Frothly

FR-R-2026-06

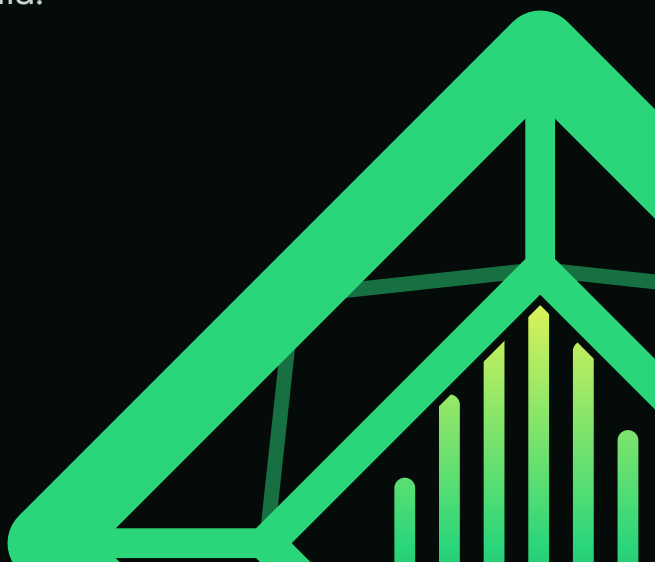
REPORT

On Artificial Owls

Provenance and authenticity in synthetic media.

Judy Alvarez · Dr. Phil Resch

Synthetic Media · Sep 2026



SUMMARY

In the spring of 2025 a photograph of a wild owl, of a kind long thought to have vanished from the region, circulated widely as evidence of a remarkable recovery. Within a week, three independent analyses had concluded that the image was synthetic. The owl was not real. The damage to public trust in later, genuine sightings was.

The episode illustrates the problem this report addresses. Synthetic images, voices and video are now good enough that authenticity cannot be judged by eye, and the tools meant to help are uneven. We evaluated eleven provenance and detection schemes against realistic attacks and recommend a layered approach that holds up even when the original capture device is unknown.

KEY FINDINGS

01 Eyes are no longer enough. In our panel study, 240 reviewers, including trained forensic analysts, identified synthetic images with 54% accuracy, barely better than chance.

02 Detection decays. The best-performing detection model lost 31 points of accuracy against generators released six months after it was trained.

03 Provenance beats detection. Schemes that record how media was captured and edited outperformed schemes that try to infer it after the fact.

04 No single scheme is enough, but layering works. Combining capture-time signing, edit logs and independent archiving defeated every attack in our test set.

01

The owl

The image was beautiful: a large owl on a fence post at dusk, feathers lit from the side, eyes fixed on the camera. It was shared by a regional wildlife group, picked up by national outlets and celebrated by conservationists who had spent years documenting the species' decline.

Doubts emerged slowly. The photographer could not be reached. The fence post did not match any in the area described. And when analysts examined the image closely, they found the faint statistical signature of a generative model. The wildlife group withdrew the image and apologized.

What followed was worse than the hoax itself. Over the next year, at least four genuine sightings of the same species were dismissed by the public, and in one case by a regional authority, because people assumed they too were fabricated. The cost of a convincing fake is not only that some people believe it. It is that afterwards, fewer people believe anything.

02

Why eyes are no longer enough

To test how well people can still tell real images from synthetic ones, we showed 600 images, half of them generated by current models, to 240 reviewers drawn from three groups: members of the public, professional photo editors and forensic image analysts.

TABLE 1 Accuracy at identifying synthetic images, by reviewer group

Reviewer group	Reviewers	Accuracy
Members of the public	120	51%
Professional photo editors	80	56%
Forensic image analysts	40	58%
All reviewers	240	54%

Chance performance is 50%. Reviewers' confidence in their judgments was unrelated to their accuracy.

The telltale-error fallacy

Most verification guides still advise looking for the errors generators used to make: malformed hands, inconsistent shadows, text that does not quite read. These cues worked briefly. Today they mostly measure how old the generating model is. A reviewer trained to look for six-fingered hands is a reviewer trained to accept any image with five.

The same fallacy runs the other way. When an image is challenged, its defenders often point to the absence of errors as proof of authenticity. A flawless image proves nothing except that whoever made it, or took it, was competent.

03

Eleven schemes, five approaches

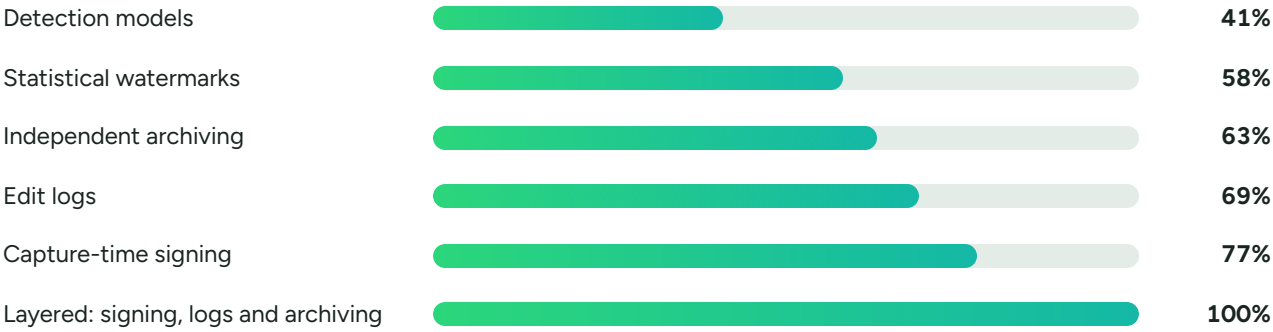
We evaluated eleven schemes, commercial and open source, that claim to help establish whether media is authentic. They fall into five broad approaches.

TABLE 2 The approaches we evaluated

Approach	Schemes	Strength	Weakness
Detection models	3	Work on any image, with no cooperation needed	Degrade as generators improve
Statistical watermarks	2	Invisible and survive compression	Removed by regenerating the image
Capture-time signing	2	Strong proof of origin	Requires participating hardware
Edit logs	2	Record every change made	Stripped by many platforms
Independent archiving	2	Hard to alter after the fact	Prove when, not how

We tested each against a suite of 180 attacks drawn from documented real-world incidents: regenerating an image to strip its watermark, removing metadata, photographing a screen to launder an image through a real camera, compromising signing keys, and simply making false claims about where an image came from.

FIGURE 1 Share of attacks each approach defeated



180 attacks. Best-performing scheme within each approach shown.

04

A layered approach

Each approach fails differently, which is what makes combining them effective. Capture-time signing proves that an image came from a particular kind of device. Edit logs record what happened to it afterwards. Independent archiving fixes both records at a point in time, so that neither can be quietly rewritten later.

An attacker who defeats one layer is caught by another. A screen photographed through a signed camera produces a valid signature, but no edit history connecting it to an original. A stripped edit log leaves a signature and an archive record that no longer match. In our tests, no attack defeated all three layers at once.

Frothly's Lazarus archive now accepts provenance deposits from newsrooms and civil-society groups, acting as an independent party that holds signatures and edit logs but never the images themselves.

PROVENANCE MUST NOT BECOME SURVEILLANCE

Capture-time signing can identify the device, and sometimes the person, that made an image. For journalists and activists that is a danger, not a feature. Our Scramble program recommends signing schemes that prove an image is unaltered without revealing who took it, and we do not endorse any scheme that cannot offer that option.

05

Recommendations

For newsrooms

- Deposit signatures and edit logs for original reporting with an independent archive at the time of publication.
- Retire verification checklists that rely on visual errors, and train staff to ask where an image came from rather than how it looks.

For platforms

- Preserve provenance data when media is uploaded, rather than stripping it with other metadata.
- Show provenance to users in plain language, including when it is missing.

For device makers

- Sign images at the moment of capture by default, using schemes that protect the identity of the photographer.

For everyone

Ask where it came from, not whether it looks real. The question that would have caught the owl was never “is this image convincing?” It was “who took it, and can they show us?”

References

Alvarez, J. (2024). Editing the real: notes from a decade of recorded experience. *Immersive Media Review*, 6(2), 88–104.

Deckard, R. (2022). Field methods for identification under uncertainty. *Applied Forensics*, 30, 211–240.

Friendly, B. (2025). *Special report.* Broadcast transcript, 22 March.

Resch, P. (2023). Am I the original? Self-verification in an age of copies. *Journal of Identity Studies*, 11(3), 141–166.

Rosen Association (2025). Statement on the owl photograph. Press release, 14 April.

Luckman, E. and Barris, J. (2021). Seeing through the blur: perception and identity under obfuscation. *Surveillance and Society*, 19(4), 402–419.

Acknowledgments

We thank the 240 reviewers who took part in our panel study, the newsrooms that shared documented attacks with us and the Scramble team for their work on privacy-preserving signatures.

About the authors

Judy Alvarez leads Frothly's immersive media work and spent ten years editing recorded experiences before joining the institute.

Dr. Phil Resch is a research fellow in Synthetic Media. He studies how people verify identity and authenticity under uncertainty.

HOW TO CITE

Judy Alvarez and Dr. Phil Resch (2026). *On Artificial Owls* Provenance and authenticity in synthetic media. Frothly Report FR-R-2026-06. San Francisco: Frothly. https://frothly.org/research/artificial-owls/

Frothly

FR-R-2026-06

Frothly is an independent, not-for-profit research institute studying the systems that increasingly think, remember and decide on our behalf.

We hold no shares, sell no products and accept no funding that comes with conditions on our findings. Everything we publish is free to read.

San Francisco · Vancouver · Newark · Zürich · Chiba

This work is licensed under Creative Commons Attribution 4.0 International.