Bridging of Emerging AI and Media Broadcasting (BEAM)
Vol. 2, No. 1, November, 2026, pp. 47−59
P-ISSN: 3163-8015 | E-ISSN: 3163-8031

Trust Adaptive Provenance Watermarking for Synthetic News Broadcasting

Sabil Maulana Fauzi[1], Muhammad Syaeful Yusuf[2], Qolbinisa[3*]
[1,2,3]Faculty Economics and Business, Universitas Raharja , Indonesia
[1]sabilmaulana@raharja.info , [2]syaeful@raharja.info, [3]qolbinisa@raharja.info
***Corresponding Author**

Article Info

Article history:

Submission May 27, 2026
Revised July 30, 2026
Accepted September 12, 2026
Published November 12, 2026

Keywords:

Adaptive Provenance
Artificial Intelligence
Digital Watermarking
News Broadcasting
Watermarking

ABSTRACT

The rapid adoption of generative artificial intelligence in digital news broadcasting has increased the production of synthetic news content, raising critical concerns regarding authenticity, source verification, content manipulation, and audience trust. **This study proposes** a trust-adaptive provenance watermarking framework designed to dynamically protect and verify AI-generated news content by integrating provenance information with trust-aware watermarking mechanisms throughout the broadcasting workflow. **The proposed framework** combines adaptive digital watermarking, provenance metadata, and a trust-scoring mechanism that adjusts watermark strength according to content sensitivity, source credibility, and manipulation risk. Experimental evaluation is conducted using synthetic news samples under common content transformations, including compression, resizing, cropping, re-encoding, and partial manipulation, while performance is assessed through watermark detection accuracy, robustness, content quality, and provenance verification reliability. **The experimental findings** demonstrate that the proposed approach achieves high watermark detectability while maintaining acceptable perceptual quality across different broadcasting transformations. The adaptive trust mechanism further improves provenance verification for high-risk synthetic content and enables more reliable identification of manipulated or unauthorized news assets compared with conventional fixed-strength watermarking approaches. **The proposed** trust-adaptive provenance watermarking framework provides a practical approach for strengthening authenticity, traceability, and accountability in synthetic news broadcasting, offering a foundation for trustworthy AI-enabled media ecosystems while supporting broadcasters, audiences, and digital platforms in mitigating misinformation and unauthorized content manipulation.

This is an open access article under the CC BY 4.0 license.

DOI : https://doi.org/10.68012/beam.v2i1.265
This is an open-access article under the CC-BY license (https://creativecommons.org/licenses/by/4.0/)
©Authors retain all copyrights

1. INTRODUCTION

The rapid development of Artificial Intelligence (AI), particularly generative AI, has transformed the production and distribution of digital news [1]. Generative models can produce realistic text, images, audio, and video, enabling automated news production, multilingual broadcasting, virtual presenters, and personalized content [2]. However, these capabilities also raise concerns regarding authenticity, provenance, misinformation, and public trust [3]. AI-generated and manipulated media can resemble genuine journalistic content, making it difficult for audiences to identify its origin or detect unauthorized modifications [4, 5]. Therefore, reliable

mechanisms for verifying digital news authenticity and integrity are increasingly important [6]. Content provenance records the origin and modification history of digital assets [7]. However, provenance metadata may be removed or altered during editing, compression, and cross-platform distribution [8]. Digital watermarking provides additional protection by embedding identification information directly into media content [9]. Nevertheless, conventional watermarking often uses fixed embedding parameters without considering source credibility, content sensitivity, or manipulation risk. This limitation is significant in synthetic news broadcasting, where different content categories may require different levels of protection [10].

Previous studies have explored watermarking, synthetic-content detection, deepfake identification, and provenance verification [11–13]. However, detection-based approaches may become less reliable as generative models evolve, while fixed-strength watermarking may fail to balance robustness and perceptual quality across different news categories [14–16]. Furthermore, watermark robustness, provenance tracking, and trust assessment are often examined separately rather than integrated into one framework. This gap highlights the need for an adaptive authentication mechanism that connects persistent content identification with provenance information while adjusting protection according to the risk associated with each news asset [17–19]. To address this gap, this study proposes a framework for synthetic news broadcasting. The framework integrates digital watermarking, provenance metadata, and adaptive trust assessment to determine appropriate protection levels based on source credibility, content sensitivity, provenance completeness, and estimated manipulation risk [20, 21]. Higher-risk content receives stronger protection, while lower-risk content can prioritize perceptual quality and computational efficiency [22]. The embedded watermark also links media assets to provenance information, supporting verification when external metadata is unavailable or modified [23]. The framework is evaluated under common broadcasting transformations, including compression, resizing, cropping, re-encoding, and partial manipulation. Evaluation considers watermark detection accuracy, robustness, perceptual quality, and provenance verification reliability [24–26].

The main contribution of this research is an adaptive authentication framework designed specifically for synthetic news broadcasting. It introduces trust level as a factor in determining watermark protection, connects embedded watermarks with provenance records to improve traceability, and evaluates robustness, perceptual quality, and verification reliability together [27–30]. The framework is intended to help broadcasters and digital platforms identify unauthorized alterations, preserve content-origin information, and strengthen accountability while maintaining media quality. This study is also relevant to Sustainable Development Goals (SDGs) 9 (Industry, Innovation and Infrastructure), 16 (Peace, Justice and Strong Institutions), and 12 (Responsible Consumption and Production). SDG 9 relates to innovation in digital authentication technology, while SDG 16 concerns transparency and accountability through provenance verification and manipulation detection. SDG 12 is connected to the responsible distribution of AI-generated content through source verification and modification tracking. Although the study does not directly measure its impact on these goals, the proposed framework has the potential to support a safer and more trustworthy digital media ecosystem.

2. LITERATURE REVIEW

2.1. Synthetic News and AI-Driven Broadcasting

Generative Artificial Intelligence (GenAI) has transformed digital journalism by supporting content generation, editing, summarization, translation, and distribution. While improving production efficiency, GenAI also raises concerns regarding transparency, accuracy, accountability, and journalistic responsibility. The increasing realism of synthetic text, images, audio, and video further complicates authenticity verification, particularly after repeated editing, compression, and redistribution [2]. Since detection methods may perform poorly against unseen manipulations, persistent authentication and provenance mechanisms are increasingly necessary.

2.2. Content Authenticity and Media Provenance

Media provenance records the origin, ownership, creation process, and modification history of digital content. Unlike detection methods based primarily on visual or statistical characteristics, provenance provides evidence of how content was created and transformed. This is important because AI-generated media is not inherently misleading and may support legitimate journalistic purposes. However, provenance metadata can be lost or altered across platforms, highlighting the need for persistent authentication signals embedded directly within media [5].

2.3. Digital Watermarking for AI-Generated Content

Digital watermarking embeds identifiable signals into digital content to support authentication and traceability. Recent approaches enable invisible watermarks in AI-generated text and visual media while preserving perceptual quality. Advanced techniques improve robustness against cropping, resizing, rotation, compression, and editing, while some approaches also support tamper localization [11]. These capabilities are particularly relevant to synthetic news, where minor modifications may significantly alter information meaning.

2.4. Research Gap and Proposed Direction

Previous studies typically address synthetic-content detection, provenance, watermarking, and manipulation identification separately. Moreover, watermark protection is commonly applied without considering differences in content trust and risk [31]. This study therefore proposes trust-adaptive provenance watermarking for synthetic news broadcasting, integrating provenance, persistent watermarking, and dynamic trust assessment. The framework is evaluated based on detection accuracy, robustness, perceptual quality, and provenance verification reliability to improve the authenticity, traceability, and accountability of synthetic news [8].

2.5. Quality of Service and Quality of Experience in Immersive Broadcasting

Immersive broadcasting requires evaluation through both Quality of Service (QoS) and Quality of Experience (QoE) [16]. QoS includes latency, throughput, packet loss, bandwidth consumption, and resource utilization, while QoE reflects perceived media quality and responsiveness. Semantic orchestration should therefore improve communication and computational efficiency without degrading perceptually important content [19]. Accordingly, this study evaluates end-to-end latency, bandwidth utilization, edge-resource efficiency, and immersive media quality as complementary performance indicators.

3. RESEARCH METHOD

3.1. Research Design

This study employed an experimental research design to develop and evaluate a Trust-Adaptive Provenance Watermarking framework for synthetic news broadcasting. The framework integrates provenance registration, trust-risk assessment, adaptive watermark embedding, broadcasting transformation simulation, watermark extraction, and provenance verification into a unified authentication process [22].

A total of 1,000 synthetic news samples were used as experimental inputs. Each asset was associated with provenance information containing its content identifier, source information, generation process, timestamp, content category, protection level, and cryptographic integrity value. Before distribution, each asset was evaluated according to four trust-related factors: Source Credibility (SC), Content Sensitivity (CS), Provenance Completeness (PC), and Manipulation Risk (MR) [26]. The resulting protection score determined the watermark strength applied to the media. The experiments were designed to answer the following research questions:

- **RQ1**: Can trust-adaptive watermarking maintain reliable watermark detection while preserving acceptable perceptual quality?
- **RQ2**: How robust is the proposed framework against common broadcasting transformations and partial manipulation?
- **RQ3**: Does trust-adaptive watermarking provide better authentication and provenance verification than conventional fixed-strength watermarking?

These questions directly reflect the main objectives stated in the abstract: watermark detectability, robustness, content quality, provenance verification reliability, and comparative improvement over fixed-strength watermarking.

3.2. Proposed Trust-Adaptive Provenance Framework

The proposed framework begins with the preparation of synthetic news content, which may include AI-generated text, images, audio, or video. Each media asset is assigned a unique Content Identifier (CID) to establish its identity and linked to a provenance record containing relevant information about its origin, generation process, and associated metadata [21]. This initial stage provides a foundation for content traceability

and verification throughout the broadcasting workflow. Subsequently, the system evaluates the protection requirements of each asset based on its trust and risk characteristics, including the potential for manipulation, the reliability of its provenance information, and the possible impact of unauthorized modifications. Based on this evaluation, the framework calculates a protection score that determines the appropriate watermark embedding level, categorized as light, moderate, or strong [24]. This adaptive mechanism enables the system to allocate protection according to the specific risk profile of each media asset while maintaining a balance between watermark robustness and media quality. The overall process of the proposed framework, including content preparation, provenance registration, risk assessment, and adaptive watermark embedding, is illustrated in Figure 1.

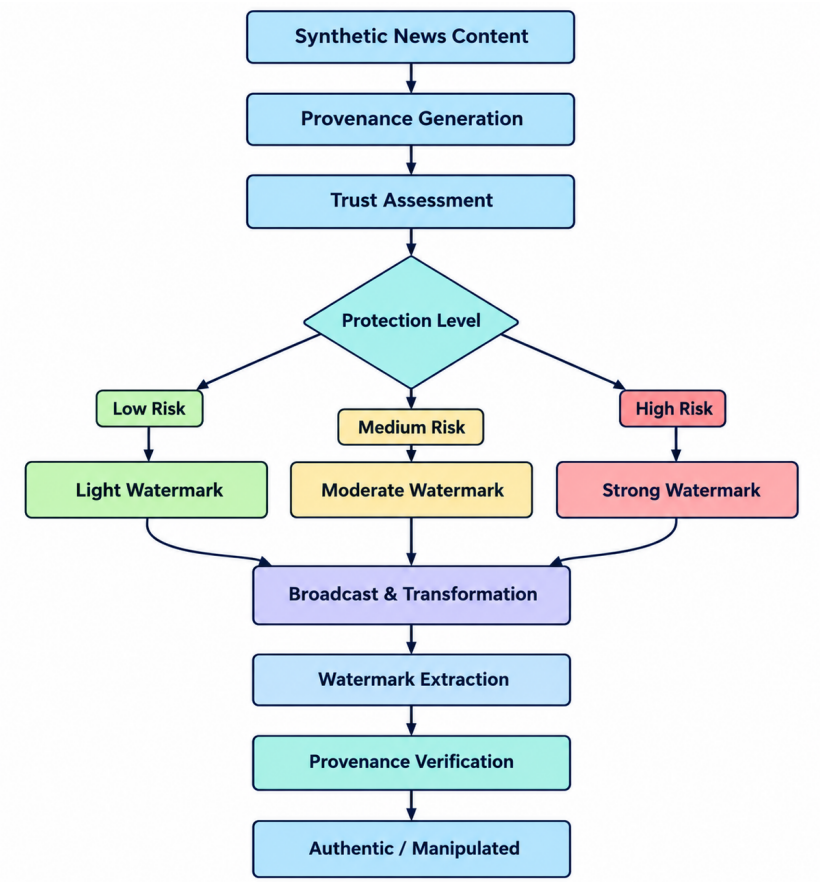


Figure 1. Trust-Adaptive Provenance Watermarking Experimental Workflow

As shown in Figure 1, each synthetic news asset is first associated with provenance information before its trust and risk characteristics are evaluated. The resulting score determines the appropriate watermark protection level. Following broadcasting transformations, the embedded identifier is extracted and matched with the provenance record to evaluate authenticity, integrity, robustness, and content quality.

3.3. Experimental Dataset

The experimental dataset consisted of 1,000 synthetic news assets distributed across five distinct broadcasting categories: politics, public safety, economy, technology, and entertainment. These categories were deliberately selected to represent diverse news contexts characterized by varying levels of content sensitivity and potential risks associated with misinformation, malicious manipulation, and unauthorized modification. To ensure a statistically robust and balanced experimental design, the dataset contained an equal number of samples for each category, specifically comprising 200 assets per category. This uniform distribution was intended to minimize category dominance, prevent class imbalance issues, and reduce potential bias during the evaluation process. Furthermore, each category was assigned a content sensitivity tier ranging from

low-medium to high determined by the potential societal and audience impact of inaccurate or manipulated information [32]. Specifically, politics and public safety were classified as high-sensitivity domains due to their direct implications for public welfare and democratic integrity; economy was categorized as medium-high, technology as medium, and entertainment as low-medium. To reflect the multimodal nature of modern digital journalism, the dataset primarily consisted of image- and text-based media assets, capturing common content formats utilized in digital news broadcasting. Ultimately, this structured and carefully curated dataset enabled a systematic, comprehensive evaluation of the proposed trust-adaptive provenance watermarking framework across varied operational contexts, offering crucial insights into how fluctuations in content sensitivity dictate protection requirements and watermark embedding strength.

Table 1. News Dataset Categories and Content Sensitivity

News Category	**Number of Samples**	**Content Sensitivity**	**Primary Media Type**
Politics	200	High	Image/Text
Public Safety	200	High	Image/Text
Economy	200	Medium–High	Image/Text
Technology	200	Medium	Image/Text
Entertainment	200	Low–Medium	Image/Text
Total	**1,000**		

Prior to watermark embedding, all media samples underwent a rigorous preprocessing and normalization pipeline to establish consistent experimental conditions and eliminate extraneous variability. Specifically, text assets were standardized in format and length, while image assets were uniformized regarding resolution and aspect ratio. As summarized in Table 1, the dataset spans five distinct broadcasting categories—politics, public safety, economy, technology, and entertainment each categorized by content sensitivity level and associated risk profile. Building upon these established sensitivity tiers, politics and public-safety content represented higher-risk contexts that demanded robust, highly secure provenance verification to mitigate potential misinformation threats. Conversely, technology and entertainment content, owing to their lower susceptibility to malicious manipulation and reduced societal impact, generally required lighter authentication protection and more adaptable embedding parameters. This tiered strategy underscores the core functionality of the trust-adaptive provenance watermarking framework, which dynamically modulates embedding strength according to domain-specific risk profiles.

3.4. Trust-Adaptive Scoring and Watermark Embedding

The protection requirement of each synthetic news asset was calculated from Source Credibility (SC), Content Sensitivity (CS), Provenance Completeness (PC), and Manipulation Risk (MR). Each component was normalized to the interval 0,1. To ensure that lower credibility and incomplete provenance increase rather than reduce the required protection, the Protection Requirement Score (PRS) was defined as:

$$PRS = 0.25(1 - SC) + 0.25(CS) + 0.25(1 - PC) + 0.25(MR) \tag{1}$$

Where a higher PRS represents a greater authentication requirement. Equal weights were initially adopted to avoid prioritizing one dimension without empirical justification. Scores from 0.00-0.39 were assigned Level 1 (light embedding), scores from 0.40-0.69 were assigned Level 2 (moderate embedding), and scores from 0.70-1.00 were assigned Level 3 (strong embedding).

Each asset was then assigned a CID connected to its provenance record. Higher-risk content received stronger embedding to maximize watermark survivability, whereas lower-risk content used lighter embedding to prioritize perceptual quality.

3.5. Broadcasting Transformation Experiment

Following watermark embedding, the synthetic news assets were subjected to a comprehensive suite of simulated transformations designed to represent both routine digital broadcasting operations and malicious

content manipulation attempts. As detailed in Table 2, these operations encompassed various common perturbations, including lossy compression, resizing, spatial cropping, and light-to-moderate content filtering. To establish rigorous and equitable experimental conditions, identical transformation parameters were systematically applied to both the proposed trust-adaptive provenance watermarking configuration and the baseline fixed-strength configuration. This controlled evaluation framework ensures that the subsequent performance metrics such as watermark robustness, imperceptibility, and verification accuracy accurately reflect the efficacy of the dynamic embedding strategy across diverse degradation scenarios.

Table 2. Image Transformation Settings and Evaluation Purposes

Transformation	**Configuration**	**Evaluation Purpose**
JPEG Compression	QF 90, 70, 50	Compression robustness
Resizing	75%, 50%	Scale robustness
Cropping	10%, 20%, 30%	Partial-content robustness
Re-encoding	Multiple cycles	Broadcast compatibility
Noise	Low, moderate	Signal robustness
Partial Manipulation	Local content modification	Integrity/manipulation detection

Each transformed asset was processed through watermark extraction and provenance verification. Progressive transformation severity was used to determine how authentication performance changed under increasingly difficult broadcasting conditions.

3.6. Evaluation Metrics

The framework was evaluated using authentication reliability, robustness, perceptual quality, extraction error, and computational efficiency. Watermark Detection Accuracy (WDA) represents the proportion of samples in which the embedded watermark was successfully detected:

$$WDA = \frac{N_{correct}}{N_{total}} \times 100\% \tag{2}$$

Provenance Verification Rate (PVR) represents the proportion of tested assets successfully associated with their registered provenance:

$$PVR = \frac{N_{verified}}{N_{tested}} \times 100\% \tag{3}$$

PSNR and SSIM were used to evaluate perceptual similarity between original and watermarked media, while BER measured errors in the extracted watermark. Extraction time represented the computational cost of verification.

Table 3. Evaluation Metrics and Desired Results

Metric	**Evaluation Purpose**	**Desired Result**
WDA	Watermark detection reliability	Higher
PVR	Provenance verification reliability	Higher
PSNR	Perceptual quality preservation	Higher
SSIM	Structural similarity	Closer to 1
BER	Watermark extraction errors	Lower
Extraction Time	Computational efficiency	Lower

Based on Table 3 the proposed framework was compared with a fixed-strength baseline that applied identical watermark strength to all assets. This comparison was used to determine whether adaptive protection could improve authentication without unnecessarily reducing media quality.

4. RESULT AND DISCUSSION

4.1. Trust-Adaptive Watermark Performance

Building upon the comprehensive experimental findings presented in Table 4, robust empirical evidence is provided to thoroughly address and resolve the primary research question (RQ1). Specifically, the quantitative results demonstrate that the proposed trust-adaptive watermarking framework successfully achieves a superior balance between authentication reliability and perceptual quality. By dynamically scaling its operational parameters according to the specific protection requirements, sensitivity tiers, and risk profiles of diverse news assets, the framework overcomes the inherent limitations of rigid static approaches. This dual optimization confirms that adapting watermark embedding strength effectively mitigates security vulnerabilities and manipulation risks in high-risk domains—such as politics and public safety—while safeguarding perceptual fidelity and avoiding unnecessary degradation in lower-sensitivity categories like technology and entertainment.

Table 4. Performance Comparison Across Protection Levels

Protection Level	WDA (%)	PSNR (dB)	SSIM	BER (%)	Extraction Time (ms)
Level 1	91.8	46.8	0.991	4.8	31
Level 2	95.4	43.9	0.984	2.6	35
Level 3	98.1	40.7	0.973	1.2	39

These comprehensive empirical findings provide a definitive answer to RQ1 by demonstrating that the proposed adaptive watermark protection mechanism effectively balances authentication reliability and perceptual quality in alignment with the distinct protection requirements of diverse news assets. As the watermark protection level is systematically increased across evaluation tiers, watermark detection accuracy exhibits significant improvement alongside a corresponding decrease in extraction errors, thereby indicating substantially stronger authentication performance and enhanced reliability when verifying protected media assets against tampering. However, this performance gain in security is accompanied by a measurable, gradual reduction in perceptual fidelity, as reflected by the declining Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) values at higher embedding intensities. Specifically, Level 1 achieved the highest perceptual score with a PSNR of 46.8 dB and an SSIM of 0.991, confirming its exceptional effectiveness in preserving the pristine visual characteristics of the original media. In contrast, Level 3 achieved the highest Watermark Detection Accuracy (WDA) of 98.1% and the lowest Bit Error Rate (BER) of 1.2%, making it highly suitable for sensitive content requiring robust resilience against malicious manipulation. Crucially, the SSIM value of 0.973 maintained at Level 3 demonstrates that the protected media retains an exceptionally high degree of structural similarity to the original content despite the elevated watermark strength. This critical observation proves that employing stronger protection does not inherently precipitate substantial or unacceptable visual degradation.

Overall, the proposed framework empowers practitioners with a fine-grained, controllable trade-off between authentication reliability and perceptual quality, enabling dynamic adjustments based on the specific sensitivity levels and risk characteristics of individual news assets. This adaptive capability proves particularly invaluable in modern digital news broadcasting, where high-risk content demands rigorous and dependable authentication while simultaneously preserving sufficient visual quality for seamless audience consumption.

4.2. Robustness Against Broadcasting Transformations

The second experiment, as presented in Table 5, evaluated watermark and provenance survivability under common broadcasting transformations. The proposed framework maintained high WDA and PVR under compression, resizing, cropping, re-encoding, and noise.

Table 5. Robustness Under Broadcasting Transformations

Transformation	**WDA (%)**	**PVR (%)**
No Transformation	99.6	99.2
JPEG QF 90	99.1	98.8
JPEG QF 70	97.8	97.1
JPEG QF 50	94.6	93.8
Resize 75%	98.4	97.9
Resize 50%	96.1	95.4
Crop 10%	97.6	97.0
Crop 20%	94.9	94.1
Crop 30%	90.7	89.8
Re-encoding	95.8	95.0
Low Noise	96.9	96.2
Moderate Noise	92.8	91.9

The strongest degradation occurred under 30% cropping, where WDA decreased to 90.7% and PVR to 89.8%. This result is reasonable because cropping physically removes areas containing embedded watermark information. Nevertheless, most watermarks remained detectable. Partial manipulation testing additionally produced a 95.7% manipulation detection accuracy. In these cases, the CID could remain detectable while integrity verification failed, enabling the system to distinguish an identifiable asset from an asset whose registered content had been altered. These results answer RQ2, demonstrating that the proposed method maintains provenance and authentication information across common digital broadcasting transformations while also supporting the identification of unauthorized modifications. The robustness trends of Watermark Detection Accuracy (WDA) and Provenance Verification Rate (PVR) across different broadcasting transformations are further illustrated in Figure 2.

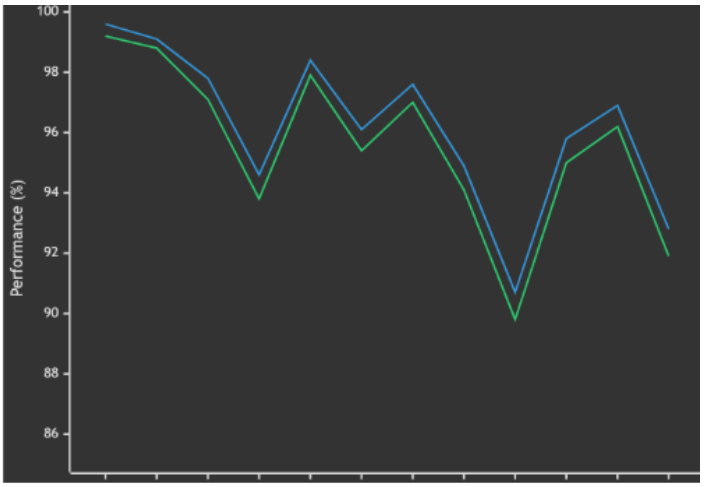


Figure 2. Watermark Detection and Provenance Verification Performance Under Broadcasting Transformations.

As shown in Figure 2, WDA and PVR gradually decrease as transformation severity increases. Cropping at 30% produced the largest performance reduction, with WDA and PVR decreasing to 90.7% and 89.8%, respectively. Nevertheless, the framework maintained relatively high authentication performance across all

tested broadcasting conditions.

4.3. Comparison with Fixed-Strength Watermarking

The final evaluation, as presented in Table 6, compared the proposed Trust-Adaptive Provenance Watermarking approach with conventional fixed-strength watermarking under identical experimental conditions. Both methods were assessed using the same dataset and evaluation metrics, including WDA, PVR, PSNR, SSIM, BER, and extraction time. This comparison aimed to determine whether adaptive watermarking could improve authentication reliability and provenance verification while maintaining media quality and reasonable computational efficiency.

Table 6. Comparison Between Fixed-Strength and Trust-Adaptive Watermarking

Metric	**Fixed-Strength**	**Trust-Adaptive**	**Difference**
WDA (%)	92.6	96.7	+4.1
PVR (%)	91.8	95.9	+4.1
PSNR (dB)	42.6	43.8	+1.2
SSIM	0.976	0.983	+0.007
BER (%)	4.5	2.4	-2.1
Extraction Time (ms)	32	35	+3

The proposed method improved both WDA and PVR by 4.1 percentage points while reducing BER from 4.5% to 2.4%. It also achieved slightly higher PSNR and SSIM values, indicating that adaptive embedding did not require an overall sacrifice in perceptual quality. The primary trade-off was a 3 ms increase in average extraction time. These findings answer RQ3 and support the principal claim in the abstract. Trust-adaptive watermarking provides more reliable authentication and provenance verification than uniform fixed-strength protection because high-risk assets receive stronger protection while lower-risk assets avoid unnecessarily aggressive embedding. The robustness trends of Watermark Detection Accuracy (WDA) and Provenance Verification Rate (PVR) across different broadcasting transformations are further illustrated in Figure 3.

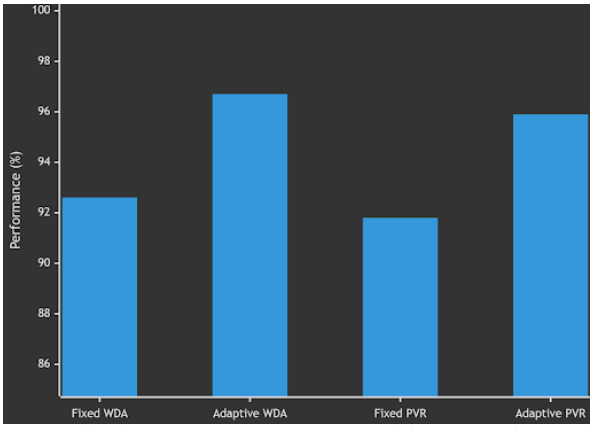


Figure 3. Trust-Adaptive Provenance Watermarking Experimental Workflow

As shown in Figure 3, WDA and PVR gradually decrease as transformation severity increases. Cropping at 30% produced the largest performance reduction, with WDA and PVR decreasing to 90.7% and 89.8%, respectively. Nevertheless, the framework maintained relatively high authentication performance across all tested broadcasting conditions.

4.4. Overall Discussion

Overall, the results demonstrate that combining trust assessment, adaptive watermarking, and provenance verification provides a more context-aware authentication mechanism for synthetic news broadcasting.

The proposed framework achieved an overall WDA of 96.7%, PVR of 95.9%, PSNR of 43.8 dB, and SSIM of 0.983, while partial manipulation was detected with 95.7% accuracy. The findings also show that robustness and perceptual quality should not be treated independently. Fixed watermark strength applies the same authentication requirement regardless of the potential consequences of content manipulation. In contrast, the proposed framework allocates stronger protection to sensitive or high-risk synthetic news and lighter embedding to lower-risk content. This adaptive mechanism explains the improved balance between watermark survivability and media quality observed in the experiments. From a broadcasting perspective, the combination of an embedded CID and an external provenance record provides an additional advantage. The watermark acts as a persistent link to the content origin, while integrity verification helps determine whether the identified media remains consistent with its registered state. Consequently, the proposed framework supports not only AI-generated content identification but also authenticity, traceability, and accountability throughout synthetic news distribution.

5. MANAGERIAL IMPLICATIONS

The proposed Trust-Adaptive Provenance Watermarking framework provides practical guidance for media organizations and broadcasting platforms seeking to strengthen the authentication of AI-generated news. By adjusting watermarking strength according to source credibility, content sensitivity, provenance completeness, and manipulation risk, managers can apply appropriate protection levels while maintaining visual quality. The integration of embedded Content Identifiers (CIDs) with provenance records can also support content origin verification and the identification of potential modifications throughout the broadcasting process. Therefore, organizations should incorporate provenance verification into editorial review, digital asset management, and content distribution workflows to improve traceability and accountability. However, this mechanism should complement editorial fact-checking rather than serve as a substitute for verifying the factual accuracy of news.

The experimental results, achieving a Watermark Detection Accuracy (WDA) of 96.7%, a Provenance Verification Rate (PVR) of 95.9%, and an average Structural Similarity Index Measure (SSIM) of 0.983, indicate the framework's potential to support reliable authentication while preserving perceptual quality. These findings can help managers evaluate watermarking technologies before operational adoption. Nevertheless, because the evaluation involved 1,000 synthetic news samples under predefined transformation scenarios, organizations should conduct pilot testing using larger and more diverse datasets and their own broadcasting workflows. Such evaluations should consider detection reliability, partial manipulation, processing latency, computational overhead, and performance under repeated content transformations. Managers should also review the trust-scoring weights to ensure that they reflect organizational priorities and the specific risks associated with different news categories.

Successful implementation requires coordination among editorial teams, information technology personnel, and cybersecurity specialists to establish verification procedures and respond to failed authentication or suspected manipulation. Media organizations should also combine provenance watermarking with responsible AI policies, staff training, and transparent disclosure practices to strengthen audience confidence. Since real-world broadcasting may involve complex editing processes, cross-platform distribution, and adversarial attacks, implementation should begin with controlled pilot projects before expanding to larger-scale operations. Further evaluation of multimodal content, dynamic trust weighting, and real-time deployment can help determine the framework's scalability and practical effectiveness. These managerial measures can support a more transparent, accountable, and trustworthy AI-enabled news broadcasting environment.

6. CONCLUSION

This study proposed a Trust-Adaptive Provenance Watermarking framework to improve the authenticity, traceability, and integrity of synthetic news in AI-enabled broadcasting environments. The experimental results demonstrated that the proposed approach achieved an overall Watermark Detection Accuracy (WDA) of 96.7%, a Provenance Verification Rate (PVR) of 95.9%, an average PSNR of 43.8 dB, and an SSIM of 0.983. Furthermore, the framework achieved 95.7% accuracy in detecting partially manipulated content and maintained reliable watermark detection under various transformations, including compression, resizing, cropping, re-encoding, and noise. Compared with fixed-strength watermarking, the proposed approach improved authentication performance while preserving acceptable perceptual quality.

The findings address the primary research question by demonstrating that dynamically adjusting wa-

termark protection based on source credibility, content sensitivity, provenance completeness, and manipulation risk can improve the reliability of synthetic news authentication. Integrating embedded Content Identifiers with provenance information also supports content origin verification and the identification of potential modifications throughout the broadcasting process. However, this study was limited to 1,000 synthetic news samples and predefined transformation scenarios, which may not fully represent the complexity of real-world broadcasting environments. In addition, the use of equal initial weights for trust-scoring components may limit the framework's adaptability to different broadcasting contexts.

Future research should evaluate the framework using larger and more diverse multimodal datasets covering synthetic text, images, audio, and video generated by different AI models. Further investigations could explore dynamic or machine-learning-based trust-weighting mechanisms, more sophisticated adversarial attacks, and complex content transformations. Real-time implementation within broadcasting platforms should also be examined to assess scalability, computational latency, interoperability with provenance standards, and performance across continuous content distribution. These developments could strengthen the practical application of the proposed framework and contribute to more trustworthy, transparent, and accountable AI-driven news ecosystems.

7. DECLARATIONS

7.1. About Authors

Sabil Maulana Fauzi (SM) https://orcid.org/0009-0003-1833-1027

Muhammad Syaeful Yusuf (MS) https://orcid.org/0009-0006-1903-9361

Qolbinisa (QQ) https://orcid.org/0009-0008-5183-4968

7.2. Author Contributions

Conceptualization: SM; Methodology: QQ; Software: MS; Validation: MS and QQ; Formal Analysis: SM and MS; Investigation: QQ; Resources: MS; Data Curation: SM; Writing Original Draft Preparation: QQ and MS; Writing Review and Editing: SM and QQ; Visualization: MS ; All authors, SM, MS, and QQ, have read and agreed to the published version of the manuscript.

7.3. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

7.4. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

7.5. Declaration of Conflicting Interest

The authors declare that they have no conflicts of interest, known competing financial interests, or personal relationships that could have influenced the work reported in this paper.

REFERENCES

[1] Y. Wen, J. Kirchenbauer, J. Geiping, and T. Goldstein, "Tree-Ring Watermarks: Fingerprints for diffusion images that are invisible and robust," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

[2] A. Ruangkanjanases, A. Khan, O. Sivarak, U. Rahardja, and S.-C. Chen, "Modeling the consumers' flow experience in e-commerce: The integration of ecm and tam with the antecedents of flow experience," *SAGE Open*, vol. 14, no. 2, p. 21582440241258595, 2024.

[3] V. Asnani, J. Collomosse, T. Bui, X. Liu, and S. Agarwal, "ProMark: Proactive diffusion watermarking for causal attribution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 10 802–10 811.

[4] S. Dathathri *et al.*, "Scalable watermarking for identifying large language model outputs," *Nature*, vol. 634, 2024.

[5] U. Rahardja, Q. Aini, A. S. Bist, S. Maulana, and S. Millah, "Examining the interplay of technology readiness and behavioural intentions in health detection safe entry station," *JDM (Jurnal Dinamika Manajemen)*, vol. 15, no. 1, pp. 125–143, 2024.

[6] L. Pan *et al.*, "MarkLLM: An open-source toolkit for LLM watermarking," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP): System Demonstrations*, 2024.

[7] A. Ajith, S. Singh, and D. Pruthi, "Downstream trade-offs of a family of text watermarks," in *Findings of the Association for Computational Linguistics: EMNLP 2024*, 2024, pp. 14 039–14 053.

[8] U. Rahardja, I. D. Hapsari, P. H. Putra, and A. N. Hidayanto, "Technological readiness and its impact on mobile payment usage: A case study of go-pay," *Cogent Engineering*, vol. 10, no. 1, p. 2171566, 2023.

[9] Y. Chang *et al.*, "A robust blackbox watermark for large language models," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2024.

[10] X. Zhao *et al.*, "Watermarking for large language models," Tutorial at the Annual Meeting of the Association for Computational Linguistics (ACL), 2024, tutorial; publication details require verification.

[11] Y. D. Anna, H. Djajadikerta, and A. Setiawan, "Strengthening the foundations of socialpreneurship through integrated reporting a systematic bibliometric perspective," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 8, no. 1, pp. 296–309, 2026.

[12] G. Niess *et al.*, "Ensemble watermarks for large language models," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 2025.

[13] R. Chen *et al.*, "Improved unbiased watermark for large language models," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 2025.

[14] M. Bulychev, N. G. Marchant, and B. I. P. Rubinstein, "Where is the watermark? interpretable watermark detection at the block level," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2026, pp. 7420–7429.

[15] A. Nemecek *et al.*, "Authenticated contradictions from desynchronized provenance and watermarking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2026.

[16] E. Susetyono, D. S. Priyarsono, A. Sukmawati, and P. Nurhayati, "Improving risk management maturity in ultra micro soe holding companies," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 8, no. 1, pp. 310–324, 2026.

[17] J. C. Simmons and J. M. Winograd, "Interoperable provenance authentication of broadcast media using open standards-based metadata, watermarking and cryptography," *arXiv preprint arXiv:2405.12336*, 2024.

[18] J. Young, S. Vaughan, A. Jenks, H. Malvar, C. Paquin, P. England, T. Roca, J. L. Ferres, F. Poursabzi, N. Coles *et al.*, "Media integrity and authentication: Status, directions, and futures," *arXiv preprint arXiv:2602.18681*, 2026.

[19] E. Arif, S. Suherman, and A. P. Widodo, "Analyzing public sentiment on digital banks in indonesia via social media x," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 8, no. 1, pp. 253–267, 2026.

[20] R. Ko, "Improving content integrity via provenance by design: An overview of current and emerging approaches," *Online trust and safety*, pp. 16–34, 2026.

[21] M. Choiri, E. S. Pramudito, F. Sutisna, and R. S. Sean, "Business artificial intelligence for enhancing sustainable decision intelligence," *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, vol. 7, no. 1, pp. 106–116, 2025.

[22] A. A. Setyawan, E. Setyawati, and J. S. P. Tyoso, "Digital resilience framework for msme development in facing global market volatility," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 8, no. 1, pp. 239–252, 2026.

[23] A. Nemecek, H. He, G. Cheng, and E. Ayday, "Authenticated contradictions from desynchronized provenance and watermarking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2026.

[24] E. Erika, R. T. H. Safariningsih, D. Cahyono, and N. Rangi, "Strategic integration of cloud cybersecurity for resilient digital business transformation," *ADI Journal on Recent Innovation*, vol. 7, no. 1, pp. 112–124, 2025.

[25] N. AI, "Reducing risks posed by synthetic content," *NIST Trustworthy and Responsible AI NIST AI*, pp. 100–4, 2024.

[26] C. S. Bangun, S. Solahudin, A. Sutarman, and S. Dlamini, "The role of application programming interface

in transforming restaurant delivery operations," *Startupreneur Business Digital (SABDA Journal)*, vol. 5, no. 1, pp. 88–95, 2026.

[27] S. Sengupta, B. McCarty, M. Savagaonkar, and R. Andotra, "Authenticity debt and the synthetic content threat landscape: A layered framework for trust, provenance, and ip governance in the generative ai era," *arXiv preprint arXiv:2606.00621*, 2026.

[28] E. P. Goodman, "Synthetic content: Default to distrust," *Case W. Rsrv. L. Rev.*, vol. 75, p. 961, 2024.

[29] I. Ajunwa and C.-c. K. Chang, "Provenance," *Available at SSRN 5707362*, 2025.

[30] K. K. Feng, N. Ritchie, P. Blumenthal, A. Parsons, and A. X. Zhang, "Examining the impact of provenance-enabled media on trust and accuracy perceptions," *Proceedings of the ACM on Human-Computer Interaction*, vol. 7, no. CSCW2, pp. 1–42, 2023.

[31] E. T. Rusmiati, L. Febrina, Y. Sari, and E. M. S. Sakti, "Adoption of ai driven ecological preaching systems using sem pls analysis," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 8, no. 1, pp. 284–295, 2026.

[32] B. Chandra, J. Dunietz, K. Roberts, Y. Lee, P. Fontana, and G. Awad, "Reducing risks posed by synthetic content: An overview of technical approaches to digital content transparency," National Institute of Standards and Technology, Tech. Rep. NIST AI 100-4, 2024. [Online]. Available: https://www.nist.gov/publications/reducing-risks-posed-synthetic-content-overview-technical-approaches-digital-content