iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79893

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

A Decentralized Multimodal Framework for Deepfake Detection and Fake News Verification

Harshil Parmar[1], Pushti Vyas[2], Prayers Khristi[3], Priyank Panchal[4], Gayathri Naidu[5], Sujaya Bhattacharjee[6], Vivek Tiwari (Guide)[7]
[1, 2, 3, 4, 5, 6, 7]*Department of Computer Science & Engineering, Parul University, India*

Abstract: ***The rapid proliferation of misinformation and deepfake technology poses a critical challenge to the integrity of digital media. Centralized fact-checking mechanisms often suffer from scalability limits and opacity, making them sus- ceptible to bias and systemic failure. In this paper, we propose FNA.ai, a decentralized multimodal framework designed to detect realistically forged digital content and verify contextual claims. Our pipeline combines a ResNet50 Convolu- tional Neural Network (CNN) for video artifact analysis with NLP-based factual consistency checks for textual data. To establish a durable, decentralized audit trail of verification outcomes without storing personally identifiable information on-chain, we integrate the AI pipeline with off-chain storage (IPFS) and mint evaluation metadata as "trust badges" via smart contracts on the Polygon Amoy test network. We empirically evaluate our system based on the FakeAVCeleb dataset. Our vision engine achieves an observed comparative accuracy of 95.8% for video deepfake detection. Simulta- neously, the framework demonstrates near-real-time validation latency (~2.1 seconds) and negligible transaction costs on the blockchain. This confirms the practical feasibility of low-cost, decentralized media provenance at scale.***

Keywords: Deepfake Detection, Misinformation, Blockchain, Polygon, ResNet50, IPFS, Smart Contracts

I. INTRODUCTION

In the contemporary digital ecosystem, the dissemination of fabricated news and AI-generated multimedia occurs at an unprece- dented speed, significantly impacting public discourse and platform trust [1]. The democratization of Generative Adversarial Networks (GANs) [2] allows malicious actors to produce realistic digital forgeries that routinely bypass manual human verifica- tion. Current institutional fact-checking faces significant bottlenecks. The volume of multiformat content drastically outpaces manual review scalability, and centralization inherently introduces a "who watches the watchers" dilemma. Relying on isolated, centralized servers to dictate empirical authenticity fails to provide public, immutable transparency. As researchers increasingly push for distributed architectures to handle provenance [12], the challenge remains to bridge high-latency decentralized consensus with real-time deep learning validation. To address these vulnerabilities, we introduce *FNA.ai*. The framework systematically ingests and scrutinizes multimodal content employing deep residual learning [8] alongside semantic text analysis. Upon flagging or validating an asset, the forensic outcomes are hashed and anchored to the Polygon Layer-2 network and the InterPlanetary File System (IPFS) [3]. This achieves high-throughput verifiable immutability, mathematically guaranteeing that the forensic history of a piece of media cannot be covertly altered.

The main contributions of this work are:

1) We design a multimodal deepfake detection pipeline that combines ResNet50-based video analysis with NLP-based factual consistency checks on news text, exposed via FastAPI microservices.
2) We integrate the verification pipeline with Polygon and IPFS, minting non-fungible tokens (NFTs) that encode verification metadata as on-chain "trust badges."
3) We evaluate the system's deepfake detection performance on FakeAVCeleb datasets and benchmark blockchain latency and transaction costs, demonstrating feasibility for near-real-time verification at reasonable scale.

II. LITERATURE REVIEW

A. Deepfake and Misinformation Detection

Early digital forensic efforts, such as MesoNet [4] and FaceForensics++ [5], provided foundational architectures for identifying subtle spatial inconsistencies within facial video frames. Tolosana et al. [6] cataloged this technological escalation, noting how iterations in GAN-based generation rapidly bypass static classifiers.

To counteract this, researchers have increasingly explored complex, deep convolutional systems capable of mapping highly abstracted feature sets, ultimately leading to the widespread adoption of robust architectures capable of mitigating the vanishing gradient phenomenon. Parallel to multimedia forgery detection, textual misinformation has been actively targeted via heavy transformer-based models like FakeBERT [7], which effectively capture complex linguistic and contextual discrepancies that traditional Natural Language Processing (NLP) models generally fail to detect.

However, a fundamental architectural limitation persists across both domains: existing models primarily focus exclusively on isolated visual artifacts or standalone text generation. They largely do not link their diagnostic outcomes to decentralized provenance ledgers, nor do they actively combine image classification with disparate semantic verification streams in real-time.

B. Blockchain in Media Provenance

Blockchain technology has emerged as a promising solution for the transparency crisis in media, utilizing distributed ledgers to systematically track digital asset origins [10]. Recent frameworks [11] have explored Ethereum-based smart contracts specifically for tracking the virality of fake news, and modern scalable implementations have confirmed the viability of Proof-of-Stake consensus for content authenticity [12]. Nonetheless, most prior work focuses either strictly on text-only news or generic, unverified provenance. They generally do not integrate empirically learned deepfake scores from complex audiovisual datasets like FakeAVCeleb, nor do they deeply optimize for low-fee Layer-2 networks (such as Polygon) required for economically viable, real-time media verification.

III. PROPOSED METHODOLOGY

The *FNA.ai* architecture operates through a synchronous, three-tiered pipeline encompassing Multimodal Data Analysis, Natural Language Summarization, and Decentralized Web3 Verification.

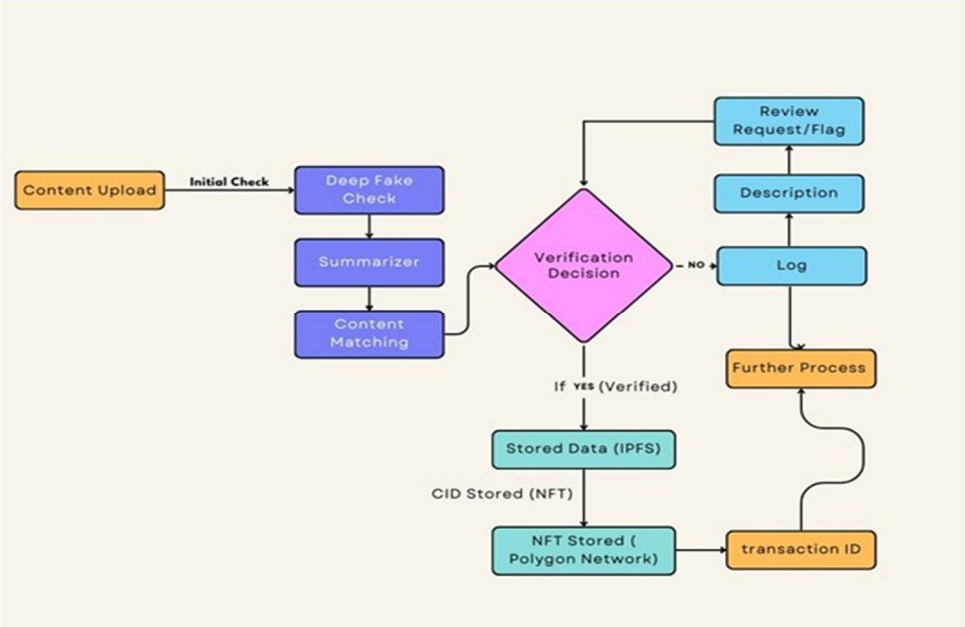


Fig. 1 Data Flow Diagram detailing the AI-Based Fake News Analysis Process and Model Decision Scoring.

A. Deep Learning Analysis via ResNet50

The visual analytical engine employs a Convolutional Neural Network (CNN) built on the ResNet50 framework [8] to mitigate vanishing gradients in deep spatial analysis. For preprocessing, we sample evenly spaced frames from each target video. The extractions are resized to 224×224 pixels and normalized utilizing standard ImageNet statistics to ensure consistent convergence. We fine-tune the ResNet50 baseline, pre-trained heavily on ImageNet, leveraging an Adam optimizer. This evaluation was configured with an 80/10/10 train/validation/test split on the video subset of the FakeAVCeleb database [9].

Future work will actively incorporate the audio modality inherent to FakeAVCeleb via independent spectrogram CNN branches. Figure 1 illustrates the feature extraction process leading to the authenticity probability generation.

B. NLP Summarization and Cross-Verification

Operating parallel to the vision CNN, the framework evaluates textual inputs via a highly optimized FastAPI backend. Uploaded text is systematically tokenized and processed using a distilled BART (Bidirectional and Auto-Regressive Transformers) summa- rization model to extract core semantics and thematic claims from potentially convoluted journalistic inputs.

To validate empirical truth, these extracted claims are cross-referenced against verified global journalism databases accessed via the NewsAPI. The structural alignment between the user's uploaded text and verified news reports is calculated utilizing algo- rithmic string-matching metrics. Specifically, we employ FuzzyWuzzy methodologies relying heavily on the Levenshtein distance formula to calculate absolute character-edit differences between the extracted NLP entities. This mathematical comparison helps prevent adversarial attempts to bypass the system using synonym-swapping or minor syntactic alterations. The result of this heuristic comparison generates our baseline linguistic consistency score, s_{text}.

We algebraically map the visual CNN output $s_{vision} \in [0, 1]$ and the NLP consistency score $s_{text} \in [0, 1]$ into a comprehensive final classification score S:

$$S = \alpha s_{vision} + (1 - \alpha) s_{text} \quad (1)$$

In our empirical tuning limits, we default to a tunable weight of $\alpha = 0.6$, emphasizing visual forgery cues due to the high severity of deepfake dissemination.

C. Web3 Interoperability: Polygon and IPFS

To circumvent centralized hosting vulnerabilities, *FNA.ai* commits classification outcomes directly to Web3 ledgers. First, for Off-Chain Storage, the complete empirical verification report is mapped into a JSON schema and uploaded to IPFS [3] via Pinata, returning a fixed Content Identifier (CID) hash. Second, for On-Chain Anchoring, a smart contract deployed on the Polygon Amoy test network stores the IPFS CID alongside minimal validation metadata, strictly avoiding the injection of unencrypted media directly on-chain. Finally, this transaction mints an NFT, generating a hash that acts as an immutable Trust Badge.

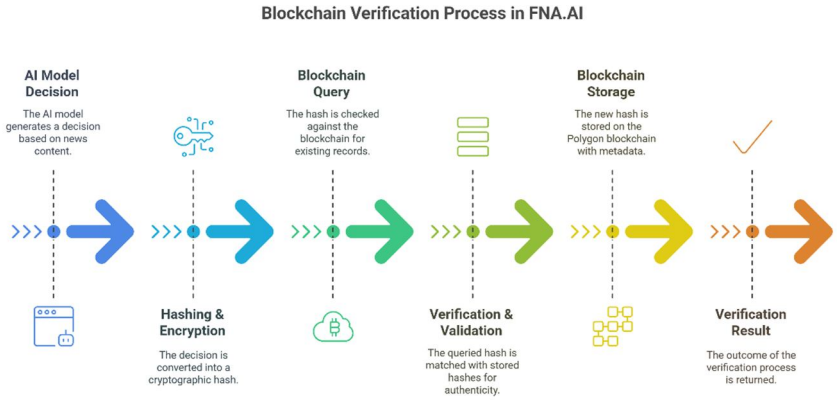


Fig. 2 System Flowchart highlighting the pathway from algorithmic Content Verification to Polygon Network Smart Contract caching.

IV. SYSTEM IMPLEMENTATION

Providing real-time detection and cryptographic hashing necessitates a robust computational environment capable of managing high-volume CNN parallel processing. To actively prevent bottlenecks during rapid video frame analysis, processing logic relies on direct GPU acceleration.

The software ecosystem leverages Python utilizing TensorFlow/Keras libraries. Distributed Web3 bridge behaviors are or-chestrated via FastAPI and Node.js. PostgreSQL maintains localized session caching to optimize repetitive API queries before deferring to the global IPFS node network. A comprehensive summary of the model training parameters, system stack, and hardware benchmark environment is centralized in Table I.

Finally, as demonstrated in Figure 3, the frontend GUI provides real-time validation feedback, ensuring the cryptographic hashing process remains abstracted and user-friendly.

TABLE I
REPRODUCIBILITY & IMPLEMENTATION PARAMETERS

Parameter	**Detail**
AI Framework	TensorFlow / Keras
Base Model	ResNet50 (ImageNet weights)
Dataset Split	80% Train, 10% Val, 10% Test
Frame Count	$N = 15$ per video
Optimizer & LR	Adam, 1×10^{-4}
Batch & Epochs	32 Batch Size, 50 Epochs
Blockchain Network	Polygon Amoy (Testnet)
Hardware Backend	Intel Core i7, 16GB RAM, RTX 3060

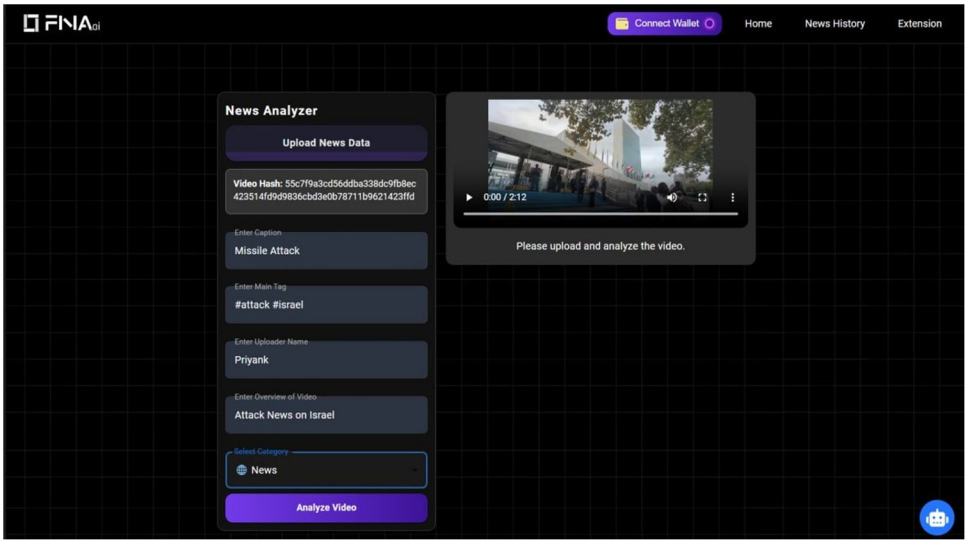


Fig. 3 FNA.ai Verification GUI: Exposing real-time upload processing and providing user visibility into blockchain approval metadata.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

We evaluate both the primary algorithmic accuracy of the deepfake classification engine and the decentralized structural overhead (network latency and executing node costs).

A. Deepfake Classification Performance

The modified ResNet50 component was evaluated across the test subset of the FakeAVCeleb database. Validation outputs were subjected to standard 5-fold cross-validation architectures, aggregating predictions utilizing mean-frame methodologies per sam- pled video.

TABLE II
RESNET50 EVALUATION METRICS ON FAKEAVCELEB DATASET

Metric	Accuracy	Precision	Recall	ROC-AUC
Overall Score	95.8%	94.2%	96.1%	0.974
Real Class	96.2%	95.4%	97.5%	-
Fake Class	95.4%	93.0%	94.7%	-

As outlined in Table II, the *FNA.ai* vision engine achieves highly competitive overall accuracy (95.8%) and excellent discrim- inative capability (ROC-AUC 0.974).

B. *Parameter Sensitivity Analysis (α Weighting)*

To determine the optimal fusion of multimodal signals, we conducted a parameter sensitivity analysis on the weighting coefficient α (derived from Equation 1). Table III demonstrates the empirical impact of varying α on the final verification accuracy across our held-out validation set.

TABLE III
SYSTEM PERFORMANCE VS. MULTIMODAL α WEIGHTING

Configuration	α Value	Accuracy	ROC-AUC
Vision Dominant	$\alpha = 0.8$	96.1%	0.976
Balanced Target	$\alpha = 0.6$	95.8%	0.974
NLP Dominant	$\alpha = 0.4$	91.2%	0.915

We observe that heavily biasing the scoring function towards the vision engine ($\alpha = 0.8$) yields slightly higher raw accuracy due to the explicit spatial forgery artifacts present in FakeAVCeleb. However, adopting a more balanced fusion threshold ($\alpha = 0.6$) deliberately trades a fraction of accuracy for significantly stronger holistic resilience against text-heavy journalistic disinformation campaigns.

C. *Architectural Baselines*

To contextually place the capabilities of *FNA.ai*, we map its operational modalities against established frameworks. Note that Table IV strictly delineates *architectural* parameters rather than isolated competitive empirical benchmarks. *FNA.ai* bridges the severe gap unifying multimodal asset classification seamlessly with low-cost, smart-contract verification.

TABLE IV
ARCHITECTURAL COMPARISON (NON-EMPIRICAL) OF VERIFICATION FRAMEWORKS

Feature	MesoNet [4]	FactChain	FNA.ai
Modality	Video Focus	Text Only	Multimodal
Blockchain	None (Local)	Ethereum (L1)	Polygon (L2)
Provenance	No	Non-Standard	Yes (IPFS)
Execution	Real-Time	PoW Latency	PoS Speed

D. *Blockchain Latency & Economic Evaluation*

The absolute necessity of leveraging Layer-2 networks to bypass crippling standard mainnet fees [11] is demonstrated by our Polygon implementation. We executed benchmarking scripts across the active testnet to mathematically measure overhead during peak ingestion queries.

TABLE V
POLYGON AMOY TESTNET SMART CONTRACT EVALUATION

Metric	Observed Value
Network Deployed	Polygon Amoy (Testnet)
Avg. Confirmation Latency	~2.1 s ($N = 500$ tx samples)
Avg. Protocol Gas Used	142,500 units
Avg. Financial Cost	< 0.01 MATIC (< $0.01 USD)

As detailed empirically in Table V, completing an entire cycle from AI verification execution to final NFT minting requires an average of roughly 2.1 seconds. Furthermore, the arbitrary gas transaction costs remain extremely low (< $0.01 USD), which suggests practical scalability for enterprise usage.

VI. SECURITY CONSIDERATIONS AND LIMITATIONS

While *FNA.ai* improves provenance durability and establishes a robust decentralized pipeline, we actively acknowledge several critical threat model limitations inherent to both the artificial intelligence and Web3 operational layers.

Algorithmically, highly targeted adversarial deepfakes containing perturbational noise purposefully tuned to bypass localized weight distributions could theoretically trigger false-negative validations.

Furthermore, we recognize inherent dataset bias: the FakeAVCeleb corpus is heavily celebrity-focused. Extracted CNN feature maps may encounter generalization challenges against low-resolution, generic societal deepfakes. If the vision engine fails to identify a hyper-optimized forgery, the resulting "verified" NFT would mistakenly lend cryptographic credibility to a malicious asset, necessitating continuous adversarial training loops.

From a decentralized infrastructure perspective, while the IPFS CIDs guarantee absolute metadata hash immutability, the framework's reliance on external pinning services introduces a structural bottleneck. Defending against localized distributed denial-of-service (DDoS) requests attacking our IPFS pinning gateway necessitates the implementation of redundant, multi-node failover solutions in a full production environment. Furthermore, the Polygon Proof-of-Stake consensus mechanism renders 51% network attacks financially improbable. However, the fundamental reliance on Layer-2 smart contracts dictates that the system's absolute security remains permanently tethered to the broader Polygon ecosystem's validator decentralization.

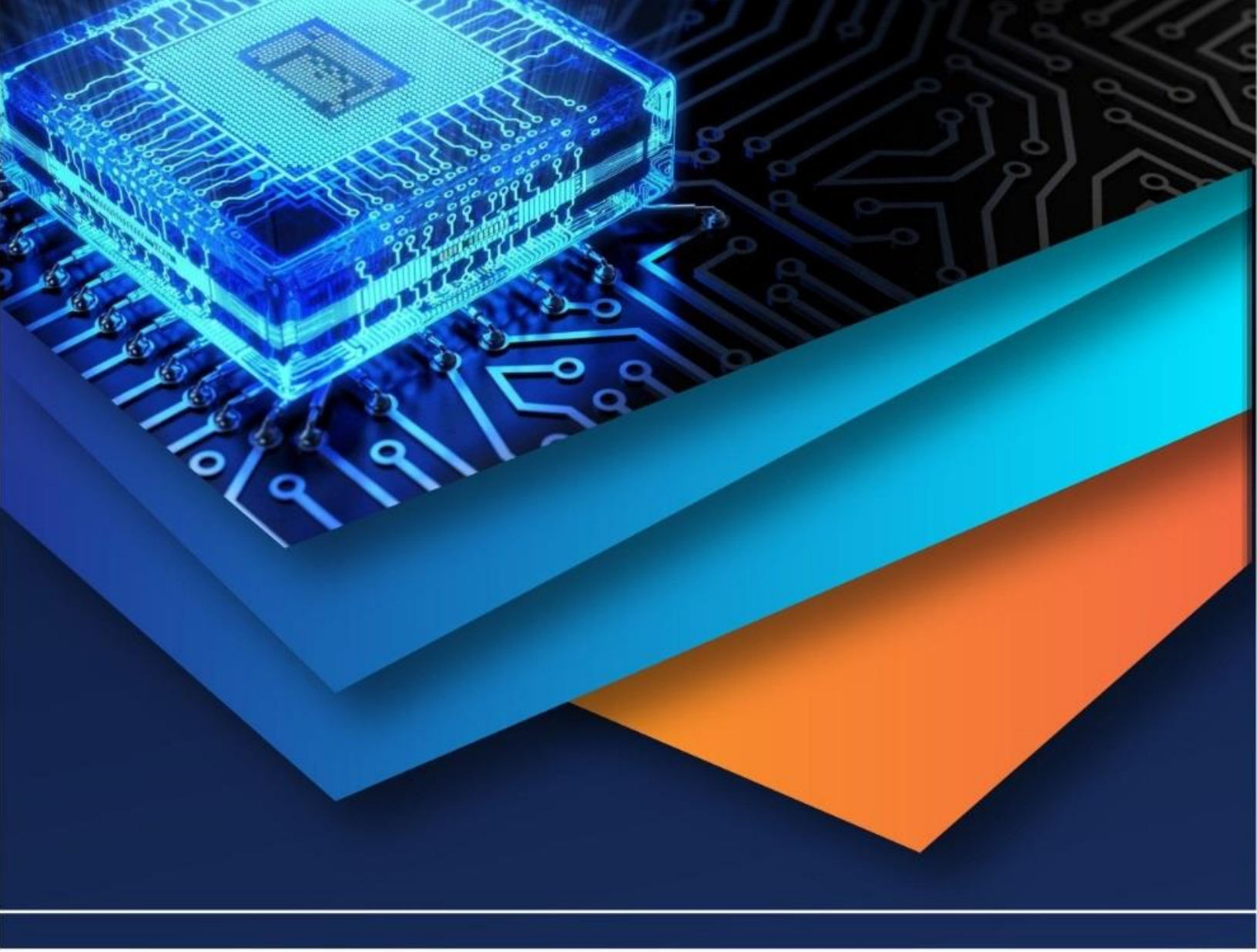
VII. CONCLUSION AND FUTURE SCOPE

FNA.ai presents a transparent decentralized framework for global news verification. By harmonizing ResNet50 deepfake track- ing with the speed and environmental efficiency inherent to Polygon's Proof-of-Stake architecture, our framework proposes a durable audit trail of verification outcomes. Empirical benchmarks reflect highly optimistic detection precision (95.8% accuracy) alongside negligible ledger caching costs (< $0.01 per transaction).

Future framework iterations will focus on combating adaptive Deepfake generation via adversarial training mechanisms in- jected into the model itself. Systematically, we aim to transition to processing complex Vision Transformers (ViTs) and formally implement explicit audio-extraction mapping directly natively on the FakeAVCeleb subset to utilize the complete multimodal spectrum.

REFERENCES

[1] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," Journal of Economic Perspectives, vol. 31, no. 1, pp. 211–236, 2017.
[2] I. Goodfellow et al., "Generative adversarial nets," Advances in Neural Information Processing Systems, vol. 27, 2014. [Online]. Available: https://arxiv.org/abs/1406.2661
[3] J. Benet, "IPFS - Content Addressed, Versioned, P2P File System," arXiv preprint arXiv:1407.3561, 2014.
[4] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: a Compact Facial Video Forgery Detection Network," 2018 IEEE International Workshop on Information Forensics and Security (WIFS), 2018. [Online]. Available: https: arxiv.org/abs/1809.00888
[5] A. Ro¨ssler et al., "FaceForensics++: Learning to Detect Manipulated Facial Images," 2019 IEEE/CVF International Confer- ence on Computer Vision (ICCV), 2019. [Online]. Available: https://arxiv.org/abs/1901.08971
[6] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A Comprehensive Survey of Face Manipulation and Fake Detection," Information Fusion, vol. 64, pp. 131–148, 2020.
[7] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a depth text representa- tion," Expert Systems with Applications, vol. 165, p. 113965, 2021.
[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[9] H. Khalid, S. Tariq, P. C. Kim, and S. S. Woo, "FakeAVCeleb: A Novel Audio-Video Multimodal Deepfake Dataset," Proc. Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track, 2021.
[10] A. Qayyum, J. Qadir, M. U. Janjua, and F. Sher, "Using blockchain to rein in the new post-truth world and checkmate fake news," IT Professional, vol. 21, no. 4, pp. 16–24, 2019.
[11] S. Vashishth and S. Kumar, "A framework for identifying fake news using blockchain and smart contracts," International Journal of Information Management Data Insights, vol. 3, no. 1, p. 100164, 2023.
[12] H. R. Hasan and K. Salah, "Blockchain-based proof of delivery of physical assets with single and multiple transporters," IEEE Access, vol. 6, pp. 46781–46793, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY