Authentitas

The World Has A Fake Problem

The Authentitas Fix. How It Works. Why It Matters.

March 2026

Contents

What's The Problem?

In May 2023, a single AI-generated image of an explosion near the Pentagon appeared on financial news feeds. No explosion had happened. Before anyone confirmed it, markets briefly erased an estimated $500 billion in value.

In April 2025, an unverified X post falsely claimed that President Trump was considering a 90-day pause on tariffs. CNBC read the headline live. The S&P 500 surged nearly 6% in ten minutes. Within an hour, the White House confirmed it was false. **An estimated $2.4 trillion had moved on a single unverified post** from an account with no identity and no accountability.

Fake videos. Fake news sites. Fake experts. Fake government announcements. All of them designed to look like the real thing. All of them getting harder to spot by the day.

Here's the part that surprises most people. The fakes aren't winning because people are stupid. They're winning because the tools we built to catch them don't work.

That's the problem. Authentitas has a fix.

How Bad Is It?

The Numbers

In 2023, there were roughly 500,000 fake AI-generated video files circulating online. By the end of 2025, that number had reached an estimated eight million. **A fake video attempt now happens every five minutes.**

There are more than 2,000 fake news websites designed to look exactly like real ones – same fonts, same layouts, same style of writing. In 2023, there were 49. That's a fortyfold increase in two years.

Global financial losses from deepfake-enabled fraud exceeded $200 million in the first quarter of 2025 alone, and rose to $347 million in the second. By the end of 2025, total deepfake-related losses had passed $1 billion.

Who's Behind This?

These aren't accidents. They're not pranksters. This is organised. It runs like a business.

The most thoroughly documented operations are state-sponsored. Russia's Storm-1516 network has conducted at least 77 campaigns across Europe since 2023. The Doppelgänger operation replicates institutions like *Der Spiegel* and *Le Monde*. China and Iran run operations at comparable scale.

But state actors aren't the whole story. The 40 US websites most responsible for spreading election disinformation in 2022 generated an estimated $42 million in advertising revenue. Political and financial incentives, it turns out, produce much the same result.

What's Actually Happening?

The Thing We Miss

When most people hear "fake content", they think of deepfake videos – politicians appearing to say things they never said.

A real problem. Not the biggest one.

Fake video is just one piece. There's also the fake news site that reports on it. The fake expert who comments on it. The thousands of automated and anonymous accounts that share it. And sometimes, even a fake fact-checking website that "verifies" it.

By the time a real institution or publication spots the fake, it's already been seen by millions of people.

The system of disinformation is designed to be faster – exponentially – than any viable response.

It's Not Only About News

We could assume this is just a massive threat to journalism. It isn't.

In academic publishing, fraudulent papers are doubling every 1.5 years – ten times faster than legitimate research. More than 250,000 cancer studies have been flagged as potentially fabricated. **Life and death decisions are being made on findings that may be entirely false.**

In December 2024, Romania's Constitutional Court annulled the first round of its presidential election over a coordinated TikTok influence operation - an EU first. Coordinated fake accounts, algorithmic manipulation, and fabricated content had distorted the result beyond the point where it could stand.

Everywhere that information carries consequences – which turns out to be pretty much everywhere – the inability to distinguish real from fake is causing real and destructive damage.

The Liar's Dividend

There's a further problem. Subtler than the fakes. And in many ways far more damaging.

Researchers call this "the liar's dividend". Once the reality of synthetic content is widely understood, bad actors – in fact, anyone who wants to – can run the same trick in reverse. Dismissing genuine footage, authentic recordings, real evidence - as "just fake news".

So the threat isn't only that people will believe false things. It's that they'll stop believing true ones.

Can't We Just Catch Fakes?

The Obvious Idea

The solution seems simple: build better detection tools. AI creates fakes; AI detects fakes. Problem solved.

It sounds like the right answer. The research proves otherwise.

The Evidence

In laboratory conditions, the best AI detection tools are impressive. They correctly identify fake content 96 to 98 per cent of the time.

In the real world, the same tools drop to roughly 50 per cent accuracy. That's the same result you'd get from tossing a coin.

Human beings can't do better. In a 2025 study, people correctly identified fake videos at a rate of 51.2 per cent - again, coin-toss level. For the most convincing fakes, human accuracy fell to 24.5 per cent. People got it wrong three times out of four.

The Columbia Journalism Review concluded that detection tools "do not yield clear, unequivocal results" and "can only calculate a likelihood".

Why It Won't Get Better

This isn't a problem that more effort will solve. Because it's structural.

Detection works by finding patterns in fake content. The people making fake content know this, so they keep changing their methods to eliminate those patterns. Every time detection improves, generation improves faster.

The people trying to catch fakes are always chasing. The people making fakes are always ahead.

And there's a fundamental cost problem, one that runs in the wrong direction. **Creating a convincing fake now costs as little as $1.33. Detecting it requires ever-more resources.** This toxic divergence continues to accelerate – and no amount of investment in detection has reversed it.

Isn't There A Better Question?

Professor Siwei Lyu, director of the Media Forensic Lab at the University at Buffalo, put it plainly in *The Conversation* in December 2025: "Simply looking harder at pixels will no longer be adequate."

He's right. But what does that actually mean?

What Detection Asks

Detection asks: "Is this fake?"

It's a hard question. The answer changes as technology improves. It requires examining millions of pieces of content one by one, after they're already in circulation. And as we've seen, it keeps getting the answer wrong.

What Authentication Asks

Authentication asks a different question:
"Does a real human being stand behind this?"

That's a much simpler question to answer. And the solution doesn't change as technology improves – because it doesn't rely on spotting patterns in the content. It relies on proof that a real person created it.

How It Works

Trust in information operates at two levels: authenticity and accountability. **We want to know who wrote something. And we want to know who takes responsibility for publishing it.** Authentication proves both – not as a matter of brand reputation or assertion, but as verifiable facts.

Authors verify their identity, one time at the point of publication, using face and government-issued ID. That verification is cryptographically bound to the content itself. This creates a permanent, tamper-proof record of who created it and who stands behind it. Anyone can check this. No one can alter it after the fact.

The absence of that record is information too. It doesn't mean the content is false. But it means no verified human, and no accountable institution, has taken responsibility for it.

Two Things Authentication Doesn't Do

Both matter, so they're worth saying clearly.

Authentication doesn't mean "no AI was involved." A journalist who uses AI to help with research or drafting, but who reviews and takes responsibility for the final piece, can authenticate that content. The variable that matters is accountability, not production method.

And authentication doesn't guarantee accuracy. It proves who created the content, not that the content is true. What it restores is the chain of human responsibility – what audiences actually need in order to make informed judgements.

Think About Hallmarking

When you buy a gold ring, there's a tiny stamp on it. That means an independent authority has tested the gold and confirmed it's real.

The hallmark doesn't tell you whether the ring is beautiful, or worth the price. It just tells you the gold is real.

The absence of a hallmark is itself a signal.

That's what authentication does for content. Not "this is true". Just: "a verified human is responsible for this". And the absence of that credential tells you something important.

Where Does Authentitas Fit?

What We're Building

Authentitas is building the authentication infrastructure described above. A system that binds verified human identity to content through government-grade biometric verification, and a cryptographic record that travels with the content and cannot be forged.

We operate as independent, third-party infrastructure – not owned by any publisher, platform, or government. Because the value of the credential depends entirely on that independence. **A hallmarking authority owned by a jeweller is worthless.**

Where We're Heading

We're starting with journalism, for two reasons.

The first is that the problem is most acute there. **Public trust in news is collapsing across most markets. Publishers are losing audiences and revenue.** The need is immediate and well-documented.

The second is that Scandinavian publishers – in Sweden, Norway, and Denmark – offer the ideal launch pad. News trust in the region is among the highest in the world. Subscription culture is strong. There is genuine collaboration between publishers.

But journalism is only the beginning.

Academic publishing, market research, legal documentation, financial research, healthcare communication, government announcements – everywhere that information carries consequences – the question "did a verified human stand behind this?" matters. Courts are already encountering fabricated evidence without consistent tools to assess it.

The infrastructure we're building answers that question, the same way, across all of these industries. The current combined value of the critical information sectors can be estimated at $375 bn.

Why This Now?

Regulation Is Happening

Governments around the world are starting to require that AI-generated content be labelled and traceable. The European Union's AI Act requires this by December 2027. China's equivalent framework came into force in September 2025. South Korea's AI Basic Act in January 2026.

These rules have a flaw, and it's important to understand it. They govern the compliant. A news organisation that labels its AI-assisted content will comply.

A disinformation operation running fake videos out of a server farm will not label anything. And detection – as we've seen – cannot force them to.

Authentication supplies what regulation alone, as it stands, cannot. It creates a positive credential for content that is genuinely human-accountable. Labelled AI content has a disclosure marker. Authenticated content has an accountability credential. Content that is neither sits in explicit uncertainty – exactly where it should be.

Social Media Have Stepped Back

In January 2025, Meta – which owns Facebook, Instagram, and WhatsApp – ended its professional fact-checking programme across all three platforms. Its chief executive acknowledged the change would mean catching "less bad stuff".

X, formerly Twitter, had already moved in the same direction. After its acquisition in 2022, professional content moderation infrastructure was dismantled and replaced with a user-driven "community notes" model.

The results are measurable, and dismal. Only 11 per cent of submitted notes reach "helpful" status. The average time for a note to achieve the required cross-perspective agreement is 15.5 hours – while disinformation spreads unchecked.

Every Election Is a Rehearsal

Elections are where disinformation operations do their best work. And 2026 and 2027 are unusually busy.

Sweden, Brazil, and the US midterms in 2026. France, Kenya, and Poland in 2027. More than forty national elections are scheduled across 2026 alone, representing over 1.6 billion voters.

That matters for two reasons. The first is obvious. Elections are high-stakes targets, and disinformation operations know it.

The second is less obvious. **Each election is also a development cycle. Techniques refined in one contest are redeployed in the next,** at greater scale and lower cost. So every election without authentication is another round of performance improvement for the other side.

The Window

Regulations are demanding more. Social media have abdicated responsibility. The gap between what the rules require and what actually happens is simultaneously widening. That shifts the burden to the people who produce trustworthy content. Which means they need tools to prove it.

There's a window – between now and roughly the end of 2028 – during which authentication infrastructure can be built, proven, and established as the standard.

After that window, the standards are set.

This is how infrastructure markets work. The people who build the rails own the rails. **The question isn't whether authentication infrastructure will be built** – the trajectory of threat, regulation, and market pressure makes that effectively certain. **The question is when. And on whose terms.**

In A Nutshell?

The world has an overwhelming fake content problem.

The technology to make fakes will continue to outrun the technology to detect them.

Authentication doesn't try to detect fakes. It proves – for any publisher, and for anyone who needs to know – that a real person, and an accountable organisation, stand behind their content.

Governments are starting to require this. Platforms have stepped back from doing it. Publishers need it to survive.

Authentitas is building this infrastructure .
The window to establish the standard is now.

Key Sources

World Economic Forum, Global Risks Report 2026 – *Ranks misinformation and disinformation as the second most severe short-term global risk.*

Reuters Institute Digital News Report 2025 – *Tracks news trust levels across global markets, from Finland at 67 per cent to the United States at approximately 32 per cent.*

Resemble AI, "2025 Deepfake Incident Report", April and November 2025.

Deepfake-Eval-2024 – *Benchmark analysis documenting detection accuracy drops of 45 to 50 per cent in real-world conditions compared to laboratory testing.*

VIGINUM, Storm-1516 Analysis (May 2025) – *French government documentation of 77 disinformation operations and the five-phase operational model.*

Columbia Journalism Review, AI Detection Assessment (2025) – *Concluded that detection tools "do not yield clear, unequivocal results" and "can only calculate a likelihood."*

Siwei Lyu, "Deepfakes levelled up in 2025: Here's what's coming next," The Conversation (December 2025) – *Leading deepfake researcher's assessment that the meaningful line of defence will shift to infrastructure-level protections.*

NewsGuard AI News Tracking (2025) – *Documents growth of AI-generated news websites from 49 in May 2023 to over 2,089 by 2025.*

Westwood, S., Proceedings of the National Academy of Sciences (November 2025) – *Demonstrates AI corruption of survey research at scale; synthetic respondents passing 99.8 per cent of attention checks undetected by every current detection method.*

Industry Valuations

News media WAN-IFRA, *World Press Trends Outlook 2024-2025, January 2025*

Academic publishing Market.us, *Global Academic Publishing Market Report, January 2026*

Market research ESOMAR, *Global Market Research Industry Report 2024, cited in Backlinko, 23 Key Market Research Statistics for 2026, January 2026*

Legal information services Mordor Intelligence, *Global Legal Tech Market Report, June 2025*

Financial data services Research Nester, *Financial Data Services Market Report, December 2025*

Medical and clinical publishing Simba Information / Freedonia Group, *Global Medical Publishing 2024-2028*

authentitas.com

Authentitas