# FALL 2024 E-NEWSLETTER

At Digital Mountain, we assist our clients with their electronic discovery, digital forensics, cybersecurity, and data analytics needs. For this E-Newsletter, and Cybersecurity Awareness Month, we discuss how deepfakes create expensive problems for organizations, a review of deepfake technology, and the slow march toward deepfake legislation.

## Building a Better Monster: Advancements in Deepfake Technology

Improving digital graphics and imagery has been solidly part of the entertainment, art, and advertising industries for decades. There are many classic examples of movies with now laughable stop-motion effects (1981's *Clash of the Titans* anyone?), with awkward early computer generated imagery (the Syncro-Vox technique used for the 1960's cartoon *Clutch Cargo*), and a plethora of movie posters with some scary photo editing (movie posters). Now, there's no excuse for digital artists to produce a flawed product because deepfake technology has come so far that separating the authentic from the fake is frighteningly difficult. The recent action film, *The Fall Guy*, even included a deepfake as a plot twist, evidencing that we can build an almost perfect digital monster with just a few clicks. In this article, we'll give you an update on deepfake technology and the quest to create an accurate analyzer.

"Anyone with basic computer skills and a home computer can create a deepfake," states a US government advisory document from 2020 (GAO asset). That statement holds true, perhaps more so, today. Applications such as FaceApp, DeepSwap, Reface, Deepfakes App, Zao (China), DeepFaceLab, and a multitude of others are helping users create both deepfake photos and videos, including some with voice-swapping, quickly, cheaply, and in some cases, for free. Face-swapping is achieved by uploading both the source data and the target data to a neural network application. The source is the person whose face is going to be changed, and the target is the face that will be swapped in. The data uploads are images and/or videos.

When data is uploaded to the app's cloud storage, the Generative Adversarial Network (GAN) both creates and analyzes the content being produced. Once the compiling and analysis reaches the default acceptance level of the application's programming, the resulting image or video is delivered back to the user. This can require minutes or hours of processing time depending on the quantity

of material uploaded, the complexity of the task, as well as the duration and quality of the final product. A still image, such as a selfie, takes less time than a video, and a video with voice requires even more time.

Many apps will return samples of the source and target data that was used to produce the final product, such as the variety of facial expressions used to build the deepfake, or the speech selections that were used to create the new soundtrack. From that, various apps provide the option to retrain the GAN in order to improve the final product. Adding an additional source and target content or selecting from previously uploaded content can often bolster the quality of a deepfake. Depending on the app, the user may also be provided with the "loss value," which is a number that corresponds to the accuracy of the swap. The smaller the loss value, the more the new product resembles the target. Like most AI applications, GANs are garbage in, garbage out systems.

The advancement of GAN technology, a form of AI, is driving the improvement in deepfake technology. In fact, the progression of deepfake technology is raising concerns. In the early days when deepfakes contained obvious flaws such as extra fingers on hands, lip movement that didn't align to the speech, mismatched earrings, and other defects in the desired product, it was easier to detect a deepfake image or video. Now, the technology has become advanced enough, and users have learned to train the GAN well enough, that loss values are decreasing and the detection process is more difficult.

Three solutions being used and improved to help with identifying deepfakes are watermarking, detection, and blockchain. A digital watermark is metadata – data about data – that may or may not be visible at the display level but is certainly machine readable and producible. Much like the metadata (called EXIF data) contained in a digital photograph, a digital watermark contains information about when and how the deepfake was created, including the application used to create the deepfake. Many creators are watermarking their original content so that if it is used in a subsequent deepfake, the original content may be traceable. There is no legal requirement to include digital watermarking on deepfakes currently, and while some applications are adding digital watermarking options voluntarily, it's not yet ubiquitous or industry standard. In February 2024, over 200 organizations from OpenAI and Google to Bank of America and JP Morgan Chase signed onto an AI consortium addressing AI safety, including watermarking (Reuters article). Like watermarking, registering a work on a public blockchain can help trace subsequent alterations or use in compiling a deepfake by including metadata that records the blockchain entry.

Detection of deepfakes is coming along as improvements in machine learning advances due to significant capital investments. The idea is to train the detector to analyze the content for the work of GAN technology that would be missed by the human eye. Much of this work dovetails with the work being done to improve detection of AI-generated images from Stable Diffusion, Midjourney, and others. The accuracy of detection technology, while greater than the average human, is still not failsafe. When we consider the development cycle of AI-based products, it's easy to understand that creation proceeds detection.

The scary things created as deepfakes aren't the only things created with deepfake technology. Like so many other technologies, it's a tool that relies on the user for guidance. We can use deepfakes to create hilarious face swaps of us and our pets, or to create monstrous videos that demean and defraud. The choice is up to us to choose wisely and not create a digital Frankenstein.

**Please direct questions and inquiries about electronic discovery, digital forensics, cybersecurity, and data analytics to info@digitalmountain.com.**

## UPCOMING INDUSTRY EVENTS

THE SEDONA CONFERENCE WORKING GROUP 11
MIDYEAR MEETING 2024
Atlanta, GA: October 29-30, 2024

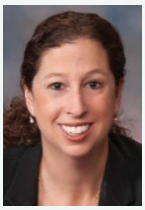
CISO-CIO FORUM
La Jolla, CA: October 30, 2024

MASTERS CONFERENCE NOVEMBER 2024
Atlanta, GA: November 12, 2024

GEORGETOWN LAW 21ST ANNUAL ADVANCED EDISCOVERY INSTITUTE
Washington, DC: November 14-15, 2024

OPENTEXT WORLD 2024
Las Vegas, NV: November 19-21, 2024

***Click here to see more upcoming events and links.***



*Digital Mountain, Inc. Founder and CEO, Julie Lewis, will be presenting at various upcoming industry events. Please send requests for speaker or panel participation for her to marketing@digitalmountain.com.*

**DIGITAL MOUNTAIN, INC.**
4633 Old Ironsides Drive, Suite 401
Santa Clara, CA 95054
866.DIG.DOCS

**www.digitalmountain.com**

***Contact us today!***

***FOLLOW US AT:***