Tauth Labs
Trusted Authentication

Tauth Labs Content Authentication Issue Brief, February 2026

Framing Digital Safety in the AI Age: Protect, Detect & Correct

Simon Erskine Locke,
Co-Founder & CEO, Tauth Labs

for our lives and democracy. Shadow content – *which we define as manipulated, deceptive or fake content* – is a rodent gnawing away at the foundations of trust in the digital landscape.

Introduction

Deepfake videos are just one of the issues making the risk of the widespread adoption of generative AI personal. But this scratches the surface of the issues companies, individuals and governments face today in a world in which we can no longer tell what is or is not true or authentic.

We hate to be Chicken Little, but if nothing is done, the sky will fall. Without trust in digital content there is no economy or democracy.

The world's leading technology and media companies recognize the threats and are working to address them. One way this is happening is through the implementation of content authentication which builds provenance into digital content to prove authenticity.

This risk to reputations and bottom lines should provide motivation enough to take action. But, unless we frame these challenges in ways in which leaders feel they can do something, we will fail. This is why we have been making the case for focusing on the Tauth Labs-originated concept of **Shadow Content** as an addressable issue, rather than AI slop, or AI disclosure at a time when everything is touched by artificial intelligence.

There is a need to go one step further. We need to frame the solutions in terms of a system-wide response, so that we do not end up offering one-dimensional silver bullets that can be dodged "Matrix-like."

Protect, Detect and Correct

We need to look at digital safety in terms of three buckets: **Protect, Detect** and **Correct.**

Protect

Under the protect rubric is cybersecurity and content authentication. While cybersecurity is generally thought of in the context of protecting “owned” or company platforms, we need to look at it more broadly. Cybersecurity strategies need to be extended to protect content when it is outside these systems, when it is shared, and when it is on third-party platforms.

This is where content authentication and cybersecurity overlap. Content authentication provides the ability to differentiate between what is authentic, what is fake, and what has been manipulated by building in information about provenance.

The technology provides a means for social media platforms, content aggregators, trading platforms, and LLMs, to differentiate and prioritize authentic content.

Detect

Detection is complementary to protection. It is where most of the focus has been to date. Detection recognizes that in a world in which the tools to generate fake or shadow content and disseminate it at scale are in the hands of the good, the bad and the downright ugly, the war against fake content has been lost. Simply put, we are living in a world in which fraudulent and synthetic content are a feature of AI. No one is going to be turning off the tap or spigot any time soon. Even with widespread adoption of content authentication, shadow content will still be prevalent.

This means companies, individuals and governments need tools to both validate what is authentic and to identify what is not. And while much of the legislative focus on helping people know what has been generated using AI is deeply flawed, it is better than nothing if labels serve as signals to make the decision not to act on content or share it.

At a much more sophisticated level – detection software can be used by companies to identify fraud and its source. As an industry leader recently shared, detection comes with a “so what” challenge. Although knowing if content has been manipulated is an end in itself, an important question today is if it is detected, what can be done about it?

This is the driver of new automated solutions being developed by companies in the detect space, and growing attention in the communications industry to building playbooks to address AI-driven crises.

In a world in which trading and information travels at light speed, prevention is a foundational strategy for dealing with threats to digital safety, but since fraudulent and shadow content are not going away anytime soon, detection is a must.

Correct

The challenge of the correct bucket is huge. Taking down fake content in a timely matter is complex, time consuming, and comes with first amendment issues. As legislators work to put in place new laws requiring technology platforms to address issues around content provenance and AI, hides are thick and kickback is strong.

Corporate communicators and their agencies are doing the important work of preparing leaders for AI crises, building out response playbooks, and solutions. Counter-programming in the form of press statements and outreach are an important part of the tool kit, but like corrections posted in the media, fixes rarely receive the attention of the original article or post. Although an ounce of prevention is always better than a pound of cure, companies need to be ready with heavy guns to address these new risks.

Technology cures are being launched and developed by detection companies. Some of the most aggressive responses include launching a denial-of-service attack on a fraudster, doxing them, and other deep tech approaches that operate with the necessary speed to address issues rapidly in “tech time.”

While having a army of bots on your side may be necessary, if we play this out, we are on the path to a Terminator-like world in which technologies are battling each other for dominance.

Takeaways

The power of the idea of **Shadow Content** is that it focuses legislators, companies, and technologists on a solvable problem. The value of ***Protect, Detect*** and ***Correct*** is that this frames the solutions in a way that are multi-dimensional and consistent with the threat.

We are focused on content authentication, but are fully aware that it needs to be part of a comprehensive approach to digital trust and safety. So, when stones are thrown at the technology, the panes of the glass house do not shatter – because they are laminated.

Content authentication is tried and tested, fundamentally robust, and a major step forward in efforts to make the digital landscape safer. This foundational technology addresses the challenge of trust in the digital economy and can be customized for numerous use cases. It is hugely consequential.

There is a role for content authentication to play across the protect, detect and correct framing we have detailed here. Implementing the technology is not only a defense mechanism, but also a business opportunity, because it is about making content more valuable and more likely to be prioritized in the technology landscape.

It passes the key test of giving clients and consumers the ability to verify that content is authentic, and, as a result, they are more likely to act on it.

About the Author

Simon Erskine Locke is co-founder and CEO of Tauth Labs which is focused on implementing content authentication technology in communications, financial services, and governments. He has written white papers and articles on content authentication, spoken at events, and is an originator of the concept of Shadow Content.

About Tauth Labs

Tauth Labs helps companies incorporate content authentication into work flows to address the growing risk of misinformation and content fraud, as well as increase its value and searchability in the new: "Verify, then trust, digital world."

Tauth, short for trusted authentication, is at the forefront of building content provenance authentication applications based on foundational C2PA technology standards developed by the world's leading technology and media companies. We are implementing customized solutions for high-value communications, financial services, and local government content.

The company is a Certificate Authority (CA) for the issuance of C2PA content credentials. Tauth uses state-of-the-art standards of security, privacy, data-protection & electronic signature regulations based on C2PA and NIST standards; robust digital watermarks for provenance authentication; a tamper-evident distributed data store (ISO 22739:2024); developer friendly SDKs; and is easily integrable with CMS platforms.

Contact us at slocke@tauth.io or visit www.tauth.io