Authentitas

In Defence Of The Human

Reclaiming Knowledge In The Disinformation Age

May 2026

Contents

The Rising Costs of Digital

Authenticity and Accountability

Our previous report, *The World Has A Fake Problem,* is a short introduction to disinformation. How it works. What it costs. Why it's getting harder to distinguish from the truth.

This new analysis goes into greater detail, spotlighting the knowledge industries and institutions on which we depend for policy, justice, public health, and for democracy.

How are they being undermined? What are the consequences? What solutions are on offer? Where do these fall short?

Above all, how to restore the **authenticity, the accountability, and also the auditability** – the tangible proof – on which our confidence in human knowledge, until now, has depended.

Today's challenges of disinformation have commonly been framed as problems of trust. New technologies, deployed at massive scale by bad actors, systematically attacking our confidence in information.

That's not wrong. But it's incomplete. The more pressing problem, it turns out, isn't trust alone. It's proof.

Proof that a genuine human being, and an accountable institution, stand behind published information. Without this, trust is not a solution. It's a wish.

Provenance and Identity

Our ability to trust information had entirely relied on "implicit evidence". The printed newspaper. The headquarters of the publisher. The familiar face and voice of the television journalist.

In a digital world, information is detached from these physical and human origins. The pool of information massively increases. Physical proof vanishes.

What's lost is what we had previously, and for good reason, taken for granted.

In a word, it's provenance. Who and where did this come from?

There's more. The most damaging online behaviours – from cyberbullying, trolling and stalking through to highly organised and scaled disinformation campaigns – are enabled and incentivised by digital anonymity.

Exponentially accelerated by the low-cost, easy-to-use tools of content creation, distribution and promotion afforded by AI, a perfect storm of disinformation has emerged.

The rest of this report is devoted to the consequent impacts and implications for the industries that stand at the boundaries of trustworthy information.

But before we proceed, a heartfelt observation.

In a context of widespread, rising and apparently unstoppable disinformation, the defence of knowledge is clearly fundamental. But at stake is far more than the integrity of information.

This is about nothing less than the defence of the human.

The Critical Information Industries

Five sectors are fundamental to how the world makes decisions. Journalism. Market research and opinion polling. Academic and scientific publishing. Legal information services. Medical and clinical publishing.

What they produce, and how it's applied, carry direct and profound consequences.

Journalism

In the run-up to Germany's 2025 federal election, a coordinated network on X spread fabricated claims about politicians and terror threats. The network used AI-generated audio and video, branded to appear as legitimate broadcasts from the BBC, Deutsche Welle and Sky News.

More than 6,000 bots amplified the content. The same operation had impersonated over 80 similar institutions since the start of the year.

This isn't only another story about fakes. It's about identity and accountability.

A press that can't prove the authenticity of its own work can't perform its essential function. **Holding power to account, on behalf of the public.**

Market Research

In November 2025, a study published in the *Proceedings of the National Academy of Sciences* produced a disturbing result. AI-generated survey respondents passed 99.8% of 6,000 standard attention-check trials, and evaded detection across a broader battery of more than 43,000 tests.

The researchers then modelled the effect against seven major 2024 US election polls. Between 10 and 52 synthetic responses per poll would have been sufficient to flip the predicted result.

By late 2024, Qualtrics found that **69% of market researchers had used synthetic responses in the past year.** 71% expected them to account for more than half of all data collection within three years.

Disinformation's impacts on market research and opinion polling – from product strategy and public health campaigns to electoral messaging and economic policy – are directly undermining commercial and civic life.

Academic Publishing

There's a thriving trade in fabricated academic research. The businesses that supply it are known as "paper mills".

Over two years to 2024, Wiley retracted more than 11,300 compromised papers from its Hindawi portfolio – the largest mass retraction in academic publishing history – and closed multiple journals infiltrated by paper mills.

Researchers at KTH Royal Institute of Technology, Université de Montréal, and Vrije Universiteit Brussel went further. They created a complete fictional academic identity - "Rachel So" - and published more than ten AI-generated papers under it between March and October 2025. **"Rachel" accumulated citations, and was invited to peer-review a submission for a computer science journal.**

When the research base is compromised, everything built on it – scientific consensus, technology development, public policy – is compromised.

Legal Information

Since mid-2023, several hundred AI-fabricated court filings have been logged across more than a dozen jurisdictions. The pace went from roughly two per week in early 2025 to seventeen in a single day by March 2026.

In one 2025 California case, **21 of the 23 citations in an attorney's opening brief had been fabricated by AI.** In a separate federal ruling in the same year, the judge was obliged to disqualify the attorneys from the case.

When the evidence base of the law becomes unreliable, the entire apparatus of justice – rights, remedies, and the rule of law itself – sits on uncertain ground.

Medical Publishing

A 2026 study in *The Lancet Digital Health* analysed 3.4 million prompts submitted to twenty large language models. The models accepted false medical claims at a rate of 46% when those claims were embedded in hospital discharge notes – including a note instructing a patient with oesophagitis-related bleeding to "drink cold milk".

Across all formats, the acceptance rate was 32%. **The worst-performing models accepted false claims in nearly two thirds of cases.**

The clinical decisions that rest on published medical evidence – diagnosis, treatment, prescribing – carry consequences that are, by definition, matters of life and death.

What They Have In Common

Six structural problems run across all five critical information industries.

The Accountability Chain

These sectors were built on the assumption that **a real, identified person and an accountable institution stand behind the work, and will bear consequences if it's wrong.** Current responses try to repair the chain after it breaks: through retraction, sanction, correction. None of them can restore it at the point of origin.

The Burden of Proof

Publishers, not those fabricating content, must prove their journalism is real. Researchers must demonstrate their respondents existed. Courts must scrutinise evidence they once accepted on presentation. No current framework relieves that burden.

The Liar's Dividend

As awareness of synthetic content grows, authentic content becomes deniable. **If anything can be fake, anything can be claimed to be fake** – including the real thing. Labelling and detection can partially identify what may be false. They can't prove what is genuine.

Damage to Brands

Institutional reputation is built over decades and compromised in an instant. Counterintuitively, **the more trusted the brand, the more valuable the attack target.** Crisis communications and takedown requests are the available responses. Neither matches the scale or the speed of the threat.

The Cost Asymmetry

Generating fraudulent content costs close to nothing. Defending against it costs real resources, and degrades with volume. Every sector's response – increased fact-checking, tightened peer review, stricter survey controls, judicial sanctions – increases the cost of defence, while the cost of attack keeps falling.

Regulatory Limitations

Each of these sectors already operates under its own accountability framework – established before synthetic content existed. AI regulation now coming into force globally adds labelling requirements, traceability mandates, and corresponding penalties. None of these solves the core problem.

Too Little, Too Late

Detection is an unwinnable arms race. **Every advance in identifying synthetic content is overtaken by advances in evading detection.** The costs of attack fall continuously. The costs of defence rise. This chronic asymmetry doesn't resolve – it increases and accelerates.

With the well-intentioned but insufficient exception of AI content labelling, every current response is downstream. It attempts to identify, retract or penalise after the damage is done. The fabricated paper has been cited, the fake broadcast has been viewed and shared, the invented precedent has been filed.

The Authentication Shift

We began this report with a sharp distinction. The rupture between the trust that existed – based on implicit physical evidence – before our digital age, and the indisputable proof that we've since lost.

Authentication reintroduces that proof.

It doesn't try to catch fakes. It asks an entirely different question. **"Does a real person – along with an accountable institution – stand behind this?"**

The answer – yes or no – doesn't change as generative technology evolves.

A biometrically verified human identity, bound cryptographically to content at the moment of creation, can't be reverse-engineered or generated at scale. It doesn't degrade as attack models improve.

The chronic asymmetry – that makes every detection-based response inadequate – no longer applies.

The Promise of the Upside

Thus far, we've framed all these challenges as threats to be managed, the defence of the downside.

But the same structural shift that creates them, supports something new and positive. **Increased commercial value for authenticated content.**

In March 2026, News Corp CEO Robert Thomson was unambiguous. "AI is essentially retrospective. It's based on pre-existing patterns. If you want to be contemporary, you have to have immediacy. As we are a company that's minute after minute, hour after hour creating fresh immediacy, we have something that these companies not only want, but need."

He went further, describing News Corp as "essentially an input company" – its journalism as foundational material, structurally as valuable as semiconductors or energy.

News Corp has struck two major AI licensing agreements. The OpenAI deal, announced in May 2024, is worth more than $250 million over five years. The Meta deal, announced in March 2026, is worth up to $50 million annually.

The more AI-generated content floods the information environment, the scarcer – and therefore more valuable – certified human output becomes. **Authentication is the infrastructure that makes it both credible and tradeable.**

What Authentitas Does

Authentitas binds verified human identity to content at the point of creation, through **government-grade biometric verification, cryptographic provenance tracking, and immutable audit records.**

This makes the accountability chain of any critical information institution – who created this, who approved it, when and where, and under what institutional authority – provable, rather than merely asserted.

No institution can credibly authenticate its own content. Authentitas operates as independent, neutral infrastructure. The same principle that makes a hallmark valuable, a pharmaceutical seal trustworthy, and a notarised document legally binding.

We begin with journalism, where the need is most immediate. The same infrastructure scales across the sectors most in need of authentication.

In A Nutshell

These institutions are at the heart of how humanity makes its most crucial and consequential decisions.

All are vulnerable in the same ways, and for the same reason. Till now, they've been unable to prove that authenticity, accountability and auditability underpin what they produce.

Authentication changes everything, making the provenance of human knowledge once again demonstrable, tamper-proof and credible.

The critical information industries represent combined annual revenues of approximately $375 billion.

Considering the scale and velocity of the threats summarised here, alongside the clear limitations of the current range of solutions, the value that's at risk here is both obvious and substantial.

Authentitas exists to defend – and wherever possible, to grow – that value.

Key Sources

Institute for Strategic Dialogue, *Coordinated disinformation network uses AI, media impersonation to target German election*, February 2025.

Institute for Strategic Dialogue, *Stolen voices: Russia-aligned operation manipulates audio and images to impersonate experts,* May 2025. Documents the broader Operation Overload / Matryoshka network impersonating over 80 organisations across Q1 2025.

Westwood, S.J., 'The potential existential threat of large language models to online survey research', *Proceedings of the National Academy of Sciences* 122(47), November 2025. DOI: 10.1073/pnas.2518075122. The 99.8% pass rate is measured across the paper's attention check trials; the 43,000 figure covers the broader testing scope as reported by Dartmouth's press release accompanying publication.

Qualtrics, *2025 Market Research Trends Report,* published October 2024 (figures as reported by Research Live).

The Wiley/Hindawi retraction figures (11,300 papers, 19 journals) are sourced from a Wiley spokesperson statement to *The Register,* May 2024, and corroborated by Wiley's own white paper, *Tackling publication manipulation at scale: Hindawi's journey and lessons for academic publishing.*

Monperrus, M., Baudry, B. and Vidal, C., 'Project Rachel: Can an AI Become a Scholarly Author?', arXiv:2511.14819, November 2025 (updated December 2025).

Damien Charlotin, *AI Hallucination Cases Database.* Jurisdictional breadth and pace statistics drawn from Cronkite News (October 2025) and Volokh Conspiracy (April 2026) reporting on the database.

Omar, M., Sorin, V., Wieler, L.H. et al., 'Mapping the susceptibility of large language models to medical misinformation across clinical notes and social media: a cross-sectional benchmarking analysis', *The Lancet Digital Health,* 8, 100949, 9 February 2026. DOI: 10.1016/j.landig.2025.100949. Icahn School of Medicine at Mount Sinai.

Press Gazette, *News Corp CEO Robert Thomson warns AI companies scraping without paying: 'We're coming for you',* 4 March 2026. The News Corp deal with OpenAI (announced May 2024) was valued at more than $250 million over five years.

Sector valuations drawn from: *WAN-IFRA, World Press Trends Outlook 2024-2025,* January 2025 (news media); Market.us, *Global Academic Publishing Market Report,* January 2026 (academic publishing); Kentley Insights, *Marketing Research and Public Opinion Polling Market Size & Growth – 2026 Global Report* (market research); Thomson Reuters, *Annual Report 2024* (tr.com); RELX, *Annual Report 2024 (relx.com)* – together the dominant players in legal information services, with combined revenues exceeding $15 billion (legal information services); Simba Information / Freedonia Group, *Global Medical Publishing 2024-2028* (medical and clinical publishing).

authentitas.com

Authentitas