Marking is not Detection

On 2 August 2026, key transparency obligations under the EU AI Act begin to apply.

Much of the discussion has focused on marking AI-generated content. That focus is understandable. If people are to distinguish between human-created and AI-generated media, some form of machine-readable provenance is essential.

But marking and detection are not the same thing.

The distinction matters because they solve different problems.

Provenance is a Layered System

Most current approaches to AI transparency rely on one or more of three mechanisms:

1. Provenance metadata, such as signed assertions and Content Credentials.
2. Embedded signals, including watermarking.
3. Verification systems, including fingerprinting and logging.

Each layer serves a different purpose.

Metadata provides information about a file's origin and history. Watermarks embed information directly into the content itself. Verification systems allow content to be compared against trusted records after it has moved through the world.

These technologies are complementary, not competing.

The Problem of Travel

Content rarely stays where it was created.

A voice actor records an audiobook sample. The file is uploaded to a marketplace. It is transcoded into a different format, downloaded, edited, excerpted, shared across platforms, and eventually appears somewhere entirely unexpected.

At each stage, provenance information may be preserved, transformed, or lost.

Metadata can disappear when files are converted or exported through systems that do not preserve it. Watermarks can be weakened by compression, editing, or repeated transformations. Neither observation is a criticism of those technologies. They were designed to solve particular problems, and they solve them well.

The challenge is that content continues to exist even when some of those signals no longer do.

The question then becomes: how do we verify what we are looking at?

The Direction of the AI Act

The EU's Article 50 Code of Practice recognises this distinction.

The Code prioritises machine-readable marking techniques such as metadata and watermarking. But it also acknowledges that these mechanisms may not always be sufficient. For that reason, it includes fingerprinting and logging facilities as an additional layer when other marking techniques prove ineffective or unavailable.

The significance is not that one technology has failed.

The significance is that the regulatory framework increasingly treats provenance as a layered system rather than a single technical solution.

Creating provenance and verifying provenance are different functions.

Verification After the Fact

A provenance mark answers the question:

"*What information was attached to this content when it was created?*"

Verification answers a different question:

"*What can we still establish about this content now?*"

Those questions are often asked at different times by different people.

A creator may attach provenance at the point of creation.

A platform, publisher, rights holder, journalist, or member of the public may need to verify provenance weeks, months, or years later.

The systems required for those tasks are not identical.

Why This Matters

As synthetic media becomes more common, transparency cannot depend on a single mechanism surviving every transformation.

Robust provenance requires multiple layers working together:

- Metadata to carry information.
- Watermarking to embed information.
- Verification systems to help establish provenance when other signals are incomplete, unavailable, or disputed.

The future of transparency is unlikely to be a choice between these approaches.

It will be a combination of them.