HUSCH BLACKWELL

THOUGHT LEADERSHIP

PUBLISHED: JANUARY 31, 2024

# Deepfakes: Preparing to Confront AI-Generated "Evidence" in Higher Education Investigations and Disciplinary Processes

## Service

Higher Education

## Industry

Education

## Professionals


ANNE D. CARTWRIGHT
KANSAS CITY:
816.983.8000
ANNE.CARTWRIGHT@
HUSCHBLACKWELL.COM

PETER C. ANDERSON
CHICAGO:
312.655.1500
PETER.ANDERSON@
HUSCHBLACKWELL.COM


Recent news items have highlighted advances in AI software that render it capable of generating fake images, video, and audio of a person nearly indistinguishable from the real thing. These fake media are commonly known as “deepfakes.” As deepfakes become more accessible and convincing, the possibility increases that colleges and universities will encounter alleged deepfakes in disciplinary investigations.

As we stand at the threshold of this rapidly changing AI era, educational institutions can move to protect the integrity of disciplinary and investigation procedures by anticipating how they will respond to potentially deepfaked evidence.

**What can institutions do to deal with deepfaked media?**[1]

**Train your investigation and factfinding team to monitor for and better identify obvious deepfakes**

Most of us have limited ability to distinguish deepfakes using our eyes alone. Regardless, some deepfakes contain clear mistakes when carefully inspected. According to the Department of Homeland Security, a deepfake research project run by MIT, and others, simple things to consider include unnatural lighting or shadows, blurry areas, inconsistent skin tone or texture, and lip movement that doesn’t match the audio. Witnesses and inconsistencies in evidence may also alert your team to the potential for AI-generated “evidence.”

**Consider the credibility of potential deepfakes**

Once the potential for deepfaked evidence has been noted, institutions must determine how to address that evidence. Investigators and/or fact finders may explore its credibility by considering traditional factors of credibility such as plausibility, corroboration, consistency, motive to falsify, contemporaneousness, and the context in which it was received (e.g., the credibility of the witness who provided it). Contradictory evidence outside the deepfake itself may call into question the credibility of allegedly false media. Deepfakes may not align with evidence such as a person's real face, digital information like file creation dates and other metadata, and other non-electronic evidence including disinterested witness statements. As with other potentially false information, parties and witnesses should be questioned about the veracity of possibly deepfaked evidence. In some situations, it may be appropriate to consult an expert on AI-generated media to weigh in on the matter.

**Consider what weight to give potential deepfakes**

In addition to considering factors related to credibility, where institutions may have grounds for suspicion that evidence has been deepfaked, but are not sure, individuals involved in investigation and decision-making processes must decide on the amount of weight to give the evidence. They may decide to consider it fully real, exclude it from consideration, or give the evidence an amount of consideration corresponding to its level of persuasiveness. Credibility should be factored into an assessment of weight/persuasiveness, as could factors such as whether the media appears based on personal/direct observation vs. hearsay or general knowledge. Although an institution may already have a policy that addresses allegedly false evidence, it may be desirable to tailor it specifically to deepfakes or clarify that AI-generated media are also covered.

**Utilize reliable deepfake detection technology**

Companies such as *Intel*, *Sentinel*, and *Sensity AI* are reportedly developing programs intended to identify deepfakes. Unfortunately, as these programs improve, deepfakes are simultaneously becoming more realistic and difficult to detect. Institutions should keep an eye out for any technology that is able to reliably detect AI-generated media, making sure to continuously verify the dependability of any programs they choose.

**Consider internal procedures**

Although transparency is important, where appropriate, it may be useful to keep some operational AI-detection procedures internal. Openly published detection protocols may make it easier for bad actors to circumvent safeguards against deepfakes. This could lead to more deepfakes being assessed as real evidence, and vice versa.

**Update your policies and procedures consistently as AI evolves**

While the advance of AI is exciting in many ways, it is important for institutions to prepare for the changes and risks that will come with it. As opportunities and challenges evolve, colleges and universities should continuously evaluate their policies and procedures to ensure they are effective and compliant with an AI-infused world, a subject we have also discussed in a previous legal update.

**Contact us**

If you have questions about deepfakes or other AI-related issues in the higher education context, please contact Annie Cartwright, Peter Anderson, or your Husch Blackwell attorney.

---

[1] Although this article focuses on addressing potential deepfakes in an investigation, the act of creating or distributing a deepfake may, in and of itself and depending on the circumstances, constitute a violation of institutional policy or law. We suspect many existing institutional policies already address such circumstances—and that AI simply provides another means of engaging in already prohibited conduct (like harassment or academic misconduct). We continue to monitor best practices in this area and recommend that institutions consider the potential for, and appropriate response to, such AI misconduct in ongoing policy revision processes.