ISSN: 2366-1313

DEEPFAKE PREVENTION USING BLOCKCHAIN AND AI FOR VISUAL MEDIA

KETHINENI VINOD KUMAR*

kethineni.vinod@gmail.com, Assistant Professor

KIMAVATH SURESHNAIK†

r190383@rguktrkv.ac.in, Bachelor of Technology

CHALAM SANJAY‡

r190155@rguktrkv.ac.in, Bachelor of Technology

Department of Computer Science and Engineering,

Rajiv Gandhi University of Knowledge Technologies, RK Valley, Andhra Pradesh, India.

Abstract: In today's digital age, the boundary between real and fabricated media is becoming increasingly blurred due to advancements in artificial intelligence. One of the most pressing concerns arising from this is the misuse of deepfakes, particularly those involving children and public figures which poses serious threats to digital privacy and identity security. This paper proposes a comprehensive, multi-layered framework designed to protect images and videos from deepfake exploitation. It integrates blockchain technology, AI-driven image analysis, and real-time watermarking to ensure authenticity and ownership verification. The system initiates protection at the point of media capture by embedding a unique, invisible watermark containing the owner's identification details. This information, along with relevant metadata, is securely stored on a blockchain ledger to establish an immutable and tamper-proof record of ownership. Each attempt to edit or manipulate protected media is logged as a blockchain transaction, and modifications are only permitted after obtaining a one-time password (OTP) from the owner, ensuring informed and traceable consent. The importance of this protection is underscored by real-world campaigns like Deutsche Telekom's powerful advertisement featuring a deepfake of a fictional nine-year-old girl named Ella. This campaign illustrates the risks of ”sharenting”, the excessive online sharing of children's images. The paper includes a visual architecture that outlines the full process, from content capture to final verification, demonstrating the integration of watermarking, blockchain-based proof of ownership, AI validation, and user-controlled permission management. By merging these technologies, the proposed model provides a secure and effective solution to combat deepfakes, reinforce digital rights, and rebuild public trust in digital media.

Keywords: Deepfake, Deepfake Prevention, OTP, Blockchain, Artificial Intelligence, Steganography, Image Analysis

I. INTRODUCTION

The rapid evolution of artificial intelligence (AI) has rev-olutionized the digital landscape, offering groundbreaking in-novations across various sectors including healthcare, finance, education, and entertainment. Among its many applications, AI-powered generative models particularly Generative Adver-sarial Networks (GANs) have enabled the creation of hyper-realistic synthetic media known as deepfakes. These technologies can generate video, audio, and images that are virtually indistinguishable from authentic content, making it difficult for the human eye or even conventional algorithms to discern real from fake. Although initially used for entertainment, film production, and gaming industries to enhance creativity and reduce production costs, deepfakes have increasingly become a tool for malicious intent. From the spread of political misin-formation to the impersonation of individuals in fraudulent activities, deepfakes pose an escalating threat to personal security, digital authenticity, and societal trust in media.

The implications of deepfake technology go beyond mis-information and fraud; they strike at the core of digital identity and consent. One of the most vulnerable demograph-ics impacted by this technology is children. In an era of ”sharenting” a cultural phenomenon where parents extensively share photos, videos, and milestones of their children on social media platforms a massive trove of personal data has been created, often without the child’s knowledge or consent. This data becomes an easy target for malicious AI systems to harvest, manipulate, and weaponize. A stark example is Deutsche Telekom’s “Ella” campaign, where a deepfake of a child speaks directly to her parents in a cinema, raising awareness about how their innocent sharing habits could lead to long-term consequences. The video emphasizes how easily a child’s digital footprint can be exploited, predicting that by 2030, over two-thirds of identity theft cases may stem from such oversharing. This not only raises ethical concerns but also demands immediate intervention through robust technological and policy-level solutions.

The reactive nature of current deepfake detection systems has proven insufficient in combating the speed and sophisti-cation of AI-generated media. Most existing methods rely on post-distribution detection identifying a deepfake only after it has already circulated online and potentially caused damage. Furthermore, these systems are often reliant on supervised learning, which requires a constantly updated dataset of fake content to train models effectively. Given the adaptive nature of AI, attackers can easily circumvent these detection techniques by modifying generation methods or training al-gorithms. What is urgently needed is a shift from reactive detection to proactive prevention.

This review proposes a comprehensive, multi-layered frame-work that emphasizes prevention by integrating blockchain, real-time watermarking, and AI-powered image verification. The core idea is to embed authenticity and ownership at the source during content creation rather than attempting to verify content after it has gone viral. Watermarking technology allows for the insertion of unique, tamper-resistant identifiers into media files without degrading quality. These digital sig-natures can be visible or invisible, but they serve as proof of origin. The ownership and metadata of the content are then securely logged on a decentralized blockchain network, ensur-ing transparency, immutability, and verifiability. This record-keeping mechanism not only confirms the original creator but also provides a legal foundation for copyright and user consent enforcement.

To safeguard the integrity of media throughout its lifecycle, the system includes an OTP-based permission layer. Every request to edit, repost, or alter the content must be validated by the original owner through a time-sensitive password, thus establishing a traceable, permissioned chain of custody. Complementing this structure is an AI module designed to scan digital content across platforms and validate whether it has been watermarked and registered on the blockchain. Unregistered files are flagged as suspicious and potentially manipulated, triggering alerts for further human or automated inspection. This ensures ongoing protection and verification, even after the media has been disseminated.

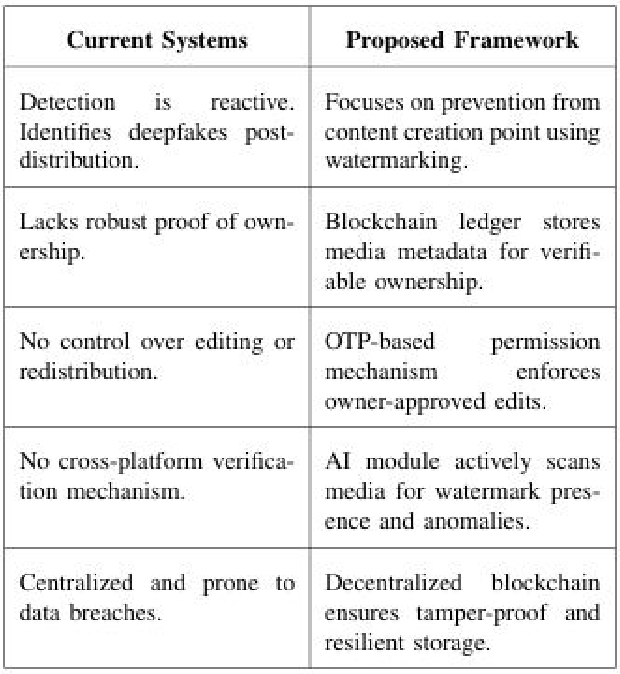
Beyond the technical dimensions, this paper also addresses the broader ethical and societal implications of deepfake technology. Misuse of media, especially involving children, women, and public figures, has significant psychological and reputational consequences. In many jurisdictions, laws and regulations have not kept pace with the rapid development of synthetic media technologies. By advocating for this integrated framework, the review not only offers a technological solution but also contributes to the policy discourse on digital privacy, consent, and media ethics.

ISSN: 2366-1313

II. LITERATURE SURVEY

The rapid advancement of deepfake generation technolo-gies has motivated extensive research into developing ro-bust detection methods. Early approaches primarily relied on analyzing visual artifacts introduced during the deepfake creation process. For example, Matern et al. (2019) proposed detecting deepfakes by identifying visual inconsistencies such as unnatural eye blinking and irregular facial movements. Similarly, Li and Lyu (2018) developed a technique focused on the frequency of eye blinking, as early deepfake models often failed to replicate realistic blinking behavior. With the advent of more sophisticated generative models, such as StyleGAN and FaceSwap, researchers shifted toward using deep learning-based classifiers. Notably, Afchar et al. (2018) introduced MesoNet, a lightweight convolutional neural network that achieved good performance by focusing on mesoscopic-level features rather than fine-grained pixel-level inconsistencies.

Recent works have increasingly explored the integration of spatial, temporal, and frequency domain analysis to enhance detection accuracy. Zhou et al. (2017) proposed Two-Stream Neural Networks that analyze both the spatial and temporal components of videos to capture subtle artifacts across frames. Meanwhile, Durall et al. (2020) highlighted that deepfakes leave distinct statistical fingerprints in the frequency domain, leading to new methods that analyze Fourier transforms of images. Despite these advancements, a common challenge re-mains: many detection systems are reactive, aiming to identify manipulated media after its circulation. This limitation has driven emerging research toward proactive methods, such as real-time watermarking, blockchain-based content authentication, and AI-integrated verification systems, to strengthen the trustworthiness of digital media ecosystems.



Current Systems	Proposed Framework
Detection is reactive. Identifies deepfakes post-distribution.	Focuses on prevention from content creation point using watermarking.
Lacks robust proof of own-ership.	Blockchain ledger stores media metadata for verifi-able ownership.
No control over editing or redistribution.	OTP-based permission mechanism enforces owner-approved edits.
No cross-platform verifica-tion mechanism.	AI module actively scans media for watermark pres-ence and anomalies.
Centralized and prone to data breaches.	Decentralized blockchain ensures tamper-proof and resilient storage.

TABLE I: Comparison of Current Systems vs. Proposed Framework

III. RELATED WORK

Recent advancements in deepfake detection have leveraged deep learning techniques to identify forged media with high accuracy. Several studies have explored convolutional neu-ral networks (CNNs) and recurrent neural networks (RNNs) for spotting inconsistencies in facial expressions, blinking patterns, or head movements that are common in synthetic videos. One of the notable works, MesoNet by Afchar et al., introduced a lightweight CNN-based architecture specifically designed for detecting deepfakes by analyzing mesoscopic-level features in videos. Another significant contribution, FaceForensics++, provided a large dataset for training and benchmarking detection algorithms, helping improve the per-formance of forensic analysis tools. These models, however, are often limited by their reliance on known manipulation artifacts and may struggle when faced with more sophisticated or unseen deepfake generation techniques.

To address such challenges, some researchers have incor-porated attention mechanisms and ensemble learning to im-prove model generalizability. For instance, XceptionNet-based architectures combined with frequency domain analysis have shown promising results in detecting subtle anomalies that are often missed in the spatial domain. Moreover, studies have begun integrating biometric signals, such as heartbeat detection through skin color changes, as proposed in DeepRhythm, to differentiate real from fake faces. Despite these improvements, the reactive nature of these systems highlights the need for proactive frameworks that can embed authenticity into the content creation process, which remains an underexplored but essential direction in preventing the misuse of deepfakes.

One of the primary approaches to combating deepfakes has been the development of detection algorithms. Researchers have focused on detecting inconsistencies in deepfakes by analyzing various artifacts introduced during the manipulation process. These methods can broadly be divided into two categories: image-based and video-based deepfake detection.

Image-based detection typically involves identifying irreg-ularities in facial features, lighting, or inconsistencies in pixel patterns. Tools like XceptionNet (Chesney and Citron, 2019) and other deep learning-based models have been shown to effectively identify manipulated images by examining deep facial features and analyzing temporal changes in lighting or skin tone. Such models can spot anomalies by focusing on the high-frequency components of the image, which are often

ISSN: 2366-1313

disrupted during the deepfake generation process.

Video-based detection includes techniques like facial ex-pression analysis and frame-by-frame scrutiny, often using re-current neural networks (RNNs) or 3D convolutional networks. Several papers, including those by Matern et al. (2020), have employed such techniques to track and detect inconsistencies in facial movements or gaze, which are typically altered when creating deepfakes.

Although these detection methods show promise, they are reactive, identifying manipulated content only after it has been created and distributed. This highlights a critical gap in their research, as proactive mechanisms to protect content from manipulation before it occurs are still lacking.

IV. PROPOSED MODEL

In this research, we propose a novel framework that focuses on the prevention of deepfake creation and dissemination rather than only detection after distribution. Our model in-troduces a real-time digital watermarking mechanism at the point of content creation. Each image or video captured is embedded with a unique, invisible watermark linked to the creator's credentials and a blockchain ledger. This watermark serves as an indelible proof of ownership and ensures that any tampering or unauthorized editing can be easily identified.

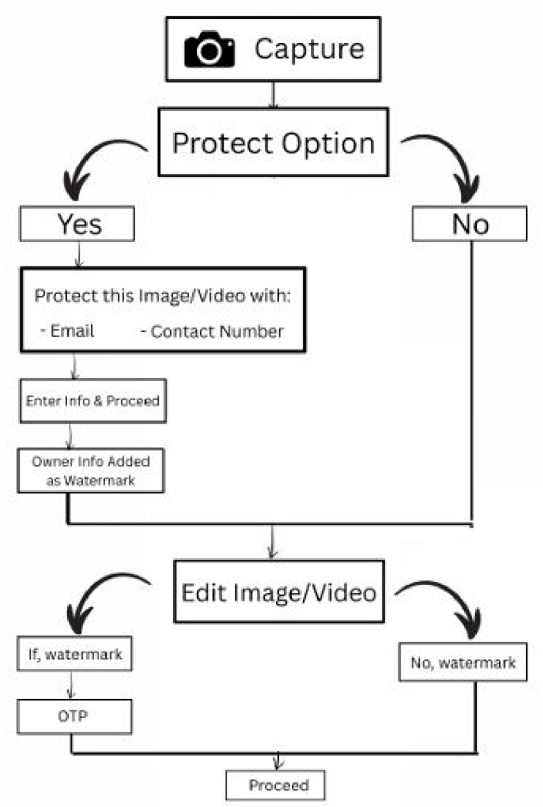


Fig. 1: Architecture for Deepfake Prevention Using Blockchain and AI for Visual Media

Furthermore, metadata associated with the content, such as timestamp, device ID, and creator details, is securely stored on a decentralized blockchain network, ensuring tamper-proof and verifiable ownership records.

To enhance content security further, we integrate an OTP-based (One-Time Password) permission mechanism that al-lows only the original owner to authorize edits or modifi-cations. Before any significant alteration is applied to the media, an OTP request is sent to the registered owner, en-suring owner-approved editing. Additionally, an AI-powered verification module continuously scans media for watermark presence and anomalies across various platforms, thereby enabling cross-platform authentication. By combining real-time watermarking, blockchain-backed ownership proof, and AI-driven anomaly detection, our proposed model offers a proactive, decentralized, and robust solution to counter the growing threat of deepfakes.

A. Watermarking for Media Authentication:

Watermarking is another commonly used technique for dig-ital content protection, including against deepfakes. It involves embedding a distinctive identifier within the content, allowing the owner or creator to prove authenticity and ownership. Tra-ditional watermarking methods, however, have been ineffective in preventing deepfakes since many image manipulations, such as those involving GANs, can erase or modify watermarks undetectably.

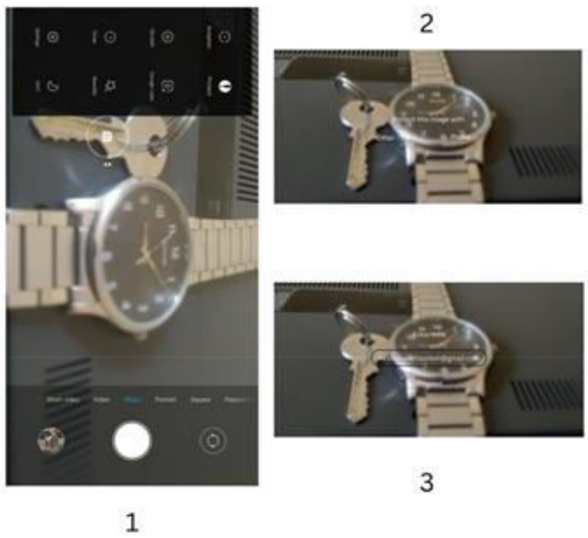


Fig. 2: First set of images consisting of the flow of the Proposed Model

A more advanced approach involves robust watermarking, which ensures that the watermark is difficult to remove even after the content has been altered. For example, techniques such as frequency domain watermarking (Vallabha et al., 2020) embed a watermark in the frequency components of the media, making it more resilient to manipulation. However, these methods still face limitations, particularly in terms of efficiency and visibility of watermarks.

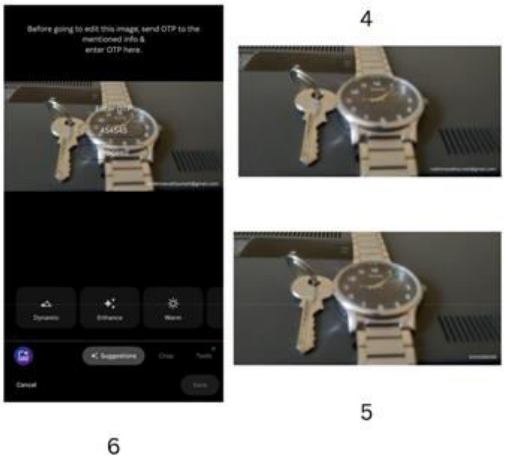
Recent advancements have proposed the use of AI to enhance watermarking. AI-powered watermarking systems can learn optimal embedding strategies that are resistant to deep-fake generation processes. For instance, a study by Liu et al. (2021) suggests using adversarial networks to optimize watermark placement to ensure that the watermark survives deepfake manipulation, even in GAN-generated content. How-ever, these techniques are still in their early stages, and more research is needed to ensure robustness against the latest deepfake technologies.

B. Blockchain for Digital Content Protection:

Blockchain technology has emerged as a powerful tool for ensuring the integrity of digital media and preventing tampering. By providing a decentralized, immutable ledger, blockchain enables transparent tracking and verification of ownership and modification histories. Several studies have explored the integration of blockchain to authenticate digital media and ensure its integrity.

A notable example is the work by Wang et al. (2020), who propose using blockchain to store media metadata such as creation time, owner information, and any changes made to the media. This ensures that every modification made to an image or video is logged as a transaction, and the original content can be verified through blockchain's tamper-proof record. Such systems are increasingly being adopted in areas like digital art and copyright protection, where provenance and ownership are crucial.

Additionally, blockchain has been integrated with water-marking to create a dual-layer security system. For example, systems like MediaChain (2017) combine digital watermarking with blockchain to link media content with verified metadata stored in a decentralized ledger. This hybrid approach not only ensures proof of ownership but also provides an immutable record of media changes, further reducing the risk of deepfake misuse.



Before going to edit this image, send OTP to the mentioned info & enter OTP here.

Fig. 3: Second set of images consisting of the flow of the Proposed Model

C. AI and Blockchain Integration:

ISSN: 2366-1313

The integration of AI with blockchain has been a recent trend aimed at providing more dynamic and responsive so-lutions for media protection. AI can analyze digital contentin real-time, identifying whether it has been tampered with or is a deepfake, while blockchain ensures the authenticity of the media through immutable records. Some studies, such as those by Zhang et al. (2021), explore using AI for real-time detection of media manipulation, while blockchain can be used to store AI decisions and metadata securely, providing a verifiable audit trail of all actions performed on the content.

AI-powered systems are also being employed to identify images that have been altered by deepfake technologies. These systems can scan content for signs of manipulation, flagging suspicious content and ensuring that only authentic, verified media is shared. By combining these AI-driven approaches with the decentralized transparency provided by blockchain, it is possible to create a more robust, proactive solution to deepfake detection and prevention.

V. EXPERIMENTAL FRAMEWORK

The proposed framework presents an innovative, multilay-ered security model that proactively safeguards digital images and videos from deepfake misuse. By integrating real-time watermarking, blockchain-based media registration, and AI-powered verification, the system not only deters unauthorized tampering but also offers a transparent and traceable approach to digital content management. To assess the feasibility and effectiveness of this concept, a detailed experimental prototype was conceptualized and developed, simulating the end-to-end workflow from media capture to secure verification and controlled modification.

The process begins at the point of content generation. A custom-built camera module or a secure mobile application initiates the watermarking process immediately after a photo or video is captured. The embedded watermark, which is designed to be imperceptible to the human eye, encodes critical metadata such as the identity of the creator, timestamp, GPS coordinates, and device ID. Advanced watermarking tech-niques such as Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT) are utilized to ensure robustness, making the watermark resilient to common transformations such as cropping, compression, format conversion, and re-sizing—conditions frequently encountered in online sharing environments.

Once the watermark is embedded, the system automatically generates a cryptographic hash of the original media file. This hash, along with the associated metadata, is uploaded and stored on a blockchain ledger. A private Ethereum blockchain was chosen for the prototype implementation due to its cus-tomizable smart contracts and secure, decentralized nature. These smart contracts automate key functions, including media registration, permission handling, content integrity verification, and edit approval mechanisms, eliminating human error and ensuring transparency.

One of the hallmark features of this framework is its OTP-based permission and modification control. When a third party attempts to edit, share, or reuse a registered media file, the system initiates a transaction that triggers a smart contract. The original owner is notified and must verify the request using a one-time password (OTP), sent securely through a registered channel (e.g., email, SMS, or secure app notification). Only after successful verification is the requested action approved, and the new state of the content is recorded on the blockchain. This process not only ensures ownership control but also builds an auditable trail of every interaction with the media, enhancing traceability and legal defensibility.

To augment the system's proactive defense capabilities, an AI-based content verification engine is deployed. This module utilizes convolutional neural networks (CNNs) and deep learning models trained on datasets of authentic and manipulated media. It constantly scans newly uploaded or shared media across supported platforms, checking for embedded watermarks and analyzing pixel-level inconsistencies indicative of AI-generated alterations. If the content lacks a valid watermark or exhibits signs of tampering, it is flagged for review or automatically quarantined. This helps maintain a secure and trustworthy digital media ecosystem.

VI. CONCLUSION AND FUTURE WORK

In this paper, we introduced a proactive and integrated framework aimed at combating the increasingly complex and harmful phenomenon of deepfakes. By synergizing three crit-ical technologies real-time digital watermarking, blockchain-based metadata storage, and AI-powered media analysis we propose a system that not only detects manipulated content but actively prevents it from spreading in the first place. This represents a significant paradigm shift from traditional approaches that are predominantly reactive in nature, often identifying deepfakes only after their distribution, when dam-age may already have occurred.

Our framework is built on the principle of “protection at the point of creation.” As soon as a photo or video is captured especially in contexts involving children, celebrities, or sensitive personal moments it is embedded with an invisible, tamper-proof watermark. This watermark encodes crucial metadata such as the content owner’s identity, timestamp, and location (if permitted). This data is not merely stored locally but is hashed and immutably recorded on a private blockchain. This blockchain entry functions as a digital certificate of authenticity, enabling any future verification of ownership or tampering.

To reinforce content control, the system employs a one-time password (OTP)-based permission mechanism for any modification or redistribution of protected media. Only users who can verify themselves through pre-registered contact information can alter the media. All such requests and approvals are logged via smart contracts, ensuring transparency and accountability. An AI-based engine supports this framework by scanning media for the presence of valid watermarks and detecting anomalies that may signal deepfake alterations. This layered, multifaceted approach not only flags unauthorized media but also significantly raises the technical and procedural barriers for creating and spreading fake content.

The real-world relevance of this solution cannot be over-stated. In an age where “sharenting” the act of parents sharing images and videos of their children online has become widespread, children are becoming unintentional victims of identity theft, misinformation, and synthetic media manipu-lation. Campaigns like Deutsche Telekom’s deepfake public service video highlight the urgency of this issue. Our frame-work, by design, is tailored to protect such vulnerable digital content from unauthorized use and exploitation.

From a broader perspective, this system paves the way for a more secure, transparent, and trustworthy digital ecosystem. It fosters digital accountability by embedding authenticity into the content itself, rather than relying solely on external moderation or post-distribution detection.

Limitations of Our System: While watermarking and AI-based protection systems present significant advantages in se-curing digital media, they also face certain inherent limitations. Watermarking techniques, although useful for embedding own-ership information, can sometimes suffer from robustness issues, particularly when the media undergoes transformations such as compression, resizing, or editing. Similarly, AI-based verification systems, though capable of detecting anomalies and tampering attempts, often require substantial computational resources, making real-time verification a challenge in scenarios with large volumes of data or limited processing capabilities. On the other hand, blockchain-based solutions, renowned for their tamper-resistant and decentralized nature, encounter scalability concerns when managing and validating massive amounts of media transactions, potentially leading to network congestion and increased operational costs.

Given these challenges, it is evident that the mere integration of blockchain, AI, and watermarking, while highly promising, is not without its complexities. There is a pressing need to refine and optimize these technologies to ensure they can work together effectively at scale. A robust, proactive, and multi-layered defense framework is essential—one that not only secures content at the point of creation but also continuously monitors its integrity across platforms where digital media is widely shared and modified. Future research must focus on enhancing the efficiency, scalability, and adaptability of such integrated systems to truly combat the evolving threats posed by deepfakes and other forms of digital forgery.

Advanced Steganographic Techniques: Future iterations can implement state-of-the-art steganographic algorithms to hide owner-identifying information such as phone numbers or email addresses within the media in a way that is undetectable during regular viewing but becomes critical during access control verification. This ensures privacy while enabling secure ownership validation.

Context-Aware Access Control: Intelligent AI models can be developed to dynamically assess the context in which a modification request is made. For instance, editing requests originating from previously unknown devices or IP addresses may be subject to stricter verification than those from trusted environments.

Biometric and Decentralized Identity (DID) Integration: One exciting avenue for advancement is the incorporation of biometric-linked ownership verification. This would bind media files to the biological identity of the created using facial recognition, fingerprint, or retinal scans, adding an additional layer of non-repudiable proof of ownership. By incorporating this verification mechanism could be particularly valuable in protecting personal media of children, influencers, journalists, or public figures. This strengthens the user’s control over their content while adhering to privacy-by-design principles.

Privacy-Preserving Verification Messages: When unau-thorized users attempt to access or edit protected media, the system can display secure prompts such as “This media is protected. Only the owner or someone with verified credentials may modify this file.” This deters tampering without revealing any sensitive data embedded in the watermark.

Social Media and Platform-Level Integration: Strategic collaboration with popular social media platforms, content hosting services, and cloud storage providers can ensure that only verified and protected content is allowed to be uploaded or distributed. This would drastically reduce the prevalence of unauthorized media and provide end-to-end security from creation to consumption.

Scalable Blockchain Solutions: Scalability is a key consid-eration for real-world adoption. To accommodate higher vol-umes of content and lower transaction latency, future versions of this system can explore Layer-2 blockchain technologies, such as optimistic rollups or zk-rollups or sidechains (like Polygon or Arbitrum), to handle metadata registration and verification more efficiently. The current implementation on a private Ethereum blockchain offers strong security and flexibility, but to support millions of users and media files, transitioning to Layer 2 solutions will be critical. These solutions reduce transaction costs, increase throughput, and maintain blockchain integrity, making large-scale deployment more feasible.

User-Friendly Interfaces and Adoption Campaigns: A successful framework must also prioritize usability. Intuitive mobile applications, browser plugins, or integration into ex-isting camera apps can help non-technical users adopt this technology easily. Public education and awareness campaigns can promote the responsible use of such tools and the dangers of unprotected digital sharing.

Looking ahead, the given future works can drive further research and development to enhance the effectiveness of this proposed model.

REFERENCES

[1] L. Verdoliva, “Media Forensics and DeepFakes: An Overview,” IEEE J. Sel. Topics Signal Process., vol. 14, no. 5, pp. 910–932, 2020.

[2] Y. Mirsky and W. Lee, “The Creation and Detection of Deepfakes: A Survey,” ACM Comput. Surv., vol. 54, no. 1, pp. 1–41, 2021.

[3] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, et al., "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," Inf. Fusion, vol. 64, pp. 131–148, 2020.

[4] N. F. Johnson, Z. Duric, and S. Jajodia, "Information Hiding: Steganog-raphy and Watermarking—Attacks and Countermeasures," Springer, 2001.

[5] J. Fridrich, "Steganography in Digital Media: Principles, Algorithms, and Applications," Cambridge University Press, 2009.

[6] X. Liu, Y. Sun, L. Xu, et al., "Digital Image Watermarking and Steganography: Fundamentals and Techniques," Springer, 2019.

[7] M. Barni and F. Bartolini, "Watermarking Systems Engineering: En-abling Digital Assets Security and Other Applications," Marcel Dekker Inc., 2001.

[8] Y. Zhao, Y. Zhang, and H. Lu, "Deep Learning Based Steganalysis for Image," IEEE Access, vol. 6, pp. 28004–28014, 2018.

[9] P. K. Sharma and J. H. Park, "Blockchain based Hybrid Framework for Secure Data Storage in Internet of Things," IEEE Trans. Ind. Informat., vol. 15, no. 6, pp. 3160–3168, 2019.

[10] S. Wang, J. Wang, Y. Yuan, et al., "A Survey on Blockchain for Big Data: Approaches, Opportunities, and Future Directions," Future Gener. Comput. Syst., vol. 100, pp. 564–575, 2019.

[11] Z. Zheng, S. Xie, H. Dai, et al., "An Overview of Blockchain Tech-nology: Architecture, Consensus, and Future Trends," Proc. IEEE Int. Conf. Big Data (BigData), pp. 557–564, 2017.

[12] S. Singh and N. Singh, "Blockchain: Future of Financial and Cyber Security," Procedia Comput. Sci., vol. 132, pp. 90–97, 2018.

[13] J. C. Westland, "Deepfake Technology: The Rise of Synthetic Media and Its Impact on Trust," Bus. Horiz., vol. 63, no. 4, pp. 509–519, 2020.

[14] D. Chai, H. Sun, and Y. Jiang, "Secure Image Watermarking Based on Deep Neural Networks," IEEE Access, vol. 8, pp. 92020–92030, 2020.

[15] A. Rossler,¨ D. Cozzolino, L. Verdoliva, et al., "FaceForensics++: Learning to Detect Manipulated Facial Images," Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 1–11, 2019.

[16] N. Yu, L. S. Davis, and M. Fritz, "Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints," Proc. IEEE Int. Conf.

ISSN: 2366-1313

Comput. Vis. (ICCV), pp. 7556–7566, 2019.

[17] S. Agarwal, T. El-Gaaly, H. Farid, and S. Lim, "Protecting World Leaders Against Deep Fakes," Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops, pp. 38–45, 2019.

[18] D. Guera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS), pp. 1–6, 2018.

[19] T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep Learning for Deepfakes Creation and Detection: A Survey," arXiv preprint arXiv:1909.11573, 2019.

[20] Y. Li, M.-C. Chang, and S. Lyu, "Exposing DeepFake Videos by Detecting Face Warping Artifacts," arXiv preprint arXiv:1811.00656, 2018.

[21] H. R. Hasan and K. Salah, "Combating Deepfake Videos Us-ing Blockchain and Smart Contracts," IEEE Access, vol. 9, pp. 61174–61193, 2021.

[22] Y. Fu, H. Li, and R. Zhang, "A Blockchain-Based Digital Forensics Investigation Framework for Deepfake Video Detection," Proc. Int. Conf. Cyberworlds (CW), pp. 272–279, 2020.

[23] Z. Zhang, J. Zhao, and H. Wang, "Blockchain-Based Steganography: Toward Secure and Decentralized Data Hiding," IEEE Access, vol. 9, pp. 16621–16631, 2021.

[24] Y. Ding, X. Luo, and Q. Li, "AI-Powered Watermarking for Multimedia Security: State of the Art and Future Directions," ACM Trans. Multimed. Comput. Commun. Appl. (TOMM), vol. 17, no. 1s, pp. 1–25, 2021.

[25] A. Goel, R. Puri, and S. Aggarwal, "A Comprehensive Survey on Deep-fake Detection Techniques: AI Approaches, Datasets, and Challenges," Multimed. Tools Appl., vol. 82, pp. 12503–12538, 2023.

[26] L. Verdoliva, "Media Forensics and DeepFakes: An Overview," IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910–932, 2020.

[27] H. Kim, J. Park, and S. Lee, "Blockchain-Based Secure Image Prove-nance System for Deepfake Prevention," Future Generation Computer Systems, vol. 123, pp. 192–204, 2021.