

# Maya Protocol

## Authenticity Layer of the Internet

Maya Labs

June 1, 2024

### Abstract

The rapid evolution of AI-generated digital media technologies has made it increasingly difficult to verify the authenticity of media content. Existing solutions for content authentication, such as AI detection and cryptographic attestations, are no longer adequate. AI detection often produces unreliable results and struggles to keep pace with generative AI advancements. Cryptographic attestations, on the other hand, place too much trust in third party entities, which can compromise the integrity of the verification process.

Maya Labs is building the first Authenticity Layer, leveraging zero-knowledge technology to provide seamless, secure verification ultimately eliminating the need for third-party intermediaries. By leveraging advanced cryptographic techniques, the protocol ensures that verification proofs are bound to digital content, enabling anyone to verify authenticity while preserving privacy. The use of blockchains creates an immutable audit trail that is publicly verifiable, fostering a trustless and open system for media verification.

Combining these cryptographic techniques with economic mechanisms, the Maya Protocol provides a robust and scalable authenticity layer for all digital content across the internet. This innovative approach aims to protect against the growing threat of misinformation and AI-generated deepfakes enhancing global trust and transparency.

## 1 Introduction

The inception of the internet revolutionized how information was cataloged and accessed, with Hypertext Transfer Protocol (HTTP), spearheaded by Tim Berners-Lee, serving as the backbone for data exchange on the World Wide Web. This transformed the internet into a repository of knowledge accessible to the masses. However, as the internet landscape evolved from an academic haven to a vast digital expanse inhabited by a diverse array of users including advertisers, anonymous entities, and malicious actors, the pressing need for a security layer became apparent. To address this, Netscape Communications introduced Secure Sockets Layer (SSL), which was later succeeded by Transport Layer Security (TLS). These technologies enhanced the web's integrity by encrypting communications and transitioning standard HTTP websites to more secure HTTPS counterparts.

The rise of camera phones and social media platforms has sparked an unparalleled explosion in digital media production. This is evidenced by staggering statistics: 4.5 billion images are created daily, and 250 hours of video content are uploaded every minute. This deluge of digital content has not only become a cornerstone of modern entertainment and communication but has also blurred the lines between reality and artificiality in media. Projections indicate that by 2026, AI-generated content will constitute 90% of all online material, making it increasingly difficult to distinguish between what is real and what is fabricated. This reality underscores the urgent need for a solution that can assure the authenticity of digital media, similar to how HTTPS has become synonymous with providing the authenticity of messages exchanged between the server and client.

We present the Maya Protocol, a decentralized protocol designed to redefine the authenticity of digital content across the internet. Inspired by the transformative role of HTTPS in enhancing web security, the Maya Protocol aims to become the cornerstone for ensuring the authenticity of digital media. The protocol seeks to make the authenticity of every digital media piece transparent and verifiable, addressing the challenges posed by the growing prevalence of AI-generated content in our digital ecosystem.

## 2 Problem

The digital age has brought about significant advancements in media creation and dissemination. However, with these advancements come critical challenges, particularly in verifying the authenticity of digital content. The current landscape is plagued by several key issues:

- **Proliferation of Fake Content:** The rise of AI-generated media technologies has made it easier than ever to create highly realistic fake images, videos, and audio. This has led to an increase in the spread of misinformation, which can cause societal unrest and political polarization. Additionally, businesses are losing billions of dollars due to manipulated media, which is eroding trust and damaging brand reputation.
- **Inadequate Verification Solutions:** Existing solutions for verifying media authenticity, such as AI detection and cryptographic attestations, are insufficient. AI detection methods often produce false positives and struggle to keep up with the rapid advancements in generative AI. Cryptographic attestations rely too heavily on trusting the signer, which can be compromised, leading to questions about the integrity of the verification process.
- **Lack of Content Privacy:** Many current verification methods do not adequately preserve the privacy of the original media or the intermediate edits. This is crucial, as raw images or videos often contain sensitive information that needs to be protected while still allowing for authenticity verification.
- **Need for Decentralized Trust:** There is a growing need for a decentralized approach to media verification that does not rely on centralized authorities. Decentralized solutions can provide a more robust and tamper-proof method for ensuring the authenticity of digital content.

These challenges underscore the urgent need for a new, innovative solution to verify the authenticity of digital media in a reliable, privacy-preserving, and decentralized manner. The Maya Protocol aims to address these issues by leveraging advanced cryptographic techniques and blockchain technology.

## 3 Current Solutions

In the evolving landscape of digital content verification, two primary approaches have emerged: AI detection and cryptographic provenance-based methods. Both have their merits but also significant limitations that hinder their effectiveness.

### 3.1 AI Detection

AI detection methods utilize artificial intelligence algorithms to identify manipulated or AI-generated content. These detection algorithms analyze various aspects of digital media to identify inconsistencies or signs of manipulation. Common techniques include:

- **Image Analysis:** Algorithms examine pixel patterns, compression artifacts, and other characteristics that may indicate tampering. By comparing these patterns to those of genuine content, the algorithms can flag potential manipulations.
- **Metadata Examination:** AI systems analyze metadata associated with digital files, such as timestamps, device information, and editing history, to detect anomalies that suggest content has been altered.
- **Contextual Analysis:** Advanced AI models can understand the context of the content, such as facial expressions in images or coherence in video scenes, to identify discrepancies that might indicate manipulation.

While these systems have made strides in recognizing fake media, they are inherently flawed due to unreliable outpaced detection.

- **Detection is Unreliable:** AI detection systems are unreliable and often produce false positives, incorrectly flagging legitimate content as manipulated. This undermines trust in the detection process and can lead to significant credibility issues for content creators and platforms.

- **Generative AI Outpaces Detection:** The rapid advancement of generative AI technologies has outpaced the development of detection methods. As generative AI becomes more sophisticated, creating increasingly realistic fake content, detection systems struggle to keep up, rendering them less effective.

## 3.2 Cryptographic Attestations

Cryptographic attestation methods involve digital signatures or certificates to verify the authenticity of content. These methods utilize cryptographic techniques to bind an attestation, such as a digital signature, to a piece of content. This process typically involves the following steps:

- **Digital Signatures:** A digital signature is created using a private key, which is unique to the signer. This signature can be attached to a piece of content, providing a way to verify that the content has not been altered since it was signed.
- **Certificates:** Certificates issued by a trusted authority can attest to the identity of the content creator and the integrity of the content. These certificates are often used in conjunction with digital signatures to provide additional layers of verification.
- **Public Key Infrastructure (PKI):** PKI involves a hierarchy of trust, where multiple layers of certificates and digital signatures are used to establish a chain of trust from a root authority to the end user.

While cryptographic techniques provide a foundational level of security, they also come with notable drawbacks.

- **Reliance on Signer:** Cryptographic attestations require significant trust in the signer or issuer of the certificate. If the entity providing the attestation is compromised or acts maliciously, the integrity of the attestation is nullified. This centralized trust model does not adequately ensure the authenticity and integrity of digital content in a decentralized ecosystem.
- **Single Point of Failure:** The centralized nature of trust in cryptographic attestations poses a significant risk. If the trusted entity is compromised, all content verified by them becomes questionable. This creates a single point of failure, which can be exploited by malicious actors. Additionally, if signing keys are leaked, they can be used to sign random content from the internet, which will still be verifiable. This can cause significant harm to reputation and further question the integrity of the verification process.
- **Lack of Decentralization:** Cryptographic attestations often rely on a few central authorities to issue and manage certificates of authenticity. This lack of decentralization can lead to vulnerabilities and potential misuse of power by these authorities, undermining the overall trust in the system. These central authorities can restrict access to verification or even charge for verification, further limiting the openness and accessibility of the verification process.

These limitations highlight the need for more robust solutions to ensure the authenticity of digital content. The Maya Protocol aims to address these issues by leveraging advanced cryptographic techniques and blockchain technology.

## 4 Maya Protocol

Maya Labs is building the first Authenticity Layer, leveraging zero-knowledge technology to provide seamless, secure verification ultimately eliminating the need for third-party intermediaries. The Maya Protocol is a decentralized protocol to verify and ensure the authenticity of digital media content such as images, audios and videos. Our solution enables users to achieve trusted and verifiable authenticity for their high-value digital media content by cryptographically proving its integrity. This allows anyone to verify the content's authenticity while ensuring that no one can alter it.

### 4.1 Features

The Maya protocol offers a robust framework for securing and verifying data. Its features include authenticity, complete privacy, trustless verification, public verifiability, transparency, immutable provenance records, credible neutrality, and composability.

#### **4.1.1 Authenticity**

The Maya Protocol guarantees the authenticity of digital media by ensuring that every piece of content is immediately signed at the moment of capture, and any edits made to the original media are properly verified and recorded. It uses advanced cryptographic techniques like zero knowledge technology to certify the origin and integrity of media files, allowing users to distinguish genuine content from manipulated or counterfeit versions.

#### **4.1.2 Privacy**

Privacy is paramount in the Maya Protocol, allowing users to maintain the confidentiality of their original media while still providing proof of its authenticity. This is crucial because raw images or videos are often edited before publication, sometimes to enhance their presentation or decrease file size, and at other times to omit sensitive details from the original content.

#### **4.1.3 Trustless Verification**

Trustless verification within the Maya Protocol eliminates the need for centralized parties to confirm the authenticity of digital media. The decentralized nature of the protocol means that the verification process is carried out on-chain using smart contracts, fostering a secure environment where trust is built into the system itself.

#### **4.1.4 Public Verifiability**

The Maya Protocol offers public verifiability, allowing anyone to independently verify the authenticity and integrity of digital media. This open verification process contributes to a transparent ecosystem where content can be trusted and verified by the wider community without compromising security.

#### **4.1.5 Transparency**

Transparency in the Maya Protocol is achieved through a transparent and open ledger of proofs of content authenticity. This feature ensures that the interactions within the protocol are visible to all participants, promoting accountability and trust in the system.

#### **4.1.6 Immutable Provenance**

Immutable provenance records are a key feature of the Maya Protocol, providing a history of digital media from creation to its current state that cannot be altered. While these records are immutable, meaning they cannot be changed once created, this does not equate to permanence; rather, it ensures that the integrity and origin of content are preserved over time.

#### **4.1.7 Credible Neutrality**

The Maya Protocol provides Credible neutrality by de-risking the system from centralized powers and ensures a widely accessible public good. The decentralized and crypto-economic nature of the protocol guarantees that all content is treated equally, fostering a balanced and censorship-resistant digital media environment.

#### **4.1.8 Composability**

Composability in the Maya Protocol refers to the ability to integrate and combine different features and services seamlessly. Composability is provided through integrations with multiple apps and services such as social networks and news platforms. This flexibility allows for the creation of an ecosystem of complex applications and services built on top of the protocol, enabling innovation and the development of new paradigms within the digital media landscape.

## **5 Technology**

### **5.1 Zero-Knowledge (ZK)**

Zero-Knowledge (ZK) technology is at the core of the Maya Protocol, providing unparalleled integrity for digital media.

- **Verifiable Integrity:** ZK provides verifiable integrity, also known as glass-to-glass integrity, ensuring that the authenticity of digital content is maintained from the moment of capture to its display on a visual display.
- **Internet Scale Media Verification:** With significant advancements in ZK technology, it now provides the necessary infrastructure that can support verifiability of all kinds of media. It will secure all digital media interactions by generating cryptographic proofs that verify the integrity and authenticity of content without revealing any additional information.

**Zero-Knowledge technology** is the endgame for media authenticity because it offers a robust, scalable, and privacy-preserving solution to verify the provenance and integrity of digital media. Additionally, ZK also eliminates the risk of private key leaks associated with cryptographic attestations and ensures trustless verification, enhancing security and integrity.

## 5.2 Blockchain

Blockchain technology underpins the **Maya Chain**, offering a transparent and immutable audit trail for digital media.

- **Immutable Audit Trail:** The Maya Chain functions as a transparent and immutable audit trail, recording all provenance and authenticity information about digital media. This ensures that the history of any piece of content can be traced and verified publicly, enhancing trust and accountability.
- **Provenance:** By maintaining detailed records of provenance and authenticity proofs, the blockchain ensures that any alterations or edits to digital media are accurately documented and verifiable. This capability is crucial for maintaining the integrity of content in a decentralized environment.
- **Decentralizing Trust:** Blockchain technology is essential for decentralizing trust and reputation, moving away from reliance on centralized institutions to provide authenticity. By distributing trust across a decentralized network, the Maya Protocol enhances the security and reliability of digital media verification, making it resilient to tampering and fraud.

The **Maya Chain** serves as a coordination layer, providing a single source of truth for the authenticity of all digital media content on the internet. It enforces a common standard for media authenticity, such as which proof system to deploy for verifying digital content, ensuring consistency across various platforms.

Additionally, the **Maya Chain** provides composability with multiple apps and services that want to use Maya for media authenticity, such as reputation systems and decentralized AI, enabling seamless integration. The use of a custom blockchain solution allows for easy upgrades, such as adding new cryptographic proof systems, and can be optimized for scalability to meet the evolving needs of digital media verification.

# 6 Mechanism Design

The Maya Protocol integrates reputation-based mechanisms alongside cryptographic verifications to ensure media authenticity and integrity.

## 6.1 Reputation System

The protocol includes a reputation-based system where users are rewarded or penalized based on validation results. This system operates through social consensus games, where participants collectively vote on the authenticity of flagged media content. Social verification is particularly crucial when source content passes all cryptographic verification checks but may still be questionable, such as a recording of a screen displaying generated video or photo.

Creators are rewarded for authenticating genuine content, which boosts their on-chain reputation score. Conversely, if tampered content is flagged, the creator faces penalties and their reputation score decreases.

Each creator authenticating content with the protocol has an on-chain reputation score cryptographically linked to a provable identity, such as a Twitter or Instagram ID. This linkage ensures accountability and traceability, enhancing the overall integrity of the system.

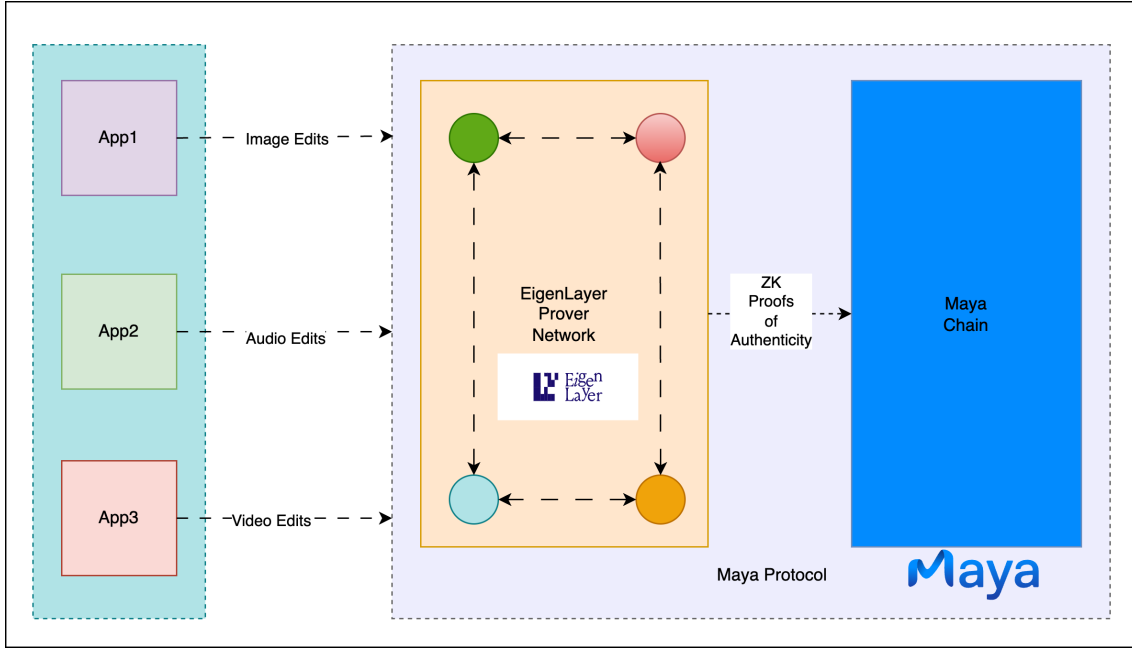


Figure 1: High-level architecture diagram of the Maya Protocol.

## 6.2 Social Consensus Mechanisms

Social consensus mechanisms leverage the collective intelligence of the community for media verification:

- **Voting Systems:** Participants vote on the authenticity of flagged media content, reflecting the consensus of a diverse group of users rather than a few centralized entities.
- **Reputation-Weighted Voting:** Votes are weighted based on reputation scores, giving more influence to trusted validators with higher reputations.
- **Dispute Resolution:** Secondary validation processes address disagreements or disputes, ensuring fair and transparent resolutions.

By combining cryptographic verifications with reputation-based mechanisms, the Maya Protocol creates a robust system for maintaining the authenticity and integrity of digital media, discouraging malicious actors and fostering a trustworthy digital ecosystem.

## 7 Use Cases

The Maya Protocol has a wide range of applications wherever the integrity and authenticity of media are paramount. These applications span across various sectors, including legal and financial institutions, professional content creators, AI platforms, media companies, and news platforms. In this section, we discuss three key use cases.

### 7.1 AI Transparency

Generative AI platforms must comply with regulations and maintain transparency in their training data to build credibility and trust.

- **Impact:** Compliance with regulations, ethical and legal sourcing of training data, traceability of AI-generated content, and maintaining credibility are critical.
- **Maya's Benefits:**
  - Ensures that training data is ethically and legally sourced, providing verifiable proof of its origins and authenticity.
  - Offers verifiable proof of origins and authenticity for AI-generated content (AIGC), ensuring that models are not used to create malicious deepfakes.

## 7.2 Professional Content Creators

For professional content creators, especially social media influencers, maintaining the authenticity and integrity of their content is crucial.

- **Impact:** The reputation of their personal brand and the trust of their significant communities are at stake.
- **Maya's Benefits:**
  - Protects creator brand IP and reputation by ensuring all content is verified and authentic.
  - Enhances audience trust and integrity, leading to stronger community engagement.
  - Provides legal safeguards for content disputes by maintaining a verifiable history of content authenticity.
  - Ensures cross-platform authenticity, allowing creators to maintain their integrity across different social media platforms.

## 7.3 Legal

In the legal sector, the authenticity of digital evidence is essential to prevent false accusations and ensure the integrity of testimonies and the chain of custody.

- **Impact:** The integrity of digital evidence, the ability to prevent plausible deniability, and the integrity of testimonies and the chain of custody are at stake.
- **Maya's Benefits:**
  - Provides tamper-proof certificate of authenticity for digital evidence, ensuring its integrity from capture to courtroom presentation.
  - Maintains a verifiable chain of custody for all digital evidence, preventing any alterations or tampering.
  - Ensures the authenticity of testimonies, safeguarding the legal process from false claims and accusations.

## 8 Conclusion

The Maya Protocol represents a transformative approach to ensuring the authenticity and integrity of digital media in an increasingly AI-generated digital world. By leveraging advanced cryptographic techniques, particularly Zero-Knowledge (ZK) proofs, and the transparency of blockchain technology, the protocol builds a future where every piece of digital media content is verifiable, providing users with sufficient contextual information to make informed decisions.

The features of the Maya Protocol, including credible neutrality and trustless verification, create a robust framework for verifying digital content. The Maya Protocol addresses the critical shortcomings of existing solutions such as AI detection and cryptographic attestations, which often produce unreliable results and depend on centralized authorities. By using advanced cryptographic techniques like ZK proofs and leveraging blockchain technology, the Maya Protocol ensures the integrity and authenticity of digital content without relying on third parties. These features cater to various applications, from protecting the reputations of professional content creators to ensuring compliance and transparency for AI inference platforms, and maintaining the integrity of digital evidence in legal contexts.

As generative AI technologies continue to evolve, creating hyper-realistic photos and videos that are indistinguishable from reality will become increasingly inexpensive and accessible. This makes it nearly impossible for audiences to discern genuine content from manipulated media. The Maya Protocol stands at the forefront of this movement for content integrity, providing a scalable and secure solution that empowers users to trust the digital content they encounter. It is designed to be deeply integrated into the very fabric of the digital ecosystem, ensuring that every story told, every picture shared, and every voice heard is genuine.

## 9 Additional Resources

1. Dan Boneh's Talk: <https://www.youtube.com/watch?v=fF9VrtiwQ00>
2. CLI Tool: <https://github.com/0xmayalabs/maya-cli>
3. Benchmarks: <https://docs.mayalabs.tech/>
4. Blog: <https://blog.mayalabs.tech>
5. ZK-IMG: <https://arxiv.org/abs/2211.04775>
6. Article: <https://medium.com/@boneh/using-zk-proofs-to-fight-disinformation-17e7d57fe52f>
7. C2PA: <https://c2pa.org/>
8. Maya ZK Editor: <https://app.mayalabs.tech>