A Proposed Adaptive Framework for Multimodal Media Integrity Verification Using Reinforcement Learning and Blockchain-Based Provenance

Nithya sri M[1], Pooja T[2], Riyapriti G[3], Vishnupriya T[4]
[1,2,3,4]*Department. of Computer Science and Engineering Vivekanandha College of Engineering for Women (Autonomous) Tiruchengode, Namakkal, Tamil Nadu, India*

***Abstract*—The explosion of digital media and the online sharing of information has made it ever more difficult to distinguish what is real from what is fake, manipulated or misleading. Existing verification techniques are mainly centralised, use static classification models and generally only provide reports on whether the content is real or fake. This paper presents a Hybrid Decentralised Framework for Media Integrity Verification based on Natural Language Processing (NLP), Reinforcement Learning (RL) and Blockchain technology. Users can submit text, URLs, files, images and audio one at a time or in batches.An NLP-based analysis is utilised to extract semantic, contextual and linguistic credibility features, while an RL agent based on the proposed Dynamic Reward-Guided Verification Algorithm (DRVGA) adapts its verification policy using credibility-based reward feedback. Cryptographic hashing is used to generate a unique fingerprint of each submission, while blockchain records with smart-contract validation keep verification outcomes tamper-resistant, traceable and auditable. The outputs are combined into a credibility score and a final decision of Authentic or Fake. For misleading content, a content preview provides a reference or suggested version of the information. The paper reviews fifteen related studies, analyses the limitations of existing systems, and describes the architecture, modules, and system requirements of the proposed framework. The framework can support digital forensics, journalism, social media, and secure digital content verification.**

***Index Terms*—media integrity verification, fake news detection, natural language processing, reinforcement learning, blockchain, DRVGA, cryptographic hashing, credibility scoring.**

I. INTRODUCTION

The rapid growth of the Internet, social media, online news platforms, and digital communication has made it possible to create and share information within a very short period of time.

Although this has improved access to information, it has also increased the spread of fake, misleading, manipulated, and unreliable content. Such information can create confusion among users, influence public opinion, damage the reputation of individuals or organizations, and reduce trust in digital media. Reliable and efficient identification of potentially misleading information has therefore become an important research challenge.

Cconventional misinformation detection methods commonly employ computational techniques such as machine learning, deep neural networks, and language processing to identify misleading information. These systems examine the features of content and categorize it based on patterns identified in a set of previously gathered system. But misinformation constantly evolves in terms of its appearance and content. New misinformation can include altered wording, altered images, dubious audio, or information from dubious sources. A system that is able to process various types of digital contents and adjust to patterns of information changes is needed. Most current systems also maintain a single central authority for keeping its data, models and decisions, thus introducing a single point of failure and little protection against manipulation of the verification record.

To cope with these issues, this paper introduces a Hybrid Decentralized Framework for Media Integrity Verification. The framework allows for semantic and contextual analysis of the information using NLP, adaptive navigation and decision-making through RL, and secure, transparent, and traceable verification records using Blockchain. These components are connected together by a Dynamic Reward-Guided Verification Algorithm (DRVGA) that calculates a

credibility-based reward to direct the RL agent. The proposed system accepts different forms of input, including text, web links, documents, images, and audio recordings. It also enables multiple items to be processed together and presents the analysis results through an interface that allows users to preview the submitted content before evaluation. When the submitted content is flagged as potentially false or deceptive, the system displays relevant sources or an alternative version for comparison, helping users identify and evaluate questionable information. This is not a guarantee of automatic conversion to a guaranteed truth, but rather it is considered as verified when it is backed by reliable verification evidence.

This work addresses the following problem: Given digital content (via text, URL, file, picture or audio, which include both authentic and potentially fake or misleading content), maintain a trustworthy and auditable record of the decision, assist the user to understand the trusted version of the information, and enable the verification strategy to evolve over time as misinformation evolves. The goals of the work are then to:

• Develop content verification processes based on artificial intelligence and decentralized verification systems to make it more reliable, transparent, and secure

• Adaptively make decisions, allowing the verification policy to learn from credibility-based reward feedback

• Ensure tamper-proof and traceable verification records, which can be audited, are maintained

• Deliver clear results using a credibility score, supporting information, and a content preview.

The key contributions of the proposed study are outlined below.

• A combination of NLP, RL and Blockchain for verification services including text, URL, file, image and audio verification – even batch processing for many input types.

• Dynamic Reward-Guided Verification Algorithm (DRVGA), which uses language consistency, cross-modal alignment and blockchain validation results to generate a credibility-based reward.

• A tamper-resistant storage, provenance and auditable history of decisions through a hash-based integrity mechanism and a blockchain record layer with smart-contract validation.

• A credibility score, final Authentic or Fake decision, content preview that provides a reference or suggested version of the misleading content.

• A review of 15 recent studies, structured analysis of the weaknesses of the current systems, and the research gaps identified which drive the proposed design.

The remaining sections of this paper are structured in the following manner.Section II examines existing research relevant to the proposed approach, while Section III analyzes currently available methods and discusses their limitations.. The proposed framework, its architecture, modules, workflow and DRVGA are presented in section IV. The system requirements are discussed in Section V. The benefits, comparison with existing systems, security issues and challenges are discussed in section VI. Section VII presents the proposed evaluation approach; Section VIII provides examples of applications and Section IX wraps up the paper and proposes future research.

II. RELATED WORK

In this section, we provide a review of 15 recent studies related to fake-news detection, misinformation detection and manipulated-media detection. The studies are categorized to multimodal fusion approaches, language-model and feature based approaches, and adaptive, secure, and manipulation approaches. Each study's technique, modalities and principal limitation are summarized in Table I.

A. Multimodal Fusion and Attention-Based Detection

In recent multimodal writing, one of the reoccurring concepts is that a fabricated story doesn't always remain coherent throughout all its channels. The framework in [1] takes this further by aggregating text, audio and video frames and transferring knowledge across different domains, which is beneficial if the topic of the test data is not the same as the training data. One of the weighty features of this pipeline is the price, and the need for large multimodal corpora, with nothing in the design to allow the system to adapt its strategy after deployment.

Most other work focuses on a headline and image and asks the question of whether or not the image is the right choice for the headline. In [2], co-attention is leveraged in addition to image–text matching signals from a pretrained vision-language model and this knowledge is distilled between the branches of the network. CAF-ODNN [3] weights complementary text

and image cues, via an attention-fusion stage, and tunes a deep network according to grid search. The approach in [4] combines features in an autoencoder followed by a cross-attention layer revealing mismatches between the two media. Such systems are limited to text and still images, a claim that is conveyed by speech or video is not within their scope, and [2] and [4] do not leave a verifiable record of their outputs.

A second group concentrates on how features from different encoders are combined. The dual-stream network in [5] uses ERNIE-M encoders and ShuffleNet encoders in combination with multi-head self-attention. The model in [6] proposes a gate to control the contribution of each modality after the cross-modal message aggregation. Both SARD [7] and [8] adopt a heterogeneous news graph to introduce social context and CLIP-style contrastive learning to align texts and images, with the former augmenting the latter to make the combined representation more difficult to fool. They demonstrate that the cross-modal agreement is a good indicator of manipulation. Each of these, however, has a number of large networks that are stacked, incompletely handles audio and video, and has a final verdict that is not subject to refinement via feedback or to storage in a verifiable manner.

B. Feature-Based Detection

The idea of text-centred tasks is to examine the extent to which the meaning of a language model can be recovered from the words. In [9] Transformer-based encoders (e.g., MPNet and SentenceT5) are benchmarked for fast screening of social-media posts and tested on unseen data as a useful baseline for testing their generalization ability. GBERT [10] combines the GPT and BERT models, enabling the use of both generative and bidirectional context for the classifier. In ECMT [11] the problem is formulated in terms of the entities in a story and domain-specific and domain-independent signals are separated by using Wikipedia-derived embeddings. The work in [12] shows the reverse: Classical machine learning, aided by the metadata and social context, is a competitive baseline.

These four have a common failing – scope. None of them can review an attached photo or video, none of them retain a record of their own decisions, and after training they aren't left with the ability to learn from the results of their performance. This is not an appropriate model for misinformation that continually morphs.

C. Adaptive, Secure, and Manipulated-Media Detection

There are only a few studies attempting to make the process of verification adaptive or trustworthy. Transformer models, combined with reinforcement-based optimization are used in [13] to achieve multimodal detection in a more refined manner, but they do not aim at securing the provenance of the work or continuous decentralized checking. Most similar to the current study is the work in [14] that combined deep learning, reinforcement learning, and blockchain together in an ensemble, with its drawbacks being an increasing complexity of the system and the challenges of scaling it in real-time. Lastly, the study in [15] surveyed deep-learning detectors for deepfakes and manipulated media, emphasized the importance of having proper training data, and noted the impact of new generation methods on detection accuracy. It's about manipulated media and not end-to-end verification of news.

D. Research Gaps

Table I shows that there are two general trends. The first one enriches the representation of news item by multimodal fusion and attention, thus giving more evidence of inconsistency between text and media, but increasing the computational complexity. The second focuses on Transformer and hybrid language models with the aim of improving the understanding of language, which is able to capture the context but limited to text. In 11 of 15 studies, text is the primary modalities, or text and image. While the reviewed studies indicate that adaptive and secure verification has been significantly less explored compared to detection accuracy, just two studies [13, 14] use reinforcement learning and only one [14] combines reinforcement learning with blockchain.

TABLE I. Overall, it is found that there are gaps in the consistency across the review. Most multimodal studies consider only text–image pairs, and deal with video and audio partially. In most detectors, the detector once trained does not adapt in a reward based manner. There are very few studies that offer tamper-resistant verification records or provenance, and those that combine multiple technologies are complex and scalability challenges. Also, nearly none of them makes the user avail of a reference version of the information when the content is misleading. The proposed framework will solve these problems by merging

multi-modal analysis, adaptive verification with DRVGA, records based on blockchain, and a preview of the content.Summary of Reviewed Studies

Ref.	Year	Technique / Model	Modalities	Principal Limitation
[1]	2025	Multi-domain, multimodal feature framework with domain knowledge transfer	Text, audio, video frames, comments	High computational cost; needs large multimodal datasets; no RL reward mechanism
[2]	2024	Matching-aware co-attention with mutual knowledge distillation	Text, image	Limited to text–image data; no complete video; no blockchain provenance
[3]	2024	CAF-ODNN: complementary attention fusion with optimized DNN	Text, image / captions	No complete video or audio; fusion and optimization add complexity
[4]	2025	Bimodal Transformer, autoencoder fusion, cross-attention	Text, image	No video or audio; high resource needs; no blockchain verification
[5]	2024	Dual-stream fusion network (ERNIE-M, ShuffleNet) with MHSA	Text, image	Video and audio not treated as complete modalities; several deep models
[6]	2024	Cross-modal message aggregation and gated fusion (BERT, Faster R-CNN, ResNet-50, VGG-19)	Text, image	Several feature extractors add complexity; no blockchain integrity verification
[7]	2024	SARD: CLIP contrastive learning and semantic alignment	Text, image, social context	No RL-based dynamic adaptation; multiple sources increase complexity
[8]	2024	Similarity inference with adversarial network (BERT, Text-CNN, VGG-19)	Text, image	No complete video analysis; no continuous RL adaptation
[9]	2024	Transformer models (MPNet, SentenceT5, T5)	Text	Text only; no blockchain; no continuous RL adaptation
[10]	2024	GBERT: hybrid GPT–BERT model	Text	Text only; no blockchain; not designed for continuous learning
[11]	2024	Entity-centric multi-domain Transformer with Wikipedia embeddings	Text, entities	No video processing; no blockchain provenance
[12]	2024	Machine learning on text, metadata, and social context	Text, metadata	No advanced image/video analysis; no RL adaptation
[13]	2025	Transformer models with reinforcement-based optimization	Multimodal	Provenance not the focus; no continuous decentralized verification; extra compute
[14]	2025	Multimodal fusion, ensemble, deep learning, RL, and blockchain	Multimodal	System complexity; real-time scalability is difficult
[15]	2024	Review of deep-learning deepfake detection	Manipulated media	Not complete fake-news verification; accuracy drops for new generation techniques

III. EXISTING SYSTEMS AND THEIR LIMITATIONS

A. Existing Approach

Traditional media integrity verification system use static machine learning or deep learning system and centralized architecture. They typically employ a supervised learning model, where classes like

Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Support Vector Machines (SVM) and Transformer-based classifiers are trained on labeled data to discriminate between real and fake content. The user enters a news article, which is preprocessed and then fed into an NLP or deep learning model trained on a static set of news articles and returns either a fake or a real result as seen in Fig. 1.

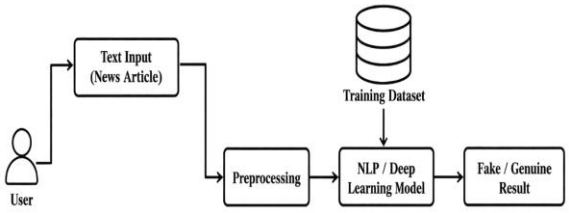


Fig. 1. Existing system diagram.

Such systems have a single server and/or authority to collect, store, train models and make decisions. Once deployed, the model does not change unless new data is being gathered and the model is retrained using the new data. Misinformation techniques are continuously changing and models trained on a static dataset will face concept drift and thus experience a decrease in performance when new patterns of misinformation are introduced. Traditional classifiers, however, are not adaptive to real-time feedback of credibility, have no way of keeping records of decisions made or decision histories transparent to the users, and are often black boxes whose decisions are difficult for users to interpret.

B. Limitations

Existing media integrity verification systems have the following shortcomings:

• Centralized architecture: A central server means that if it fails, there is a point of failure, if it is compromised, it hinders the entire verification process. Risk of cyber-attacks. Centralized servers store vast quantities of data and are desirable for hacking into and tampering with.

• Static learning: Once a model has been trained on a fixed set of data, it will continue to run the same behavior until it is retrained with fresh data.

• Concept drift: When new forms of misinformation are not in the historical training data, performance decreases. • Limited real-time adaptability. Supervised classifiers are not continuously learning from real time feedback and/or do not update the decision policy.

• Poor generalization: Good benchmark performance may not transfer to a variety of unknown and/or changing misinformation.

• Lack of transparency: Black-box models make predictions without explicitly stating what is causing them.

• No tamper-proof records: Centralized systems don't automatically give immutable records of verification decisions and it's hard to maintain an auditable history on them.

• Limited data provenance: It is not possible to track, verify, and confirm the origin and history of digital content with a reliable decentralized mechanism.

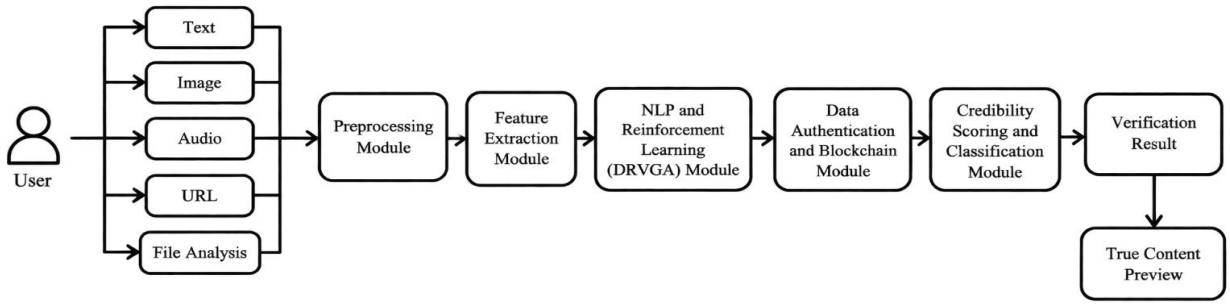
• Limited cross-modal verification: There is no effective examination of the consistency of different forms of information (text and associated media).

IV. PROPOSED FRAMEWORK

A. Overview

The proposed system is the Hybrid Decentralized Framework for Media Integrity Verification to address the drawbacks of centralized and static misinformation detection systems. Combines NLP, RL, and Blockchain technologies to offer a safe, adaptable, and transparent validation process. It supports text analysis, URL analysis, file analysis, image prediction and audio prediction, and offers batch processing to be able to process multiple items at once. This makes it not be limited with the conventional fake news classification based on texts.

The NLP component applies semantic and contextual analysis, and detects features related to linguistic inconsistencies, misleading stories, and credibility. Cross-modal consistency analysis analyses the interaction between the textual information and related media. The RL component is used to incorporate adaptive decision-making; the RL agent is not set with a fixed classifier, but rather updates its verification policy with reward feedback provided by DRVGA. The blockchain element is responsible for keeping a cryptographic representation of the information being verified, which is an immutable historical record, and smart contracts perform validation processes automatically. The proposed system's end-to-end flow from the input options, preprocessing, feature extraction, NLP and RL (DRVGA), data authentication, and data verification from blockchain, to the credibility score, verification result and content preview, is depicted in Fig. 2.



B. Proposed system diagram.Design Rationale

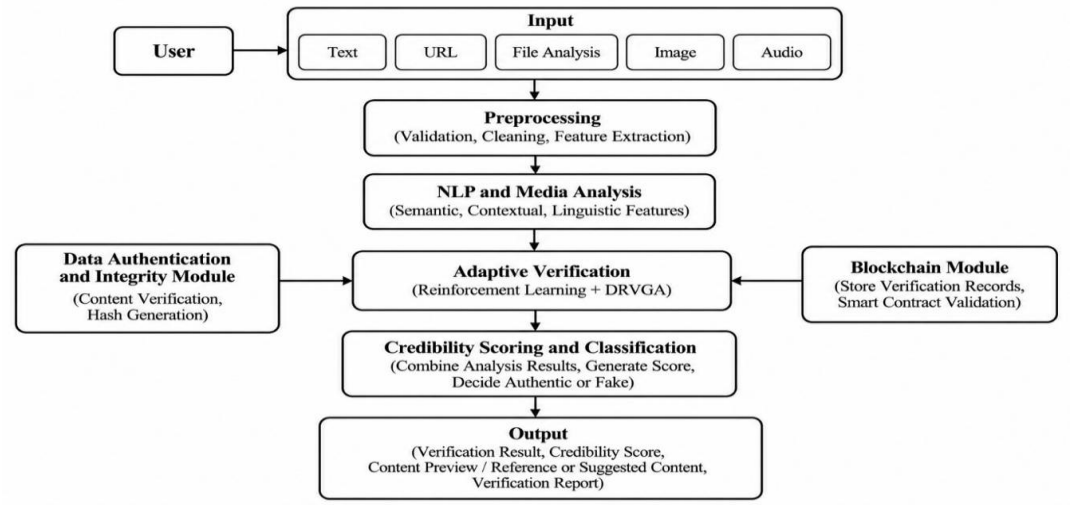
The three core technologies address three different shortcomings of current systems. NLP offers semantic and contextual analysis of content, as misleading information may not contain any obvious false keywords and its meaning may rely on the interactions between different parts of the text. The great thing about reinforcement learning is that the agent adapts: its policy not just during training, but when new patterns of misinformation appear. Blockchain ensures trust: hashes and verification records held throughout a distributed ledger are easily accessible, tamper-proof, and open to audit; decentralized consensus reduces reliance on any single authority.

The three technologies are linked by DRVGA. Its reward is based on the linguistic evidence generated by NLP, on the cross-modal coherence of the different parts of the content, and on the outcome of validation generated by the blockchain layer, thus influencing the RL policy in both analytical and integrity aspects. This is the main difference between the proposed framework and a basic fake-news classification model.

C. System Architecture

The architecture (see Fig. 3) starts by gathering and preprocessing digital content – text, URLs, files, images and audio. Preprocessing validates the content, cleans it and extracts features. The next step, the NLP and media analysis, gathers semantic, contextual, and linguistic characteristics, as well as any misleading patterns. The extracted information is then fed into the adaptive verification stage where the verification policy is optimized based on the credibility-based reward feedback by guiding it through DRVGA.

Two supporting modules are now connected. The data authentication module calculates the hash of the content for validation and the blockchain module maintains records of content validation and validates the smart contracts. Lastly, the results from the NLP, RL, and blockchain verification stages are integrated at the credibility scoring and classification stage into a credibility score and a final Authentic or Fake classification. The user will receive a verification result, credibility score, content preview/reference/suggested content, and a verification report.



D. System architecture diagram.Input Modalities and Preprocessing

The system has an easy-to-use interface that allows for the submission of various forms of digitized information. Text analysis is based on the information embedded in the text and the utilization of NLP tools to analyze the linguistic, semantic and contextual features of the text. URL analysis enables information retrieved from the Web to be analyzed. File analysis allows the user to upload supported digital files, where the file format is checked before the file is processed. Image prediction and audio prediction takes the verification process into visual and audio information, so as to have the analysis not only based on the textual content.

For both items and collections of digital content the framework can be used for batch processing, that is several content items can be analyzed at once in a single processing. After submission, preprocessing checks the submitted content, cleans it and extracts the required features based on the type of content. These features are presented in a formalized structure, enabling the same modules in the downstream processes to process all types of input, and allowing cross-modal consistency analysis to be performed when there is text with accompanying media.

E. Module Description

The framework is divided into 6 modules, as illustrated in Fig. 4, and the processing and outputs of each module are summarized in Table II.

1)Media Collection and Organization: This module will gather and structure the media that the user provides. It loads Text, URL, File, Image, Audio and Batch Processing sources, validates the data and the file format, cleans and preprocesses the data, and extracts and structures the relevant content information in a structured manner for the analysis modules.

2)Fake Media Detection Using NLP: In this module NLP techniques are used for detecting potentially misleading and manipulated information. It conducts semantic and contextual analysis, detects misleading narratives, linguistic inconsistencies, credibility-revealing linguistic features, sentiment analysis and content consistency analysis (where applicable), and analyzes supported text, image and audio content. Features extracted are fed into the adaptive verification module.

3) Adaptive Verification using Reinforcement Learning RL: The RL agent employs the features derived in the analysis stage and learns with the reward feedback based on credibility. The decision policy can adjust to the evolving trend of misinformation rather than being hard-coded, supported by DRVGA's dynamic reward guidance, which is based on verification-related factors.

4) Data Authentication and Integrity Verification: This module verifies the integrity of the data and ensures its authenticity during verification process. The integrity reference used is a cryptographic hash of the input content. The module will validate integrity before final classification, cross check consistency between processing modules, and maintain integrity reference used for blockchain validation.

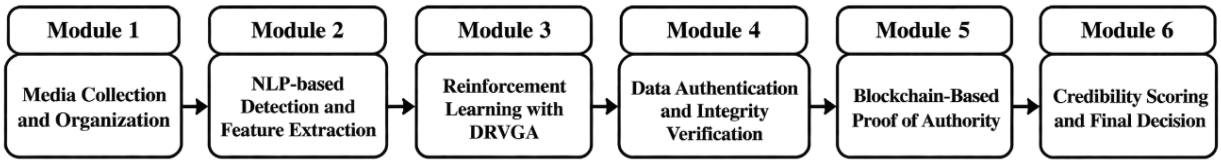
5) Blockchain Based Proof of Authority: This module is used for recording verification information that cannot be tampered with on the blockchain. It generates verification records and cryptographic information, logs verification in an immutable manner, offers traceability of verification information and implements smart contracts to automate validation, thereby ensuring transparency and accountability.

6) Credibility Scoring and Final Decision: This module combines the NLP and RL outputs, blockchain outputs and authentication output to calculate the credibility score and returns Authentic / Fake result. It shows an explanation or supporting verification text and offers the preview of content and/or a reference version or suggested version of the content (as applicable).

TABLE II. SUMMARY OF FRAMEWORK MODULES

Module	Main Processing	Output
1. Media collection	Validation, cleaning, preprocessing, and organization of text, URL, file, image, audio, and batch inputs	Structured content
2. NLP detection	Semantic and contextual analysis; linguistic feature extraction	Credibility-related features
3. RL with DRVGA	Policy learning with the credibility-based reward	Adaptive verification decision

Module	Main Processing	Output
4. Authentication	Hash generation; integrity and consistency checks	Integrity reference
5. Blockchain	Smart-contract validation; immutable storage	Tamper-resistant audit record
6. Scoring and decision	Combination of NLP, RL, integrity, and blockchain outputs	Score, Authentic / Fake result, content preview



F. Modules flow diagram.End-to-End Verification Workflow

The following steps describe the general verification of an item where verification is possible:

- The user inputs text, a URL, a file, an image, audio or a set of items through the interface.
- The content is validated, cleaned, organized and features that are relevant are extracted.
- The integrity reference is a cryptographic hash of the content.
- The features of semantic, contextual, and linguistic credibility are extracted by using NLP and media analysis.
- The RL agent decides on a verification decision and the DRVGA calculates the reward based on the credibility.
- The verification record is validated using smart contracts and recorded on the blockchain.
- The credibility score is computed and content is categorized as Authentic or Fake.
- The user gets the result, the score, a content preview and a verification report.

G. Integrity Hashing and Blockchain Record

The data authentication module calculates a cryptographic digest for each item m submitted.

$$h = H(m) \quad (1)$$

where h is a cryptographic hash function (e.g., SHA-256) and h is the unique digital fingerprint of the content. The digest gets changed if there's a change in the content, which makes the comparison of the digest at each processing stage possible to ensure the same content was analyzed by the various modules. This digest is then combined with the outcome of the verification and a timestamp and written onto the blockchain. The history of verifications can be tracked and audited without having to rely on any single central authority, as records on the chain are tamper-resistant and distributed, and smart contracts can validate each record automatically before accepting it.

H. Dynamic Reward-Guided Verification Algorithm

DRVGA rewards the RL agent with a credibility-based reward, following each verification decision. The reward is composed of three elements of the framework: language consistency, cross-modal alignment, and the outcome of the blockchain validation. The reward at step t is given by

$$R_t = w_1 L_t + w_2 C_t + w_3 B_t \quad (2)$$

where all the weights w1, w2 and w3 are normalized in [0, 1] for linguistic consistency, cross modal alignment and blockchain validation, respectively. The weights can be adjusted to reflect the importance of the various factors in the case of a particular content type.

The agent then adjusts its verifier accordingly to the reward it observes, and

$$\pi_{t+1} = U(\pi_t, s_t, a_t, R_t) \quad (3)$$

At is the action for verification performed, st denotes the state extracted from the state, and U represents the policy update rule of the selected RL method. DRVGA can be characterized as:

• Take current state as feature vector from the NLP and media analysis stage. Use the existing policy to select the action that will be taken on verification.

• Compute the linguistic-consistency and cross-modal-alignment scores for the decision. Get blockchain validation result of the record's related block.

• Calculate the reward using (2) and update the policy using (3).
• Send the new verification output to the credibility scoring module.

I. Credibility Scoring and Content Preview

The final credibility score is the result of merging the three results of the verification:

$$S = \alpha PNLP + \beta PRL + \gamma VBC \quad (4)$$

In this scenario, PNLP and PRL are the credibility rating of the NLP and RL module, respectively, VBC is the result of validating the blockchain and integrity, and $\alpha + \beta + \gamma = 1$. The score is then compared with a threshold τ, which leads to a final decision: the content is considered Authentic if $S \geq \tau$ and Fake if $S < \tau$. The score is also presented to the user to give an indication of confidence in the decision.

The outcome is provided to the user as a verifiable report that is easy to understand. The final classification, or credibility score, is included in the report, an explanation or supporting verification information is included in the report, the relevant integrity and blockchain information (such as content hash, record reference) is included in the report, and the content preview (where applicable) is included in the report.

If the content is determined to be potentially fake or misleading, the content preview displays the appropriate information in a more readable format, and, if applicable, a reference or recommended version of the content. This enables the user to detect unreliable information and what the reliable version might be. The preview is an interpretation or reference and can only be regarded as verified content when there is reliable verification evidence to support this.

V. SYSTEM REQUIREMENTS

System requirements explain what hardware and software resources are required to design, run, test and implement the framework. The system incorporates AI, NLP, RL, and blockchain technologies, which means sufficient computing power and development tools are necessary. The minimum hardware requirements are presented in Table III and the software requirements, in Table IV.

TABLE III. HARDWARE REQUIREMENTS

Hardware Component	Minimum Requirement
Processor	Intel Core i5 or above
RAM	8 GB minimum; 16 GB recommended
Storage	256 GB SSD or above
GPU	NVIDIA GPU
Internet connection	Required

TABLE IV. SOFTWARE REQUIREMENTS

Software Component	Requirement
Operating system	Windows / Linux / macOS
Programming language	Python
Machine learning framework	TensorFlow / PyTorch
NLP libraries	NLTK / SpaCy
Reinforcement learning tools	OpenAI Gym
Blockchain platform	Ethereum / Hyperledger
Database	MySQL / MongoDB
IDE	VS Code

The tools that are listed correspond naturally with the modules of the framework. The common programming language used is Python. The NLP, image and audio models are supported by TensorFlow or PyTorch, and text preprocessing and linguistic analysis are supported by NLTK or SpaCy. OpenAI Gym gives the RL agent and reward of the DRVGA an environment in which they can be defined and trained. The records are kept in the Ethereum or Hyperledger blockchains and smart contracts, and the metadata and verification results for submitted content are stored in a relational database (MySQL) or non-relational database (MongoDB) that are not part of the blockchain. The GPU is needed primarily for training and operating the learning models and an Internet connection is needed for URL analysis and for communication with the blockchain network.

VI. DISCUSSION

A. Advantages of the Proposed Framework

The framework has the following benefits over traditional detection systems:

• Decentralized verification: Single authority reliance is reduced, reducing the risk of single point failure.
• Tamper-proof records and secure provenance: Blockchain storage secures verification information from being changed by unauthorized parties and keeps track of the source and history of verification information.
• Adaptive learning and better adaptability : The RL agent can change its verification policy based on the credibility of the reward, and it can follow the evolution of the information pattern.
• Semantic analysis and contextual analysis: NLP detects potentially misleading stories in the meaning and context of the text.
• Multi modal analysis and batch processing: The text, URL, file, image and audio inputs are analyzed in a single framework, and numerous items can be processed simultaneously.
• Clearly indicated verification and credibility scoring: Users can see the confidence level behind a decision by seeing the recordable and credibility score.
• Content preview and increased trust: Users can view a misleading version or a reference version, and adaptive analysis, along with a decentralized record, adds to the confidence in the process.

B. Comparison with Existing Systems

Table V shows a comparison of the proposed framework to traditional centralized systems on the primary design aspects.

TABLE V. EXISTING VERSUS PROPOSED SYSTEMS

Feature	Existing Systems	Proposed Framework
Architecture	Centralized server or authority	Hybrid, decentralized verification
Learning	Static; retrained on new data	Adaptive RL with DRVGA rewards
Input types	Mostly text	Text, URL, file, image, audio, batch
Records	No inherent tamper-proof records	Blockchain with hashes and smart contracts
Provenance	Limited	Traceable, auditable history
Transparency	Black-box decisions	Credibility score and verification report
Cross-modal check	Limited	Cross-modal consistency analysis
User output	Fake / genuine label	Result, score, content preview, report

C. Limitations Addressed by the Framework

Table VI shows the correlation between each of the limitations identified in Section III and the feature of the proposed framework that addresses it.

TABLE VI. LIMITATIONS AND CORRESPONDING FRAMEWORK FEATURES

Existing Limitation	Addressed by the Proposed Framework
Centralized architecture	Decentralized verification using blockchain records
Cyber-attack vulnerability	Distributed, tamper-resistant records and hash-based integrity checks
Static learning	RL agent that updates its policy from reward feedback
Concept drift	Dynamic DRVGA reward that follows changing misinformation patterns
Limited real-time adaptability	Continuous credibility-based reward feedback
Poor generalization	Multimodal evidence combined with an adaptive verification policy
Lack of transparency	Credibility score, supporting information, and verification report
No tamper-proof records	Blockchain storage of hashes with smart-contract validation
Limited data provenance	Traceable and auditable blockchain history of decisions
Limited cross-modal verification	Cross-modal consistency analysis between text and media

D. Security Considerations

The framework has three levels of security. On the content level the cryptographic hash can ensure that any change in the content that is submitted between processing stages is detected. Indeed, on the record level, the digest and the verification result is stored on the blockchain, which makes any subsequent change to the verification history obvious and enables the audit of the decisions. Decentralized verification eliminates the single point of failure (SPOF) caused by a central server at the architecture level. Note that the blockchain does not endorse the correctness of the decision, it only ensures that the record of the decision is correct. As it

is the correctness of the decision depend on the quality of the NLP and RL analysis, the credibility score shows to the user and the supporting information is presented as a reference and not as certain truth.

E. Challenges and Limitations

There are a few problems arising from the literature and from the design. The number of modalities and the incorporation of RL training, however, add to computational demands, while sufficiently large and appropriate datasets are needed for effective multimodal analysis. Blockchains may also become slower to process transactions and store verification records as the number of records increases, therefore, the scaling issue must be addressed carefully. The incorporation of NLP, RL, and blockchain also adds to the complexity of this system. Lastly, the content preview is a reference or suggested version, and must be supported by supporting evidence in order to be considered verified information.

VII. PLANNED EVALUATION METHODOLOGY

The design of the framework is presented in this paper, while its quantitative evaluation will be carried out during implementation. The evaluation will be used to evaluate both the quality of the detection achieved, as well as the properties that set the system apart from traditional systems: adaptability, integrity protection, and blockchain overhead. There are the planned evaluation aspects and metrics listed in Table VII.

TABLE VII. PLANNED EVALUATION METRICS

Aspect	Metric	Purpose
Classification quality	Accuracy, precision, recall, F1-score	Measure Authentic / Fake decisions per input type
Adaptability	Reward convergence; performance on new patterns versus a static baseline	Test the response to concept drift
Cross-modal analysis	Detection of text–media inconsistencies	Measure the value of cross-modal evidence
Integrity protection	Hash mismatch detection on modified content	Confirm detection of tampering
Blockchain overhead	Transaction latency, storage growth	Assess scalability of record keeping
Batch processing	Processing time per item	Assess efficiency for large collections

Classification quality will be separately reported for text, image and audio inputs to see the contribution of each modality. For a comparison of adaptability, a previously unseen misinformation pattern will be introduced and the RL based verification policy compared with a static classifier, while the convergence of the DRVGA reward will be tracked. Content will be altered to test integrity protection and then the integrity check of the content stored in the digest will be checked for. The overhead of Blockchain will be assessed in terms of transaction latency, the amount of storage needed as the number of verification records grows to the extent that the Blockchain can scale to real world use.

VIII. APPLICATIONS

The framework can help in various areas that promote the authenticity, integrity, traceability, and reliability of digital media. A hash-based fingerprint and blockchain records would offer digital forensics with a verifiable proof that content is not altered after submission. Journalism: Editors have the ability to confirm the accuracy of claims, images, and audio before it goes to press, and record who has checked what and how. Journalism: Editors are able to confirm the accuracy of claims, images and audio prior to publication, and maintain an auditable record of the checks performed. Batch processing: On social media it's possible to analyze large amounts of posts and the preview of the content makes it easy for readers to see which part of a post is suspect. Within secure digital content verification in general, entities can verify documents and media based on a repository of documents that they know are tamper resistant rather than a single central repository.

IX. CONCLUSION AND FUTURE WORK

A. Conclusion

This paper proposed a novel approach combining NLP, RL and the Dynamic Reward-Guided Verification Algorithm, data authentication, and Blockchain technology in a Hybrid Decentralized Framework for Media Integrity Verification. As opposed to traditional methods that rely on static classification model and central storage, the framework supports text, URL, file, image, audio and batch input, learns and changes its verification policy based on credibility, and logs verification results in a tamper-resistant, traceable way. The output of each stage is aggregated together to generate a 'credibility score' and a final 'Authentic' or 'Fake' result and supporting information; a 'content preview' is provided as a reference and/or suggested version of the content, if any, however this does not constitute a guaranteed fact unless verified independently.

The review of fifteen related works demonstrated that the majority of existing works focus on either text or image data, are static after training, and rarely offer the secure provenance of multimodal data, whereas the proposed combination of multimodal analysis, adaptive verification and decentralized records fills the above gaps. The next step is implementation of the framework as well as a quantitative evaluation.

B. Future Work

There are a number of enhancements that can be made to improve functionality, accuracy, scalability and real-world applicability:

• Advanced manipulated media: Expand to include deepfake videos, AI-generated photos and more complex manipulated audio.

• Advanced multimodal models: Compare text, images, sounds, and videos together to identify inconsistencies between the media more accurately.

• Multilingual support: Include multilingual NLP models to be able to verify content from various linguistic communities and regional-language sources.

• Federated learning: Enable the sharing of privacy-sensitive raw data between models, enhancing privacy.

• Better RL and DRVGA: Regularly review and update RL policy and DRVGA reward system based on patterns in misinformation and feedback.

• Blockchain optimization: Minimize blockchain delays and optimized verification hashes and audit information storage.

• Explainable AI (XAI) : Identify the language, context, multimedia and/or integrity-related factors that influenced a decision.

• Continuous improvement: Expand on the use of varied data sets, frequent model updates and ongoing assessment of new misinformation strategies.

REFERENCES

[1] "A fake news detection framework integrating multi-domain and multimodal features," ScienceDirect – Elsevier, 2025.

[2] "Multimodal matching-aware co-attention networks with mutual knowledge distillation for fake news detection," Information Sciences, ScienceDirect, 2024.

[3] "CAF-ODNN: Complementary attention fusion with optimized deep neural network for multimodal fake news detection," Information Processing & Management, ScienceDirect, 2024.

[4] "Cross-attention-enhanced multimodal fake news detection using autoencoder-based fusion and Transformer-based models," Procedia Computer Science, ScienceDirect, 2025.

[5] "Dual-stream fusion network with multi-head self-attention for multi-modal fake news detection," Applied Soft Computing, ScienceDirect, 2024.

[6] "Multimodal fake news detection based on cross-modal message aggregation and gated fusion network," ScienceDirect, 2024.

[7] "SARD: Fake news detection based on CLIP contrastive learning and multimodal semantic alignment," J. King Saud Univ. – Comput. Inf. Sci., ScienceDirect, 2024.

[8] "Multimodal social media fake news detection based on similarity and inference adversarial networks," Computers, Materials & Continua, ScienceDirect, 2024.

[9] "Defending against misinformation: Evaluating Transformer architectures for quick misinformation detection on social media," Procedia Computer Science, ScienceDirect, 2024.

[10] "GBERT: A hybrid deep learning model based on GPT-BERT for fake detection," Heliyon, ScienceDirect, 2024.

[11] "Entity-centric multi-domain Transformer (ECMT) for improving generalization in fake news detection," Information Processing & Management, ScienceDirect, 2024.
[12] "Analyzing and detecting the fake news using machine learning," in Proc. ICICAT, IEEE, 2024.
[13] "Hybrid optimization driven fake news detection using reinforced and Transformer models," Scientific Reports, Nature, 2025.
[14] "A multimodal fusion and ensemble approach for robust fake news detection through deep learning, reinforcement learning, and using blockchain to enhance security and authorization," IEEE, 2025.
[15] "A review of deep learning-based approaches for deepfake content detection," Expert Systems, Wiley, 2024.