RESTORING TRUST IN THE AGE OF DEEPFAKES: A BLOCKCHAIN-BASED PROPOSAL

THE GLOBAL CHALLENGE OF AI-MANIPULATED MEDIA

Alexander Pfeiffer, University for Continuing Education:
ORCID: 0000-0002-8689-3318

Nanditha Krishna, Independent Researcher:
ORCID: 0000-0001-5536-4993

The proliferation of AI-generated deepfake media represents a critical global challenge, undermining public trust across political, financial, and social domains. Recent advances in generative AI, particularly Google's VEO3 model, have dramatically escalated the sophistication and accessibility of synthetic media creation, enabling the generation of photorealistic videos with synchronised audio that are nearly indistinguishable from authentic content. Existing approaches, primarily focused on deepfake detection, fall short due to the evolving sophistication of artificial intelligence methods and the emergence of what researchers term the "liar's dividend"—the ability for bad actors to dismiss authentic content as potentially fake. This opinion paper introduces a technically feasible solution: the optional integration of cryptographic hashing, secure metadata anchoring (including GPS coordinates and device specifications), and optional digital identity signatures (eIDAS, EUDIS, or decentralised Self-Sovereign Identity frameworks) directly into media-capturing devices such as smartphones and cameras. These authenticity markers would be immutably anchored to public blockchain infrastructures—examples include Ethereum Layer 2 solutions, Solana, and Ardor—creating an incorruptible provenance ledger. The article examines the multiple advantages of such blockchain-based media authentication, including strengthened content integrity, clear provenance, rapid verification capabilities, deterrence of misinformation, protection of intellectual property, and resilience across decentralised platforms. Despite its promise, the implementation faces challenges related to privacy and anonymity concerns, device integration complexities, scalability, verification infrastructure, and evolving legal and regulatory landscapes. To overcome these barriers, parallel initiatives in user and stakeholder education are crucial, emphasising media literacy, transparency, and global inclusivity. By collaboratively addressing these issues, blockchain-based media authentication has the potential to significantly mitigate the impact of deepfakes and rebuild digital trust in society.

Keywords: Blockchain Media Authentication, Deepfake Detection, Digital Provenance, Synthetic Media, Content Integrity Verification, Cryptographic Timestamping,Media Forensics, Information Authenticity

Paper DOI: https://doi.org/10.48341/dx7k-e072

Format: From the Editors / Opinion-Paper

———

Introduction

In March 2022, as Russia's invasion of Ukraine started, a fabricated video of President Volodymyr Zelenskyy appeared online telling Ukrainian troops to surrender [1]. This deepfake—a video manipulated by AI to create realistic but false content—was quickly debunked, yet it stands as a stark example of how advanced AI can fuel dangerous disinformation. Around the world, similar incidents are multiplying at an alarming rate. A wave of deepfakes tied to recent elections in Europe and Asia has flooded social media, with fake videos of leaders endorsing opponents or confessing to wrongdoing [2]. In Slovakia's 2023 parliamentary election, a deepfake audio clip surfaced two days before voting, purportedly capturing pro-European candidate Michal Šimečka discussing electoral fraud with a prominent journalist, leading to speculation about whether this was "the first election swung by deepfakes" [3]. In Moldova, deepfake videos of President Maia Sandu have repeatedly appeared, including fabricated content showing her announcing her resignation and supporting pro-Russian candidates [4].

The threat extends beyond politics into finance, security, and everyday life. In 2019, criminals cloned a CEO's voice using AI and convinced a colleague to wire them €220,000 (approximately $243,000), marking the first known instance of AI voice mimicry being used for fraud [5]. More recently, a deepfake image of an "explosion" at the Pentagon went viral in May 2023, briefly causing a stock market plunge before being exposed as fake [6]. Fraudsters have even attempted real-time deepfake video calls to trick companies out of millions. From bogus videos of officials slurring their speech—like the doctored clip of U.S. Speaker Nancy Pelosi that spread widely on social media [7]—to non-consensual AI-generated pornography targeting private citizens and celebrities, manipulated media is a global menace. Most recently, an elaborate case of CEO fraud targeting a Dutch enterprise resulted in financial losses totalling €8.7 million, with the perpetrators employing advanced digital deception techniques such as deepfake videos, falsified meeting minutes, and synthetic voices. Through swift international cooperation and the alertness of Austrian financial institutions, approximately €1.5 million was recovered, illustrating both the sophistication of such crimes and the critical role of cross-border collaboration in mitigating their impact [27].

The landscape has become dramatically more challenging with the recent introduction of Google's VEO3, a state-of-the-art video generation model that represents a quantum leap in synthetic media capabilities [8]. Unlike previous AI video generators, VEO3 can create clips that are "nearly indistinguishable from real ones," complete with dialogue, soundtracks, sound effects, and adherence to real-world physics [9]. In its first week of release, users posted fake news segments in multiple languages, including fabricated death announcements and political news conferences [9]. As one expert warned, VEO3 could mean "the final death knell for truth on the internet" [10].

The common result of this technological escalation is an erosion of trust: if seeing and hearing can no longer be believed, the very notion of truth is undermined. This erosion of trust—termed the "liar's dividend" by researchers—allows bad actors not only to spread fake content but also to dismiss real footage as fake. The stakes for democracy and society are profound, as this phenomenon enables politicians and other public figures to evade accountability by falsely claiming that damaging but authentic content is fabricated [11].

Governments and tech platforms are scrambling for solutions. Regulators, mainly from the EU, have begun calling for labelling of AI-generated media, and researchers are improving deepfake detection tools. In a pioneering move to counter the growing threat of deepfakes, Denmark has announced an amendment to its copyright laws that would grant individuals legal rights over their own facial and vocal features. This legislative change aims to enhance protections against unauthorized digital imitations of a person's identity. By positioning biometric likenesses - such as one's face and voice - as intellectual property, the Danish government seeks to give individuals stronger legal recourse against AI-generated impersonations that can be used for malicious or deceptive purposes The decision to amend its copyright law by granting individuals ownership of their own facial and vocal features represents a forward-thinking approach to combat the misuse of deepfakes and protect personal identity in the digital age [29].

Northern Ireland's proposal to criminalize the creation and sharing of deepfake pornography reflects a critical step toward addressing the severe psychological harm, gendered abuse, and legal

gaps exposed by AI-driven identity manipulation, particularly in the absence of existing protections for adult victims [30].

Yet detection alone is a perpetual arms race, as AI-generated fakes grow more convincing with each technological advancement. A complementary approach is urgently needed: authentication of genuine media at the source. If we can prove a video or audio recording is real—timestamped, untampered, and originating from a verified source—then fake or manipulated content becomes easier to identify by absence of authenticity credentials.

This opinion paper introduces a technically grounded solution to authenticate audio-visual media using cryptographic techniques and blockchain technology. By anchoring the provenance of media files in tamper-proof digital ledgers, and optionally linking them to user identities, we can help restore trust in our information ecosystem. The proposal is platform-agnostic, leveraging open standards so it could be adopted across devices and integrated with various blockchains (from Ethereum and Solana to Business as a Service networks like Ardor), ushering in a future where truth in media can be verified by anyone.

A Vision for Blockchain-Based Media Authentication

Imagine if every photo, video, or audio clip captured on a device came with the option of a verifiable "seal of authenticity"—an unforgeable record of when, where, and by whom it was created. The core of this proposal is to achieve exactly this by combining three elements at the moment of capture: cryptographic hashing, trusted metadata (including GPS coordinates and temporal data), and digital signatures, all anchored to a blockchain. Crucially, this process would be built into devices like smartphones, body cameras, or DSLRs, so it runs seamlessly in the background for users. The end result is a public, tamper-evident ledger that can confirm whether a given media file is original and unaltered, or flag if anything is amiss.

How it would work:

- **One-Way Hash Creation & Timestamping:** As soon as a video, photo, or audio recording is created, the device computes a one-way cryptographic hash of the file using robust algorithms such as SHA-256 or SHA-3. This hash is essentially a digital fingerprint of the media—even the slightest change to the file (one altered pixel or audio sample) would produce a completely different hash value [12]. The hash, along with a trusted timestamp, is then prepared for recording on a blockchain. Modern blockchains or distributed ledgers can serve as global timestamp servers, since any data written to them comes with a reliable time of inclusion that cannot be retroactively altered.

  By recording the hash on a blockchain transaction, we obtain cryptographic proof of existence at a specific time. If someone later questions when a video was created, the blockchain's timestamp serves as an impartial witness that cannot be manipulated by any single party. Notably, only the hash (a short string of approximately 64 characters for SHA-256) is stored on-chain, not the media file itself—preserving efficiency, privacy, and keeping blockchain storage costs minimal [13]. Because cryptographic hashes are irreversible one-way functions, no one can derive the original content from the hash, but anyone with the content can recompute and compare the hash to verify integrity [12]. In essence, this creates a permanent, immutable record that "this exact file existed at this time" and has not been altered since.

- **GPS and Device Metadata Anchoring:** To further anchor the media in reality and provide comprehensive provenance information, certain metadata can be embedded or hashed alongside the file. This might include the GPS coordinates of the recording location (with appropriate privacy controls), the device model and camera sensor ID, lens settings, environmental data such as ambient light conditions, and network information. By capturing these details and cryptographically binding them to the media hash—for

example, by concatenating the metadata with the file data before hashing or including them in a signed manifest—we ensure the provenance of the content is recorded with high fidelity.

The blockchain entry could thus attest not just to a file's integrity, but also that "Video X was captured at latitude Y, longitude Z on Date/Time T by Camera model C with sensor ID S." This approach creates an expanded metadata watermark that is extremely difficult to forge. Missing or inconsistent metadata would indicate potential tampering or suspicious origins [14]. In practice, the device would likely include the GPS reading, device identifier, and other contextual information into the data it hashes and signs before blockchain submission.

This approach aligns with emerging standards like Adobe's Content Authenticity Initiative, where "each asset is cryptographically hashed and signed to capture a tamper-evident record" of its origins [15]. Indeed, camera manufacturers like Leica and Nikon have begun to build in such content credentials at the point of capture, highlighting the technical feasibility of embedding secure metadata [16]. Anchoring these details on a blockchain means anyone can later verify them independently—for example, a news editor could confirm that a photo's claimed GPS coordinates match the signed record on the ledger, providing an additional layer of verification beyond traditional metadata.

- **Digital User Signatures (Identity Binding):** As an optional but powerful layer, the device (or the user operating it) can digitally sign the hash/metadata record before it's submitted to the blockchain. This signature uses public-key cryptography: the creator (photographer, journalist, citizen) would have a private signing key, whose corresponding public key or certificate represents their identity. By signing the media's hash with their private key, the creator essentially provides a cryptographic attestation: "I certify I created this file." This adds accountability and source authentication on top of the file authenticity guarantee.

  The beauty of this design is that it can work with different identity systems, providing flexibility for various use cases and privacy requirements. A user could sign with a pseudonymous key for privacy protection, or use a verified digital identity for greater trust and legal standing. For instance, under the EU's eIDAS framework, a journalist might use a qualified electronic signature linked to their official ID, giving strong legal assurance of their identity and the content's authenticity [17]. As an outlook, the forthcoming EUDI-Wallet, developed under the eIDAS 2.0 regulation, is set to enable citizens across the EU to securely store and present digital credentials—such as identity documents, driving licences, and educational certificates—on their smartphones, supporting trusted cross-border authentication and qualified electronic signatures [28]. Alternatively (or in combination), decentralised identity (Self-Sovereign Identity) solutions could be used, where the signer's public key is tied to a Decentralised Identifier (DID) that others trust through a web of cryptographic attestations [18].

  In either case, the blockchain record would contain not only a hash and timestamp, but also a digital signature that viewers can verify against the author's public identity. This means that if a deepfake emerges purporting to be from Reporter Alice, one could see that it lacks Alice's signature or uses a key not registered to her, immediately casting doubt on its legitimacy. Digital signatures thus enable user authentication, complementing the media authentication provided by hashing. It's important to note this identity signing is optional—individuals concerned about privacy or reprisals might choose to remain anonymous or use organisational keys. But in scenarios where attribution is vital (e.g., official communications, journalism, scientific evidence), it can greatly enhance credibility and legal standing.

- **Blockchain Recording and Verification:** The final component is to record the hash (and any attached signature/metadata bundle) onto a blockchain infrastructure. This could be done in real-time for urgent content or asynchronously for routine media. A smartphone might batch and upload its authenticity records every time it connects to the internet, optimising for both user experience and network efficiency. The choice of blockchain

infrastructure is flexible—the solution is platform-agnostic by design, allowing for adaptation to different technical requirements and cost considerations.

The key requirements are a decentralised ledger that is tamper-proof, widely accessible for verification, and has sufficient throughput to handle global media authentication needs. For example, one could use an Ethereum Layer 2 network such as Polygon or Arbitrum to log these hashes, leveraging Ethereum's security while keeping costs and latency low [19]. Alternatively, the Solana blockchain, known for high throughput and low fees, could handle thousands of media authentication entries per second, which is useful if millions of users are recording media concurrently [20]. A different approach is exemplified by Ardor, a platform with a parent–child chain architecture; a dedicated child chain could store media hashes and metadata, benefiting from Ardor's security while allowing pruning of old records to manage blockchain bloat [21].

The New York Times' News Provenance Project piloted a scheme to record news photo metadata on a blockchain, embedding photographer identity, location, and timestamps into images and logging them on a ledger for readers to verify [22]. This proposal generalises that concept and extends it to audio and video, making authenticity attestation part of the media's creation process rather than a post-hoc addition.

Once the data is on-chain, verification becomes straightforward and accessible to anyone. Any individual, software application, or platform can look up the hash of a media file on the blockchain. If a matching record is found with valid signatures and metadata, it confirms the media's origin details and integrity. If no record is found for a video that claims to be "authentic," that absence itself becomes a red flag, especially once this system achieves widespread adoption. This creates a powerful incentive structure where authentic content is easily verifiable, while fabricated content lacks the cryptographic credentials that users and platforms come to expect.

Benefits of Blockchain-Based Media Authentication

A blockchain-based authentication framework for media would bring multiple benefits to the information ecosystem, strengthening both content integrity and source credibility whilst addressing the escalating challenges posed by advanced AI generation tools like VEO3:

- **Tamper-Proof Integrity:** The most immediate benefit is the assurance that audio or video content has not been manipulated since the moment of recording. Any alteration—from a subtle edit in an audio clip to a sophisticated deepfake swap of someone's face—will produce a mismatched hash, alerting viewers to tampering. This provides a content integrity guarantee that current watermarking or forensic techniques struggle to match, particularly as AI-generated content becomes increasingly sophisticated. As noted in recent research, uploading media and metadata to a public blockchain yields a secure version that cannot be altered without everyone noticing [23]. In other words, the ledger acts as an incorruptible audit trail: if the media's current form matches the chain record, it's demonstrably the original; if not, something has been altered.

  This capability is particularly powerful in combating subtle disinformation, where an authentic video might be slightly edited—for example, altering timestamps, splicing context, or making minor modifications that change meaning whilst maintaining apparent authenticity. Such changes would break the authenticity seal, immediately flagging the content as potentially compromised. Given that VEO3 can now generate content that is "nearly indistinguishable from real ones" [9], having cryptographic proof of authenticity becomes essential for distinguishing genuine content from even the most sophisticated synthetic media.

- **Provenance and Accountability:** Tying media to metadata and creator signatures means we can establish comprehensive provenance—the complete "story" of where a piece of

content came from, when it was created, and by whom. This is vital for journalism, science, legal proceedings, and historical records. With this system, a news photo published online could carry an embedded provenance link that provides verifiable information about its creation. A reader who sees the photo can access verification data showing, for instance, that it was "captured by Jane Doe at 14:32 GMT on Nikon Camera X, in London," with a signature from Jane's news organisation, all verifiable on the blockchain [15], [22].

This helps counteract misattribution and plagiarism whilst providing protection for content creators. It also holds creators accountable: if a piece of media is misrepresented, we know who originally published it and can trace its authentic origins. Conversely, it protects creators from false accusations of fakery. A journalist accused of posting a doctored video, for example, could point to the blockchain ledger to prove it's original—akin to having a digital notary that cannot be corrupted or influenced. The dual authentication of both media and user provides a complete chain of trust from capture to consumption.

- **Rapid Verification for Viewers and Platforms:** In a world of instant information, this system allows for near-instantaneous verification, which is crucial given the speed at which misinformation can spread. Imagine scrolling social media and seeing a startling video of a public figure. If the platform is integrated with the authenticity ledger, it could display a "Verified Authentic" badge if the media has a matching blockchain record, or flag it as "Unverified" if not. Readers and viewers wouldn't need to be cryptography experts; user-friendly indicators can convey authenticity status through intuitive visual cues.

  Media outlets and fact-checkers would gain a powerful tool—instead of hours spent tracing the source of a viral clip through traditional investigative methods, a blockchain lookup could confirm authenticity in seconds. This speed is crucial during breaking news or election seasons when fake media can spread faster than truth can catch up [24]. The system would be particularly valuable in combating the type of rapid-spread misinformation seen with the Pentagon explosion deepfake, which briefly affected stock markets before being debunked [6].

- **Deterrence of Disinformation:** Widespread adoption of authenticated media raises the cost and difficulty for disinformation actors. Currently, creating and spreading fake content is relatively easy and low-risk. However, if authentic media becomes the expected norm, with clear verification mechanisms, fake content becomes more conspicuous by its absence of credentials. This creates a powerful deterrent effect: bad actors would need to either compromise the authentication system itself (extremely difficult with proper cryptographic implementation) or accept that their content will be flagged as potentially fake.

  The deterrent effect is amplified by the fact that blockchain records are permanent and traceable. Unlike traditional media where provenance can be obscured or lost, blockchain-authenticated content maintains its verification trail indefinitely. This makes it much more difficult for disinformation campaigns to operate at scale, as each piece of fake content becomes a potential liability that can be traced and exposed.

- **Protection of Intellectual Property:** The proposed system provides robust protection for content creators' intellectual property rights. By timestamping and signing media at the point of creation, creators establish indisputable proof of ownership and creation date. This is particularly valuable for photographers, videographers, and other media professionals whose work might be stolen or misattributed. The blockchain record serves as immutable evidence in legal proceedings, providing stronger protection than traditional copyright mechanisms.

  Furthermore, it can help combat the unauthorised use of someone's likeness in deepfakes. If a public figure's authentic media is consistently blockchain-verified, it becomes much easier to identify and challenge unauthorised synthetic content that lacks proper authentication credentials.

- **Resilience Across Decentralised Platforms:** Unlike centralised verification systems that depend on a single authority or platform, blockchain-based authentication works across any platform or service that chooses to implement verification checks. This creates a universal standard that isn't controlled by any single entity, reducing the risk of censorship or manipulation by powerful actors. The decentralised nature also means the system remains functional even if individual platforms or services fail or are compromised.

  This resilience is particularly important in the context of global information flows, where content may be shared across multiple platforms and jurisdictions. A video authenticated on one platform remains verifiable on any other platform that implements the verification protocol, creating a robust, interconnected web of trust that transcends individual platform boundaries.

- **Enhanced Trust in Democratic Processes:** Perhaps most importantly, the system can help restore trust in democratic processes by providing reliable verification for political content. Election-related media, official government communications, and political advertisements could all be authenticated, making it much more difficult for bad actors to spread election disinformation or create fake political content. This is particularly crucial given the documented impact of deepfakes on elections, such as the Slovak case where fabricated audio may have influenced electoral outcomes [3].

  The system would also help combat the "liar's dividend" phenomenon, where politicians and other public figures dismiss authentic but damaging content as potentially fake [11]. With blockchain authentication, the authenticity of content becomes cryptographically verifiable, making it much more difficult to falsely claim that genuine content is fabricated.

Implementation Challenges and Considerations

Whilst the potential benefits of blockchain-based media authentication are substantial, several significant challenges must be addressed to ensure successful implementation and widespread adoption.

- **Privacy and Anonymity Concerns:** One of the primary challenges involves balancing authentication with privacy protection. The system's requirement to record metadata such as GPS coordinates, device identifiers, and potentially user signatures raises legitimate concerns about surveillance and privacy erosion. Citizens in authoritarian regimes, whistleblowers, and activists may require the ability to create verifiable content whilst maintaining anonymity to protect themselves from persecution or retaliation.

  To address these concerns, the system must incorporate sophisticated privacy-preserving mechanisms. Zero-knowledge proof systems could allow verification of authenticity without revealing sensitive metadata. Ring signatures or group signatures could enable authentication whilst obscuring the specific identity of the creator within a larger group. Additionally, the system should support multiple privacy levels, allowing users to choose between full transparency (for official communications), partial disclosure (revealing location but not identity), or maximum privacy (anonymous verification with minimal metadata).

  The implementation must also comply with privacy regulations such as the General Data Protection Regulation (GDPR) in Europe, which grants individuals rights over their personal data. This requires careful consideration of what data is stored on-chain versus off-chain, and how users can exercise their rights to data portability and erasure whilst maintaining the integrity of the authentication system.

- **Device Integration Complexities:** Integrating authentication capabilities into existing devices presents significant technical and economic challenges. Smartphones, cameras, and

other recording devices would need hardware and software modifications to support cryptographic operations, secure key storage, and blockchain connectivity. This requires coordination across multiple industries, including device manufacturers, operating system developers, and blockchain infrastructure providers.

The implementation must be designed to work across a diverse ecosystem of devices with varying computational capabilities. Whilst high-end smartphones may easily support sophisticated cryptographic operations, lower-cost devices or older equipment may require optimised implementations or cloud-assisted processing. The system must also account for devices that operate in environments with limited or intermittent internet connectivity, requiring robust offline capabilities and synchronisation mechanisms.

Furthermore, the authentication system must be designed to be backward-compatible and forward-compatible, ensuring that content created with early implementations remains verifiable as the technology evolves. This requires careful standardisation efforts and coordination among industry stakeholders.

- **Scalability and Infrastructure Requirements:** The global scale of media creation presents enormous scalability challenges. Billions of photos, videos, and audio recordings are created daily, each potentially requiring blockchain authentication. Current blockchain networks, even high-throughput systems like Solana, would struggle to handle this volume without significant infrastructure improvements.

  Solutions may include hierarchical authentication systems where individual devices create local authentication records that are periodically batched and submitted to the blockchain. Layer 2 scaling solutions, such as state channels or rollups, could handle high-frequency authentication whilst maintaining the security guarantees of the underlying blockchain. Alternative approaches might involve federated blockchain networks or hybrid systems that combine blockchain anchoring with traditional database systems for improved performance.

  The infrastructure must also be designed to handle peak loads during major news events when media creation and verification demands spike dramatically. This requires robust load balancing, redundancy, and failover mechanisms to ensure system reliability when it's needed most.

- **Verification Infrastructure and User Experience:** Creating an effective verification infrastructure requires significant investment in user-facing tools and services. Platforms, applications, and services must integrate verification capabilities in ways that are intuitive and accessible to non-technical users. This includes developing standardised APIs, user interface guidelines, and educational resources to support widespread adoption.

  The verification process must be fast and reliable, providing near-instantaneous results without requiring users to understand the underlying cryptographic mechanisms. This necessitates the development of efficient verification algorithms, distributed verification networks, and caching mechanisms to ensure good user experience even under high load conditions.

  Additionally, the system must handle edge cases gracefully, such as content that has been legitimately edited or processed whilst maintaining its essential authenticity. This requires sophisticated policies and technical mechanisms to distinguish between acceptable modifications (such as format conversion or compression) and substantive alterations that compromise authenticity.

- **Legal and Regulatory Hurdles:** The legal landscape surrounding blockchain-based authentication is complex and evolving. Different jurisdictions have varying approaches to digital signatures, blockchain evidence, and data protection, creating challenges for global implementation. Frameworks like the EU's eIDAS provide supportive environments for

digital signatures, lending legal legitimacy to authenticated media [17]. However, regulatory challenges include scepticism around blockchain technology due to concerns over energy consumption, cryptocurrency volatility, and regulatory uncertainty.

- Legal guidelines may require updating to address the admissibility of blockchain-timestamped digital evidence in court proceedings. Questions arise about the legal status of smart contracts, the enforceability of blockchain records across jurisdictions, and the liability of various parties in the authentication ecosystem. International coordination is necessary to avoid fragmented national standards that could undermine the system's global effectiveness.

  Regulatory alignment with laws such as the EU's Digital Services Act could incentivise platforms to prioritise or label verified content [25]. However, this also raises questions about potential regulatory capture, where large technology companies might gain disproportionate influence over authentication standards and implementation.

- **Economic and Incentive Considerations:** The economic model for blockchain-based authentication must carefully balance costs, incentives, and accessibility. Transaction fees for blockchain recording, infrastructure costs for verification services, and development costs for device integration all represent significant economic barriers that could limit adoption, particularly in developing regions where the need for authentication may be greatest.

  The system must be designed to be economically sustainable whilst remaining accessible to users regardless of their economic circumstances. This might involve subsidised authentication for certain types of content (such as news reporting or human rights documentation), tiered pricing models, or alternative funding mechanisms such as public-private partnerships or international development funding.

  Incentive structures must also encourage honest participation whilst discouraging gaming or abuse of the system. This includes mechanisms to prevent spam, ensure quality of authentication services, and maintain the integrity of the verification infrastructure over time.

- **Addressing Limitations and Potential Abuse:** The proposed system does not automatically identify AI-generated media that never undergoes authentication, nor does it prove the inherent truthfulness of an event depicted in authenticated content. A video might be authentic in the sense that it hasn't been digitally manipulated, but still misleading due to context, staging, or selective editing. Clear communication about these limitations is essential to prevent overreliance on authentication as a guarantee of truth rather than merely technical authenticity.

  There's also potential for abuse or misinformation about the system itself. Bad actors might create fake authentication claims, spread confusion about how the system works, or attempt to undermine trust in legitimate authentication mechanisms. Countermeasures include clear public education, transparency in system design, and robust mechanisms for detecting and responding to abuse.

  Authoritarian misuse represents another significant concern. Governments might attempt to mandate authentication in ways that enable surveillance or censorship, or create parallel authentication systems that serve political rather than truth-seeking purposes. Strong safeguards are necessary, including support for anonymous or group signatures, international oversight mechanisms, and technical designs that resist centralised control or manipulation.

  The system must also be designed to resist various attack vectors, including attempts to compromise device security, manipulate blockchain records, or create false authentication

credentials. This requires ongoing security research, regular system updates, and robust incident response capabilities to address emerging threats as they develop.

Education and Adoption:

- **The Human Factor:** Technological solutions alone cannot solve what is fundamentally a human trust problem. For a blockchain-based media authentication system to succeed, a comprehensive effort in user and stakeholder education is essential. Journalists, content creators, policymakers, and the general public all need to understand what this technology does, how to use it properly, and what its limitations are.

- **For Media Professionals and Technologists:** News organisations and reporters should be trained in new workflows that incorporate authenticity seals into their content creation processes. Just as journalists learned to use encryption tools for secure communications, they can learn to use authentication tools for their photos and videos. This might involve incorporating capture applications or camera firmware updates, and using verification services when sourcing content from the public.

- Media outlets could establish verification desks that not only fact-check content but also authenticity-check user submissions using blockchain verification tools. This represents an evolution of traditional editorial processes, adding a technical verification layer to complement journalistic investigation. Training programmes should emphasise both the capabilities and limitations of the technology, ensuring that media professionals understand that authentication verifies technical integrity rather than truthfulness of depicted events.

- Technologists and developers will need to collaborate on integrating these features into popular software and platforms. For example, photo and video editing programs might maintain content credentials transparently, recording any modifications whilst preserving the original authentication chain. Adobe's Content Authenticity Initiative already promotes such integration, ensuring edits are recorded transparently in provenance trails [15]. The development community must work together to create standardised tools, libraries, and protocols that make authentication integration straightforward for application developers.

- **For Platforms and Industry:** Social media companies and information platforms must be actively involved, as they are critical distribution points for media content. These companies should educate users about authenticity indicators and integrate verification capabilities into their platforms. For instance, a video tagged as "Source Verified" could come with a brief explainer if clicked: "This video's origin and integrity have been verified via a public digital ledger."

- Platforms might run media literacy campaigns to encourage people to verify content before sharing, analogous to existing campaigns teaching people to spot misinformation [24]. These educational efforts should be culturally sensitive and available in multiple languages to ensure global accessibility. Industry coalitions can set common user experience standards, ensuring clarity and consistency for users across different platforms and services.

- The integration should be designed to be helpful rather than intrusive, providing clear visual indicators without overwhelming users with technical details. Platforms should also develop policies for handling unverified content, balancing the need to flag potentially problematic material with avoiding censorship or suppression of legitimate content that simply lacks authentication credentials.

- **For Policymakers and Public Institutions:** Governments can play a leading role by using authenticated media in official communications, demonstrating the technology's value and building public familiarity. For example, a police department could release bodycam

footage with a public authenticity certificate, making conspiracy theories less credible and increasing public trust in official accounts of events. Election authorities might require authenticated political advertisements to prevent deepfake campaigns and increase transparency in political communications.

- Officials should educate citizens about authenticity's importance, perhaps including it in digital literacy curricula or public service announcements. This education should emphasise both the benefits and limitations of authentication technology, helping citizens understand that verified authenticity doesn't guarantee truthfulness but does provide important information about content integrity and provenance.

  Policymakers could mandate authenticity standards for sensitive areas like election advertisements or official communications, using legislative frameworks such as the EU's Digital Services Act [25]. However, such mandates must be carefully designed to avoid creating barriers to legitimate expression or enabling authoritarian control over information flows.

  International cooperation is essential to ensure that authentication standards work across borders and jurisdictions. This might involve working through international organisations such as the United Nations or World Economic Forum to develop global standards and best practices for media authentication.

- **Engaging the Public:** Ultimately, everyday media consumers need awareness of both the possibilities and limits of this system. Users should understand that "verified authentic" means content is likely untampered and from a legitimate source, whereas "unverified" content should be approached with appropriate caution. However, authenticity does not guarantee truthfulness of the depicted event—a video might be technically authentic but still misleading due to context, staging, or selective presentation.

  Public understanding grows through experience and targeted education campaigns by civil society organisations. Nonprofits combating misinformation could create accessible guides or videos explaining blockchain-based content verification clearly, demystifying the underlying cryptography whilst focusing on practical implications for media consumers. The educational approach should be empowering rather than intimidating, helping individuals develop the skills to navigate an increasingly complex media landscape.

  Educational initiatives should address common misconceptions about both blockchain technology and media authenticity. Many people may have negative associations with blockchain due to cryptocurrency volatility or environmental concerns, requiring careful explanation of how authentication systems differ from speculative financial applications. Similarly, education should address the difference between technical authenticity and truthfulness, helping users develop nuanced understanding of what verification can and cannot tell them.

Inclusivity and Global Reach

We must ensure this solution doesn't only serve technologically advanced or wealthy nations. Disinformation is a global problem, often most severe in regions lacking robust verification resources or where press freedom is limited. Therefore, authenticity tools should be open-source where possible and widely accessible across different economic and technological contexts.

Leveraging widespread smartphone adoption, even affordable devices could support authenticity hashing through optimised implementations or cloud-assisted processing. International organisations could endorse standards and promote cross-border cooperation, ensuring that authentication systems work globally rather than creating new digital divides.

Consider a scenario where citizen journalists globally use authenticity-enabled devices: their content would be immediately trusted due to common verification systems, regardless of their location or the platform they use for distribution. This could be particularly valuable in regions where traditional media is restricted or unreliable, giving citizens tools to document and share authentic information about local conditions.

The implementation must account for varying levels of technological infrastructure, internet connectivity, and digital literacy across different regions. This might involve developing lightweight authentication protocols for low-bandwidth environments, offline verification capabilities, and educational materials adapted to different cultural and linguistic contexts.

Technology adoption is inherently social, requiring advocates in newsrooms, supportive legislation, and public awareness campaigns. Educational investment must accompany technical rollout to ensure that users understand and trust the system. Properly executed, users will not feel overwhelmed by cryptographic complexity—instead, they will experience a more trustworthy media environment where information can be reliably checked, and misinformation confronted effectively.

The success of blockchain-based media authentication ultimately depends not just on technical excellence, but on building a global community of users who understand, trust, and actively participate in maintaining the integrity of our shared information ecosystem. This requires sustained commitment to education, inclusivity, and transparency from all stakeholders involved in the system's development and deployment.

Conclusion: Toward a Verifiable Future for Media

The proliferation of AI-generated fake media, dramatically accelerated by technologies such as Google's VEO3, represents an existential challenge to truth in the digital age. The ability to create photorealistic videos with synchronised audio that are "nearly indistinguishable from real ones" [9] has fundamentally altered the information landscape, making traditional approaches to content verification increasingly inadequate. In response, this article has outlined an ambitious yet achievable countermeasure: leveraging the power of cryptography and blockchain technology to prove when media is authentic.

By embedding authentication at the point of creation—through cryptographic hashes, timestamps, and digital signatures saved on decentralised ledgers—we can create a new paradigm of digital trust that addresses both current and emerging threats. This proposal is intentionally platform-agnostic and human-centric, recognising that no single technology will serve as a complete solution. Rather, it combines the strengths of several complementary approaches: the immutability of blockchains, the rigour of public-key cryptography, and the context provided by comprehensive metadata. Together, these elements can form a robust backbone for content provenance that spans across Ethereum [19], Solana [20], Ardor [21], and other emerging networks; across smartphones and professional cameras; across news studios and social networks.

The impact of such a system could be transformative in addressing the challenges posed by increasingly sophisticated synthetic media. Consider a near future in which a deepfake video of a world leader emerges during a tense election period. Instead of spreading unchecked and potentially influencing electoral outcomes, it would be quickly scrutinised by journalists and automated verification systems. Finding no matching authenticity record—no cryptographic hash or digital signature—in the public ledger, experts could label it inauthentic within minutes. The public, accustomed to looking for verification indicators, would remain appropriately sceptical of the video, and it would fail to gain traction. Conversely, when genuine but shocking footage surfaces—such as evidence of human rights violations—the authentication framework would provide immediate credibility, thwarting attempts to dismiss it as fabricated. Thus, both rampant disinformation and cynical denial of reality could be effectively countered.

This approach directly addresses the "liar's dividend" phenomenon identified by researchers [11], where the mere possibility of sophisticated fakes allows bad actors to dismiss authentic but inconvenient content as potentially fabricated. By providing cryptographic proof of authenticity, the

system makes it much more difficult for public figures to evade accountability by falsely claiming that genuine content is fake. Authenticity becomes a verifiable property rather than a matter of opinion or political convenience.

No technology can eliminate deception entirely, and new forms of AI-enabled manipulation will undoubtedly emerge as the technology continues to advance. However, shifting the information battleground to one where truth has robust technological defences represents a significant improvement over the current situation. Blockchain-authenticated media creates an environment where honest creators have powerful tools to reinforce their credibility, and audiences can quickly apply verification tests to content they encounter. It reinforces the fundamental journalistic principle of "verify, verify, verify" with modern technological capabilities that match the sophistication of contemporary threats.

Importantly, this approach safeguards core democratic values by enhancing transparency whilst respecting privacy through optional anonymity features, and by relying on decentralised networks rather than centralised authorities or censors [26]. The system's design principles prioritise user agency and resist authoritarian control, ensuring that authentication serves truth-seeking rather than political manipulation.

The path forward requires coordinated action from multiple stakeholders. Policymakers can foster supportive regulatory environments, perhaps mandating authenticity standards for sensitive content areas such as election communications whilst ensuring that such requirements don't stifle legitimate expression. Technologists need to address interoperability and security challenges, learning from pilot projects and scaling infrastructure to meet global demands. Media professionals and educators will serve as crucial ambassadors, integrating these practices into daily reporting and building public awareness of both capabilities and limitations.

Early implementations, including Adobe's Content Credentials initiative [15], the New York Times' News Provenance Project [22], and camera manufacturer initiatives [16], indicate that the technical foundations are falling into place. The challenge now is to scale these approaches whilst addressing the economic, social, and political barriers to widespread adoption.

The emergence of VEO3 and similar technologies has made this work more urgent than ever. As one expert warned, such tools could represent "the final death knell for truth on the internet" [10] if left unchecked. However, the same technological capabilities that enable sophisticated synthetic media creation can also power robust authentication systems. The question is whether society will act quickly enough to deploy these defences before the information ecosystem suffers irreparable damage.

In conclusion, cryptographic authentication of audio, photo, and video content via blockchain represents a bold and forward-looking solution to the pressing problem of synthetic media. It is visionary yet technically grounded, tapping into the fundamental qualities of distributed ledgers: transparency, permanence, and trust without reliance on single authorities. By anchoring digital content to an immutable timeline of provenance, we can reclaim a measure of certainty in the digital realm.

Truthful media should no longer be vulnerable to disruption by viral fakes or dismissal through false claims of fabrication. Instead, armed with cryptographic proof and collective resolve, we can ensure that reality has formidable defences in the digital age. This represents an investment not merely in technology, but in the resilience of our information society and the preservation of democratic discourse. The sooner we act to implement and embrace these systems, the sooner we can blunt the impact of deepfakes and raise effective shields for truth. In an age where seeing can no longer automatically mean believing, we must create new foundations for trust—and blockchain-based authentication offers a promising path forward.

Disclaimer

We utilised DeepL Write for spell-checking.

This opinion paper did not receive any specific funding from public, commercial, or non-profit organizations. The views expressed herein are based solely on the analysis of publicly available information and are intended to contribute to scientific and academic discourse. Any references to current events are drawn from sources in the public domain and are not intended to reflect the personal political opinions of the author(s).

About the Authors

DR. DR. ALEXANDER PFEIFFER: A technology and game studies enthusiast, He specializes in blending speculative fiction and academic inquiry. His work focuses on how gaming, virtual worlds, and emerging technologies shape our understanding of society and ethics.

https://www.alexpfeiffer.at | https://www.researchgate.net/profile/Alexander_Pfeiffer

NANDITHA KRISHNA graduated with a five-year integrated Master's degree (M.A.) in English Language and Literature from Amrita Vishwa Vidyapeetham, Amritapuri Campus, India. She is currently a Future Days 2025 Fellow, administered by the Copenhagen Institute for Futures Studies (Copenhagen, Denmark), Media Lab Bayern (Munich, Germany), and Future Days (Lisbon, Portugal). She is also a Research Assistant at the Australian Research Centre for Interactive and Virtual Environments (IVE) at the University of South Australia (Adelaide, Australia), where she contributes to projects focused on interactive narratives, news games, digital art, virtual reality (VR), and creativity in artistic performance using immersive technology. Her interests span the disciplines of literary studies—particularly speculative fiction studies, media studies, and future studies—to explore the impact of digital technologies on culture and society. From 2021 to 2023, she was a Humanities, Arts, Science, and Technology Alliance and Collaboratory (HASTAC) Scholar and a Research Assistant at the Empathic Computing Lab at the University of Auckland (Auckland, New Zealand). Read more about her at:

https://nandithakrishna.home.blog/

References

[1] NPR, "Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn," Mar. 16, 2022. [Online]. Available: https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

[2] Associated Press, "AI chatbots have been used to create dozens of news content farms," Mar. 14. 7, 2023. [Online]. Available: https://apnews.com/article/artificial-intelligence-elections-disinformation-chatgpt-bc283e7426402f0b4baa7df280a4c3fd

[3] L. de Nadal and P. Jančárik, "Beyond the deepfake hype: AI, democracy, and 'the Slovak case'," Harvard Kennedy School Misinformation Review, Aug. 22, 2024. [Online]. Available: https://misinforeview.hks.harvard.edu/article/beyond-the-deepfake-hype-ai-democracy-and-the-slovak-case/

[4] Balkan Insight, "Moldova Dismisses Deepfake Video Targeting President Sandu," Dec. 29, 2023. [Online]. Available: https://balkaninsight.com/2023/12/29/moldova-dismisses-deepfake-video-targeting-president-sandu/

[5] J. Damiani, "A Voice Deepfake Was Used To Scam A CEO Out Of $243,000," Forbes, Sep. 3, 2019. [Online]. Available: https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

[6] S. Bond, "Fake viral images of an explosion at the Pentagon were probably created by AI," NPR, May 22, 2023. [Online]. Available: https://www.npr.org/2023/05/22/1177590231/fake-viral-images-of-an-explosion-at-the-pentagon-were-probably-created-by-ai

[7] Reuters, "Fact check: 'Drunk' Nancy Pelosi video is manipulated," Aug. 3, 2020. [Online]. Available: https://www.reuters.com/article/world/fact-check-drunk-nancy-pelosi-video-is-manipulated-idUSKCN24Z2B1/

[8] Google DeepMind, "Veo - Google DeepMind," 2025. [Online]. Available: https://deepmind.google/models/veo/

[9] A. R. Chow and B. Perrigo, "Google's New AI Tool Generates Convincing Deepfakes of Riots, Conflict, and Election Fraud," TIME, Jun. 3, 2025. [Online]. Available: https://time.com/7290050/veo-3-google-misinformation-deepfake/

[10] R. Circelli, "I Tested Out Google's Veo 3 AI Video Generator. The Internet Is Not Prepared for What's Coming," PCMag, Jun. 10, 2025. [Online]. Available: https://www.pcmag.com/opinions/i-tested-out-googles-veo-3-ai-video-generator-the-internet-is-not-prepared

[11] K. J. Schiff, D. S. Schiff, and N. S. Bueno, "The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability?" American Political Science Review, vol. 119, no. 1, pp. 71-90, Feb. 2025. DOI: 10.1017/S0003055423001454.

[12] National Institute of Standards and Technology, “Secure Hash Standard (SHS),” FIPS PUB 180-4, Aug. 2015. Available: https://csrc.nist.gov/pubs/fips/180-4/upd1/final

[13] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System,” 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[14] H. Farid, “Photo Forensics,” MIT Press, 2016. Available: https://mitpress.mit.edu/9780262537001/photo-forensics/

[15] Adobe, “Content Authenticity Initiative,” 2021. [Online]. Available: https://contentauthenticity.org/

[16] Leica Camera AG, “Leica M11-P: The first camera with Content Credentials,” Oct. 2023. [Online]. Available: https://leica-camera.com/de-AT/fotografie/kameras/m/m11-p-schwarz?gad_source=1&gad_campaignid=21099665978&gbraid=0AAAAABbirkNDFk2smYN20-wmFbUPvhTZv&gclid=CjwKCAjw9uPCBhATEiwABHN9K1VBFppKRYGLp_gFuLCjc1M0qxVkeSf2QIMAzslgnr-6z0V_LPmRWRoC3ksQAvD_BwE&gclsrc=aw.ds

[17] European Parliament and Council, “Regulation (EU) No 910/2014 on electronic identification and trust services for electronic transactions in the internal market (eIDAS Regulation),” Jul. 23, 2014.

[18] W3C, “Decentralized Identifiers (DIDs) v1.0,” Jul. 19, 2022. [Online]. Available: https://www.w3.org/TR/did-core/

[19] Ethereum Foundation, “Ethereum Whitepaper,” 2013. [Online]. Available: https://ethereum.org/en/whitepaper/

[20] Solana Labs, “Solana: A new architecture for a high performance blockchain,” 2017. [Online]. Available: https://solana.com/solana-whitepaper.pdf

[21] Jelurida, “Ardor Platform,” 2016. [Online]. Available: https://ardordocs.jelurida.com/Faq

[22] S. Koren,, “Introducing the News Provenance Project,” on Medium, 2019. [Online]. Available: https://open.nytimes.com/introducing-the-news-provenance-project-723dbaf07c44

[23] M. Swan, “Blockchain: Blueprint for a New Economy,” O’Reilly Media, 2015. Available: https://www.academia.edu/44112222/Melanie_Swan_Blockchain_BLUEPRINT_FOR_A_NEW_ECONOMY

[24] Reuters Institute, “Digital News Report 2024,” Oxford: Reuters Institute for the Study of Journalism, 2024. Available: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2024-06/RISJ_DNR_2024_Digital_v10%20lr.pdf

[25] European Parliament and Council, “Regulation (EU) 2022/2065 on a Single Market For Digital Services (Digital Services Act),” Oct. 19, 2022. Available: https://www.eu-digital-services-act.com/

[26] T. Berners-Lee, “Decentralised Information Management,” CERN, 1990. Available: https://home.cern/science/computing/birth-web/short-history-web#:~:text=By%20the%20end%20of%201990,This%20machine%20is%20a%20server

[27] D. Taljat, “CEO-Fraud-Bande erbeutet 8,7 Millionen Euro – Ermittler sichern Teilbetrag in Österreich,” Computerwelt IT & Business, Jun. 2, 2025. [Online]. Available: https://ittbusiness.at/article/ceo-fraud-bande-erbeutet-8-mio-euro-ermittler-sichern-teilbetrag-in-oesterreich

[28] European Commission, “The European Digital Identity Wallet,” Shaping Europe’s Digital Future, 2025. [Online]. Available: https://ec.europa.eu/digital-building-blocks/sites/display/EUDIGITALIDENTITYWALLET/EU+Digital+Identity+Wallet+Home

[29] The Guardian, “Denmark to tackle deepfakes by giving people copyright to their own features,” Jun. 27, 2025. [Online]. Available: https://www.theguardian.com/technology/2025/jun/27/deepfakes-denmark-copyright-law-artificial-intelligence

[30] BBC, “MLA says deepfakes ruin lives, as consultation on criminalisation opens “Jun. 28, 2025. [Online]. Available: https://www.bbc.com/news/articles/cwyxv9k28evo