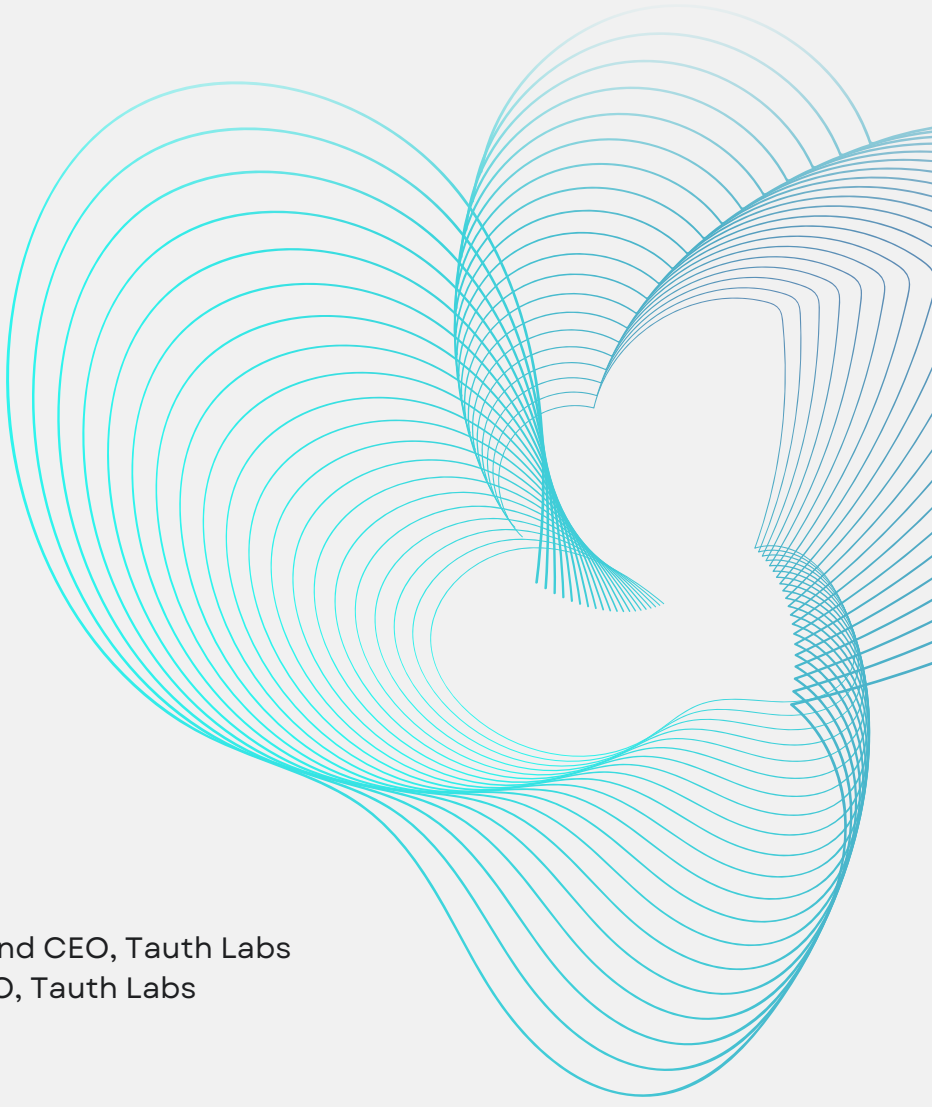
Tauth Labs

Trust, Content Authentication, and Financial Services

Protecting Companies from Reputational and Financial Risks, Clients from Fraud, and Making Communications More Trustworthy in the AI Age.



White Paper, March, 2026

Simon Erskine Locke, Co-Founder and CEO, Tauth Labs
M. Danish Bilal, Co-Founder and CTO, Tauth Labs

Introduction

Financial services firms are facing significant new challenges in the AI world.

At the same time banks, securities firms, asset managers, and insurance companies are looking at ways to deploy artificial intelligence to drive efficiencies and create new opportunities, AI is being deployed against them.

As Dimitri Zabelin, senior AI and cybersecurity research analyst at PitchBook states in a recent report:

> Attackers are increasingly using agentic AI systems, polymorphic malware, and AI-generated social engineering to automate reconnaissance, adapt intrusions in real time and persist after remediation. These techniques compress the available time to respond and undermine more traditional defense approaches.

As if securing networks was not a significant enough challenge, generative AI is also at the root of new threats to bottom lines, reputations, and clients from the menace of what we call “shadow content,” and an even bigger issue as a consequence - the erosion of trust in the content companies share with their audiences.

We define shadow content as misinformation, disinformation and fraudulent content issued around corporate announcements. The term can be applied more broadly to any content designed to look like it is from a company or individual. In this context, fake texts, emails, articles, and invoices are part of the continuum. It is turning up around quarterly earnings reports, in fake corporate announcements, and around M&A transactions.

It’s important to be clear from the outset that while for the most part in this paper we are focused on communications content (e.g. announcements, press releases, websites, articles, videos, and client communications), the challenges we outline apply to all forms of organizational communications.

With every cyberattack on a company or individual, every phishing event, and every time consumers receive, read or watch content that is fake or designed to deceive, a little bit of trust is lost.

Perceptions of trust are a relative issue. We may see trust in the system declining, but believe that trust in our organizations is separate and distinct. At the level of individual financial institutions, it may be tempting to look at trust as an issue others need to address, or because it is looked at in the context of the overall system, as a too-big-to-solve problem.

The response to risk reaches a tipping point when the cost of fraud, reputational damage, or client concerns outweigh the expense of addressing them. In this paper, we make the case that we are at such a moment.

Fraudulent content is driving a fundamental shift from “trust, but verify” to “verify, then trust.” This matters. If audiences question or no longer trust that content is authentic, they will ignore it and not act on it. Since driving behavior is the purpose of the communications and marketing industries, and critical to the functioning of the digital economy, this is an existential threat.

With leading technology consulting firms sounding the alarm around the reputational risk of AI and fraud, the industry needs to focus on addressing these risks.

Importantly, as we outline in this paper, there are solutions. Content authentication, which addresses both trust in the system and the risk of shadow content, offers one path to rebuild trust into content.

Rebuilding trust into content and protecting companies and clients from shadow content requires a multidisciplinary approach to minimize the impact of these new risks. And, while we primarily focus here on external and client communications applications, as a senior professional at one of the large banks shared, ultimately the technology will need to be built into all digital documents.

Trust in Digital Content is Declining

There has been a well-documented steady decline of trust in digital content. This is accelerating in parallel with the rise of generative AI.

Today, we take for granted the trust that has been built into the digital economy. For boomers and millennials, the fundamental trust in electronic transactions had to be earned. Trust in the internet as a source of news and information, in social media, and in the technology we use has developed over time.

This trust is at risk. In the AI era, the ability to know what is true and can be trusted is being questioned and challenged across the board.

Trust in content is multidimensional. One dimension is the message, another the brand, a further dimension is the reputation of the message carrier, and the fourth is the technology medium through which it is carried. All are important.

Content authentication builds trust into content at the level of digital files. In the same way the adoption of HTTPS integrated security certificates into websites, content authentication builds trust into each piece of content. It provides the ability to determine provenance, whether or not content has been manipulated, and incorporate valuable information related to its origin. Authentication is a fundamental shift in the technology landscape.

Shadow Content

The Journal of Accounting and Economics warned in December 2024 of fake financial content being issued around quarterly earnings announcements. This type of content is now an emerging issue around mergers and acquisitions.

Although it is not a new phenomenon, the production of shadow content is being accelerated by the adoption of generative artificial intelligence and agentic tools designed to automate its dissemination at scale.

> Shadow content is designed to mislead audiences into believing it is authentic, propagate misinformation and disinformation, and create opportunities for financial fraud and potentially move markets.

When fake announcements appear authentic and are acted upon, the financial consequences can be severe. The stock price of the company may be at risk, corporate investors deceived, clients defrauded, and reputations damaged.

The growth of AI-driven fraud against corporations is a growing concern. And, with more than 70% of Americans having been subject to cyberattacks – much of which is based on content that looks authentic - fake content in all its dimensions is one of the root causes of the erosion of trust.

With the widespread availability of generative AI tools, we are fighting a losing war to prevent shadow and fake content. The focus now needs to be on helping audiences know what is authentic, and ensuring that authentic content is prioritized on the internet.

For the financial services industry, this is a particularly consequential issue. Phishing and vishing are using what look like communications from financial institutions to secure information, execute transactions, and dupe clients into paying fake invoices.

The cost of digital fraud in 2025 in the U.S. alone was estimated to be in excess of $1 trillion. Every major financial institution is facing this threat.

Content Provenance Authentication Technology

There are a number of new technologies and approaches to building trust into content for this new AI era. Understanding the differences between the various approaches that are being taken is essential to recognizing the value of the Coalition for Content Provenance and Authenticity (C2PA) content authentication standards developed and championed by the world's leading technology and media companies.

Content provenance authentication provides a way for companies, organizations and governments to build credentials into digital documents, images, audio and video to confirm their authenticity and provenance. The technology is complex, but its use, when integrated into CMS systems or through client-specific applications can be as simple as one click for authentication and verification.

Content authentication builds a form of digital identity certificate into individual pieces of content that can be validated by recipients, search engines, LLMs and content aggregators. In our first white paper on this topic, we provided a very detailed technology overview, which we will skip here.

Content authentication using C2PA standards does not prevent the creation of fraudulent content, but it is the basis for a set of new tools that are able to identify what is authentic and what is not and, critically, to rebuild trust.

Authentication is a three-step process that includes user or organization verification, the incorporation of a digital watermark and credentials into content, and as an optional third step, using blockchain to create a unique identifier for each piece of content.

C2PA authentication works in the same way as the Federal Reserve uses multiple levels of security to protect against counterfeit bank notes. If a bank note is copied, it may look the same, but these layers of security and watermarks mean that like a $10 bill, authentic notes are clearly identifiable from fakes.

C2PA provides the most robust and interoperable set of standards to build the who and what into digital files ranging from documents to videos. The underlying open-source Linux-based technology standards enable the customization of credentials for individual companies and types of content.

Content provenance authentication is distinct from a range of other standalone technologies that are in the conversation. These include: user verification, digital watermarks, AI detection, and misinformation detection tools.

User Verification

A starting point for content authentication, user identity verification enables platforms and users to know that people or companies are who they say they are when they post information or create accounts. It does not build into content a way to confirm the authenticity of content when it is shared.

Digital Watermarks

Digital watermarking is a technology that has been available for more than a decade. When embedded into documents or images the metadata can be read to validate its origin. While digital watermarking has been widely adopted for images, this adoption has been relatively limited to niche applications. The integration of watermarks into C2PA's comprehensive authentication process, which includes the integration of digital credentials and the potential integration of blockchain to prove authenticity, significantly increases the robustness of the digital provenance built into content.

AI Detection

AI detection tools help users know if content was created or manipulated using AI. This becomes important when documents, images or video are being passed off as real. For the media, the ability to determine if an image has been manipulated is the end goal.

When applied more broadly, with AI becoming omnipresent, we need to ask at what point and level does the use of AI matter?

It is worth noting here that C2PA content authentication can be used to incorporate into a content manifest that it was in part or in whole generated using AI - consistent with disclosure requirements.

Misinformation Detection Tools

Misinformation detection tools provide ways for companies to monitor the internet to identify fake or imposter content that has been posted, and track its origin. These tools can be thought of as an evolution of media and social media monitoring. They provide organizations with a way to quickly identify issues and work to address them.

As these tools evolve, the capability to automate the process of addressing the problems they detect is being developed. Detection is also an important starting point for communications playbooks, to address crises and the repercussions of fake content.

These tools are complementary to content authentication. Each is different and valuable in its own way. The framing idea of “Protect, Detect and Correct” provides a way to look at each of these technologies as part of a systemic approach to content security.

Customization of Credentials and Content Manifests

A Tauth Labs innovation, built on the flexibility of the open source C2PA platform, provides financial institutions with the ability to implement customized authentication processes using their brand and benefiting from the trust they have established.

Since different use cases will benefit from various types of information being embedded into content, we provide the ability to customize information in the digital manifest embedded into a file.

Built into C2PA authentication are ways to incorporate credentials at the time of publishing or log changes made prior to publication. A press release would under most circumstances be authenticated at the time it is published, while a video may benefit from a manifest which includes edits that have been made during production.

Incorporating geolocation data into manifests may be a priority for some types of content or sources of data a requirement for others. Other client priorities can also be addressed that significantly add to the value proposition. The C2PA standards underlying authentication provide this flexibility.

Financial Services Use Cases

Content authentication can be embedded into all forms of digital content. It can be built into CMS tools and systems as a final step in the publishing process for documents, images, audio or video files. Companies can also use standalone applications to authenticate content outside their technology environment before it is distributed.

Outlined below are a number of use cases:

1 Corporate and Financial Communications

Authenticating press releases, financial communications and required disclosures provides a way to ensure that journalists, analysts, clients, and other relevant audiences know that the documents they receive are authentic.

2 Client Communications

Authentication provides ways for clients to determine which materials are from a specific financial institution through built-in “trust shields” in digital communications or by using browser plugins to automate this process. In the future, authentication built into texts and email systems will provide additional ways to incorporate trust into content across these communications channels.

3 Create Secure Communications Channels with Journalists, Analysts or Aggregators

One of the most significant risks of shadow content is that it is shared by journalists, analysts, and other audiences. Once fake content is shared by a credible source it benefits from the authority this conveys. Content authentication provides these key audiences with the ability to validate the authenticity of press releases, reports or articles, so they share only what is authentic.

4 Authenticating Images, Charts or Graphs

Content authentication can be used to validate the ownership and authenticity of proprietary images, charts or graphs created within an organization.

5 Detection Tools for Authenticated Content

Authenticated content that has been logged on blockchain can be the basis for tools to detect both its use and potential misuse.

6 Research Reports

The potential for the manipulation of research and analyst reports is significant along with the consequences for clients and markets. Authenticating research, white papers, and other reports as with all other forms of content not only builds trust into these critical documents, but will, over time, make them more searchable and valuable in the digital landscape.

7 Authentication of Corporate Podcasts and Videos

Tauth content authentication can be built into audio and video files, to provide a basis for determining their authenticity.

8 Invoice Authentication and Verification

Fraudulent invoices are being generated by criminals and sent to clients on an ongoing basis using increasingly sophisticated methods. The use of authentication for invoices and payments provides an additional level of payment security and reduces the risk of clients being defrauded. The WSJ highlighted one dimension of this issue in an article that showed how fake invoices were used to dupe a BlackRock unit into a $400 million loan.

9 Creating Provable Ownership of Content

Content authentication provides a way to prove ownership of digital assets, so that their use by others can be identified in relation to patents or monetization.

10 Embed Do Not Train Instructions for LLMs

If your organization is concerned about whether content is being used to train LLMs, content manifests can be customized to incorporate do not train notifications.

New Technology, New Questions

For companies and professionals new to content authentication, there are likely to be many questions about the technology. The following Q&A addresses common questions. We are available to address questions that are not covered here.

What is the C2PA standard?

The C2PA standards were developed by a coalition of the world's leading technology and media companies. Tauth Labs is a contributing member of C2PA (www.c2pa.org).

How do you validate the identity of a person or company in a way that ensures the authenticity of content provenance?

There are a variety of ways we can do this including the use of existing identity verification tools, using a corporate domain to verify identity, or social proof.

Can bad actors use C2PA to authenticate fake content?

Tauth has completed C2PA's rigorous conformance process to become a Certificate Authority with the ability to issue content credentials. The barriers to issuing C2PA credentials are high, and the layers of technology incorporated into the process are fraud resistant. Blockchain logging of content creates an immutable connection between a document and the manifest.

How do companies benefit from having the ability to view their manifest of authenticated documents on blockchain?

Blockchain provides an added layer of security to the content. If someone wants to keep an immutable record on the blockchain, they can do it.

If I send a press release to a journalist, how will they know if the release is authentic?

A journalist will be able to review the authentication credentials in a document shared with them to confirm if it is authentic. A browser extension can also be provided to automate the process of determining if a document has been authenticated and validate the content credentials. We can also provide verification tools for specific audiences based on a client's goals, as well as general audiences.

Can authentication be used to track where documents are sent and who is viewing them?

Yes. Tracking of authenticated content is a potential feature.

If companies already validate the identity of people issuing press releases or other content, why do I need to build additional content credentials into my content distribution process?

The process of issuing a press release through major platforms has a high bar for the individual authentication of the individual or company issuing the release. But it is not foolproof. When an article travels via email, texts, WhatsApp, and other channels/platforms, it becomes hard to verify its authenticity. Content authentication provides the ability to confirm the authenticity of documents, wherever they are sent. It is an additional layer of digital security.

If I post content to social media, do content credentials remain with it?

Content credentials include robust digital watermarks. Even if metadata is deleted, the authentication process provides a mechanism for audiences to determine whether a document is in fact authentic through our detection tools. With major technology companies and social media platforms part of C2PA , and regulatory standards pushing the industry toward the requirement that provenance data not be stripped from digital files, we believe we are moving toward a world in which embedded authentication will become the rule, not the exception.

Can a user see the digital watermark?

When content is shared invisible authentication credentials and a robust invisible watermark are included. Clicking on the “Trust Shield” credential in the document provides the ability to valid ate that a document is authentic. Using verification tools that identify the digital watermark or link to the blockchain content manifest provides information about the provenance of the document.

Can a watermark be faked?

Digital watermarks offer a high-level of security, but as AI becomes increasingly sophisticated, a watermark may be replicated. A blockchain hash to the original document, however, cannot be recreated; so, the document will not register as authentic.

Why should I trust another company with sensitive documents?

Content authentication does not mean the authenticator has to see the content. So, everything can be hosted within or outside a company’s secure technology environment. Authentication would, in this case, only take place at the time content is published.

Would including a link that brings people back to my website be the safest way to send content?

The goal of creating content is for it to be shared. Once it is shared, it is no longer in the custody of the company email or website. This is when there is the greatest risk for clients or consumers.

Website fraud is increasingly common as AI helps companies set up in minutes websites that mirror those of actual companies. URLs can look like they are authentic but also be faked. While bringing users back to websites offers a high degree of security, it does not address the fact that malicious actors can replicate this process for fake content.

Content credentials significantly raise the level of digital security by providing users with a secure mechanism to validate documents that are authentic.

How can I verify if a document with content credentials has been manipulated and used fraudulently?

Content authentication provides a way to track changes in a document and confirm if it has been manipulated. There are a number of ways to show this including changing colors to red or orange, depending on the severity of the fraud.

Do users need to download any software onto their computers to recognize content credentials?

No. The authentication process provides a way for recipients to see that content has been authenticated and verify this without having to download software. If clients want this process to be automated for specific users, we can provide a browser plug-in to do this.

How do you manage privacy issues?

C2PA guidelines are designed to ensure that content provenance authentication is an option for companies and individuals for privacy reasons. This is why we are focused on specific use cases where it makes sense for creators and users to have content provenance information included in a document.

Can I integrate GEO location data or other information into the content manifest?

Yes. Content manifests can be customized to address specific client use cases.

Conclusion

Content provenance authentication represents a fundamental shift in the technology landscape. In a world where a growing portion of content is AI generated and access to tools that create it and automate its distribution are in the hands of the good, the bad and the ugly, content authentication is key to helping companies, their clients and consumers differentiate between what is authentic and what is not.

Content authentication is not a technology that needs mass adoption to have value. In fact, authentication is particularly valuable for situations in which high-value content must be sent to specific individuals or groups. Recipients do not need to download software or tools to verify authenticity. The ability to do so is embedded in each piece of content that is shared. Simply mousing over or clicking on a logo or trust shield provides the receiver with the ability to know what can and cannot be trusted.

For the financial services industry there are a range of applications. By authenticating press releases, financial communications and disclosures, companies can build trust into these communications, making it more likely that they will be acted on. Authentication also provides a basis for the media, content aggregators, search engines and LLMs to prioritize content in ways that increase its value and reduce the risk that fraudulent content will be shared.

The world's leading technology and media companies are behind the C2PA technology standard. Within three to five years, we expect that most high-value content will be authenticated as industry leaders adopt the technology, clients expect it from companies and their vendors, and regulators require provenance be included. The financial services industry will be at the vanguard of the adoption of content authentication given the importance of trust to the industry.

Tauth Labs is at the forefront of developing customized implementations of the technology, as well as advising companies and technology teams on the adoption of C2PA standards.

About the Authors

Simon Erskine Locke

Simon Erskine Locke is Co-Founder and CEO of Tauth Labs. He is a recognized thought leader on content authentication having written numerous articles on the topic and spoken at industry events. He originated the concepts of “Shadow Content” and “Protect, Detect and Correct” as framing tools for a system-based view of digital and content safety.

He is an entrepreneur and writer on communications and marketing issues. He founded CommunicationsMatch™ and Townhub. He is a former head of corporate communications functions at Prudential Financial, Morgan Stanley and Deutsche Bank. He is a member of the board of the Foreign Press Association of the United States.

M. Danish Bilal

M. Danish Bilal is Co-Founder and CTO of Tauth Labs, where he leads the development of its next-generation digital content authentication platform built on blockchain and cryptographic trust standards such as C2PA.

With a background in AI infrastructure, blockchain security, and decentralized identity, Danish has worked extensively on verifiable computation pipelines, PKI frameworks, and smart contract security. He has coauthored research on blockchain-based PKI and Ethereum storage vulnerabilities, and has led deep tech initiatives spanning zero-knowledge proof systems, verifiable credentials, and AI agent orchestration for Web3.

At Tauth Labs he oversees the design of its validation and authentication modules, enabling users and publishers to embed and verify tamper-evident manifests in digital content and helping establish trust, provenance, and safety across the internet.

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 the value and searchability of content in the new: *"verify, then trust, digital world."*

Tauth, short for trusted authentication, is at the forefront of building content provenance authentication applications that are 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, and electronic signature regulations based on C2PA and NIST standards. Robust digital watermark for provenance authentication. Tamper-evident distributed data store (ISO 22739:2024). Developer friendly SDKs and easily integrable with CMS platforms.

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