



SUMMER 2026 E-NEWSLETTER

At Digital Mountain, we assist our clients with their electronic discovery, digital forensics, cybersecurity, and data analytics needs. For this E-Newsletter, we discuss how AI technology is pitting the forger and the authenticator in the battle over digital media in the legal arena.

Deepfakes: The Forger Versus the Authenticator

Every art forger and every art authenticator study the same three elements to hone their craft: brushstroke, pigment, and provenance. The forger develops skills to make a fake indistinguishable from the original. The authenticator uses identical knowledge to expose the forgery. Digital forensics investigators now face an AI-era version of that same adversarial relationship when dealing with digital media. Understanding this rapidly growing niche of digital forensics starts with a single distinction, Generative AI versus discriminative AI.



Generative AI is the forger who can also be a genuine artist. It creates or enhances content, for legitimate or sketchy purposes, in the form of digital images, audio, and video by learning the patterns, structures, and distributions in existing data, then using those to produce new material that looks or sounds authentic. Diffusion models are the workhorse of Generative AI for media, and the mechanics are worth understanding. Training happens in two process directions, forward and reverse. In the forward process, the model takes real images and progressively adds random noise across many steps until the original structure is destroyed and what remains is statistically indistinguishable from pure noise. In diffusion models, "noise" looks like television static, or "snow." If you magnify the image to the degree where you can see the pixels, what you see is a field of pixels where each one has been assigned a random color value independent of its neighbors, so there's no edge, no shape, no gradient a human eye can latch onto. Every pixel is uncorrelated with the pixel next to it, so your visual system, which is built to find edges and patterns, finds nothing to grab onto. It's literally visual "noise."

The reverse process is where the actual intelligence lives. The model is never shown a specific image being enhanced, there's nothing sitting in a database to compare. During training, the model is fed millions of images and instructed to develop a set of statistical regularities, what edges tend to look like, what a face's proportions tend to be, what fabric texture tends to look like when blurred.

Those regularities get compressed into the model's parameters. The training images themselves are never stored in the model. A set of statistical regularities is called a learned prior. A neural network, either a U-Net-based (as in the original Stable Diffusion) or a transformer-based (as in Sora), uses a learned prior to predict what's noise versus what's actually in the image. Run iteratively, that prediction sharpens a noisy image into something coherent, step by step.

That distinction matters because it's the mechanical reason "enhanced" and "recovered" aren't the same category of evidence. Take super-resolution, rather than magnifying existing pixels, the model generates new fine-grained details such as pore texture, hair strands, and fabric weave, based on what those features statistically look like in training data. Two different models given the same blurry photo will generate two different faces, both plausible, neither necessarily correct. Deblurring works the same way, it predicts the sharp edges "lost" to motion or focus blur based on learned patterns. Inpainting is the clearest case, filling a torn, watermarked, or edited-out region of an image using zero information about what was actually there, generating only what's likely, given the surrounding pixels. All three target "what's probably here" and not "what was actually here."

Discriminative AI is the authenticator. It doesn't create, it examines and classifies, asking the same question a museum authenticator asks, is this what it claims to be? Four tools do that work, each examining a different file layer, though each comes with real limitations worth knowing before relying on it. **Metadata analysis** examines information invisible in the image but embedded in a file, including camera model, GPS data, timestamps, and edit history. It's easy to strip or falsify, so it's rarely dispositive alone, but an absence or inconsistency is itself evidence.

Error Level Analysis (ELA) works with JPEG compression. The technique re-saves an image once at a known quality level and compares that re-saved version against the original. Regions that were part of the image's original compression settle into a stable "equilibrium" and appear dark in that comparison. Regions with a different compression history appear bright, standing out as anomalies, because they are inserted, edited, or spliced in after the initial file is saved. ELA is an established and widely used in forensics tool, but it has a documented blind spot, it can't reliably distinguish an edited region from a region that's naturally high-frequency, like sharp edges, fine texture, or small text, both of which appear bright in an ELA map even in a completely authentic photo. For that reason, ELA findings need corroboration from other techniques rather than standing alone.

Spectral analysis examines a file in the frequency domain rather than the pixel domain. Against GAN-generated images (Generative Adversarial Network, two networks trained against each other) this method has a well-established track record. GANs' resolution-scaling architecture creates periodic patterns in the high-frequency spectrum that real cameras don't produce, giving investigators a reliable fingerprint to search for. That track record doesn't carry over cleanly to diffusion models, though, which are the more common generative architecture in the current wave of image and video synthesis. Diffusion's denoising process doesn't reliably leave the same frequency-domain signature GANs do, and detectors trained to catch GAN artifacts frequently fail to generalize to diffusion-generated content. Spectral analysis remains a genuinely useful tool, but its reliability depends heavily on knowing which generative architecture produced the content under review.

AI-based deepfake detection flips the generative architecture back on itself, with models trained on paired real and fake datasets to recognize the subtle clues diffusion and GAN-based generation tend to leave, for example inconsistent boundaries where a swapped face meets the original frame.

C2PA (Coalition for Content Provenance and Authenticity) Content Credentials work differently from all four. Rather than detecting alteration after the fact, they establish a cryptographically signed record at capture and every edit afterward. This is not digital watermarking, but more comparable to a signed metadata record that travels with the file itself.

As the forger's tools improve, the authenticator's job gets harder. Both halves of the response must keep pace. Discriminative AI, the detection methods described above, must advance in lockstep with generative AI's capabilities, not lag behind them. And chain of custody matters more, not less, as the technology evolves. Preserving original files, documenting any AI tool that has touched evidence before preservation, involving experts early, and preserving metadata throughout aren't just best practices, they're what stands between evidence and a defensibility challenge. Technology is necessary for authentication as our human eyes are not equipped to catch the forger.

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

UPCOMING INDUSTRY EVENTS

ILTACON 2026

Nashville, TN: August 23-27, 2026

AI TECHWORLD 2026

Santa Clara, CA: September 1-3, 2026

NATIONAL EDISCOVERY LEADERSHIP INSTITUTE CONFERENCE 2026

Kansas City, MO: September 24-25, 2026

RELATIVITY FEST 2026

Chicago, IL: September 29, 2026 - October 1, 2026

THE AI CONFERENCE 2026

San Francisco, CA: September 29, 2026 - October 1, 2026

[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

Contact us today!

www.digitalmountain.com

FOLLOW US AT:

