty




# Proposed Evidence Rules Regarding Alleged Deepfakes

# #3

**Judge Paul Grimm (Ret.) and Dr. Maura Grossman** - require the challenging party to produce evidence to support the claim that the proffered exhibit is fabricated or altered

**Proposes new Fed. R. Evid. 901(c):**
*Potentially Fabricated or Altered Electronic Evidence. If a party challenging the authenticity of computer-generated or other electronic evidence demonstrates to the court that it is more likely than not either fabricated, or altered in whole or in part, the evidence is admissible only if the proponent demonstrates that its probative value outweighs its prejudicial effect on the party challenging the evidence.*

Advisory Comm. on Evidence Rules, Agenda, Proposed New Rule 901(c) to Address "Deepfakes," at 18 (Apr. 19, 2024), https://bit.ly/3JFAL2g

# Proposed Amendment: FRE Rule 901(c)

The draft concept introduces a **burden-shifting authentication process**:

- **Initial Challenge by Opponent**
  - Party challenging evidence must show **sufficient facts suggesting the evidence may have been altered or generated using AI**
- **Heightened Authentication by Proponent**
  - If that showing is made, the **proponent must demonstrate the evidence is authentic by a preponderance of the evidence**
- **Judge as Gatekeeper**
  - The court may require **additional proof of authenticity before the evidence is presented to the jury**

**Purpose of the Proposal**

- Address risks posed by **AI-generated audio, video, and images**
- Prevent **highly convincing synthetic media from misleading juries**
- Provide a **structured process for authenticity disputes**

Deepfakes today are the worst they will ever be.

Home › News › AI

# OpenAI Exec: A Year From Now, Today's ChatGPT Will Look 'Laughably Bad'

OpenAI COO Brad Lightcap suggests that future versions of ChatGPT will offer drastic upgrades. He also talks about AI's potential to replace humans jobs and stress the power grid.

By Michael Kan May 7, 2024


(Photo by Artur Widak/NurPhoto via Getty Images)

We don't know when OpenAI plans to launch GPT-5, but a company executive is already hyping up future iterations of ChatGPT as a major upgrade that put older versions of the chatbot to shame.

"I think we will look back in a year and realize how laughably bad they were," OpenAI COO Brad Lightcap said in a talk at the Milken Institute Global Conference on Monday.


Questions?

Thank you for attending!

Find me on LinkedIn or email to continue the conversation.

**Email:** sthompson@bluestarcs.com
**LinkedIn:** https://www.linkedin.com/in/sarah-thompson-ceds-b861944/

BLUESTAR