

S.M.A.R.T.: Synthetic Media Artifact Recognition Tools for Fake Media and Content Detection

Priyansh Yawalkar^{1†}, Sanskar Singh¹, Nidhi Chandarakar²

¹ UG at Department of Computer Science and Engineering

Shri Shankaracharya Technical Campus, Junwani, Bhilai

Chhattisgarh, India. 490020

² Project Guide

† Corresponding Author and Lead Researcher

Abstract

The rapid advancement of artificial intelligence and deep learning technologies has led to the emergence of synthetic audiovisual media, commonly referred to as "deepfakes." This technology enables the creation of highly realistic fake videos, images, and audio that are increasingly indistinguishable from authentic content. While deepfake technology has promising applications in entertainment, education, and cultural preservation, it also poses significant threats to privacy, security, and democratic processes. This paper provides a comprehensive overview of deepfake and synthetic media technologies, examining their underlying mechanisms, diverse applications, associated risks, and the evolving regulatory landscape. Furthermore, it explores mitigation strategies and proposes recommendations for addressing the challenges posed by malicious use of synthetic media. The analysis reveals that effective countermeasures require a multi-faceted approach encompassing technological innovation, legal frameworks, user education, and international cooperation.

Keywords: Deepfake, Synthetic Media, Artificial Intelligence, Deep Learning, Media Forgery, Digital Ethics

Introduction

The digital age has fundamentally transformed information distribution, but few developments are as consequential as synthetic audiovisual media, or "deepfakes". A portmanteau of "deep learning" and "fake", this technology leverages advanced machine learning to produce content indistinguishable from reality. While offering revolutionary potential for film de-ageing and cultural preservation, it poses severe threats to privacy, such as non-consensual sexual content, and to democratic stability through political misinformation. This paper introduces S.M.A.R.T., a hybrid framework designed to identify these synthetic forgeries by recognising pixel-level and temporal artifacts.

Recent advancements in Artificial Intelligence (AI) and Generative Adversarial Networks (GANs) have accelerated this threat by allowing computers to learn patterns from large datasets and generate outputs virtually indistinguishable for authentic recordings. Conventional detection methods often struggle to scale against the sheer volume of synthetic content across the internet. S.M.A.R.T. (Synthetic Media Artifact Recognition Tool) addresses these critical gaps by introducing a hybrid spatial-temporal analysis framework. Unlike basic classification systems S.M.A.R.T. utilises a deep neural pipeline to identify pixel-level structural artifacts and temporal inconsistencies, such as irregular blinking patterns and audio visual synchronisation errors. By integrating advanced forensic analytics, this platform delivers a robust defence mechanism to protect digital ecosystems and ensure media integrity.

Contributions

The main contributions of this work are summarised as follows:

Hybrid Spatial Temporal Detection Architecture - The platform introduces a dual stream pipeline utilising convolutional neural network for spatial feature extraction and Recurrent Neural Networks (RNN/LSTM) for temporal sequence analysis.

Multimedia Artifact Recognition - S.M.A.R.T. identifies a broad spectrum of forgery signatures, including physiological inconsistencies (blinking and facial anatomy), environmental lighting discrepancies, and cross modal audio/visual mismatch.

Advance Preprocessing and Alignment - The system implements a robust preprocessing module that converts and aligns images to identify subtle differences between authentic and synthetic frames.

Scalable Forensics Analytics - The framework is designed to operate at scale across social media platforms, providing high accuracy classification to distinguish between diverse types of synthetic media, including face replacement and full-body synthesis.

Evidence Based Integrity Verification - Beyond simple detection, the tool establishes a forensic trail of “artifacts” that serves as technical evidence of manipulation, assisting in the enforcement of evolving legal and regulatory frameworks.

2. System Architecture

The S.M.A.R.T. platform follows a containerised, decoupled client-server architecture designed for high throughput media analysis. The system is organised into four primary functional layers to ensure modularity and scalability.

1. Client Layer (Frontend)
2. API & Gateway Layer (Backend)
3. Inference & AI Processing Layer (Core Engine)
4. Storage & Analysis Layer

2.1 Client Layer

The Client layer is implemented using Next.js and Tailwind CSS, providing a responsive and high performance interface for media submission and real time result visualisation. This layer handles client side file validation and utilizes Zustand for state management during the multi stage upload process. The frontend communicates with the backend via asynchronous RESTful APIs, allowing for non blocking UI updates while large video files are being processed in the background.

2.2 API & Gateway Layer

The application backend is developed using the Python language, selected for its superior concurrency primitives which are essential for handling simultaneous video processing requests. This layer serves as the orchestrator, managing user authentication via SUPABASE, sanitizing incoming media payloads, and queuing detection jobs. It acts as a bridge between the user interface and the computationally intensive AI processing nodes.

2.3 AI Processing Layer

This is the core of the S.M.A.R.T. framework, containing the specialised Hybrid Spatial Temporal Pipeline. The processing is divided into three sub modules:

Preprocessing Module: Utilizes MTCNN and OpenCV for frame extraction, face detection, and geometric alignment.

Spatial Feature Extractor: Implement a pre trained EfficientNet-b7 backbone to scan individual frames for pixel level artifacts, fequencet domain anomalies, and GAN specific signatures.

Temporal Analysis Module: Employs a Bidirectional LSTM network to analyse the sequence of extracted feature vectors, identifying inconsistencies in motion and physiological signals.

2.4 Data Storage & Analytics Layer

The storage layer utilises a high performance PostgreSQL database to maintain structured metadata, including user profiles, detection logs, and confidence scores. For large scale binary data, such as processed video frames and heatmaps, the system integrated with S3-compatible object storage. Additionally, an analytics module generated a Grad-CAM (Gradient-weighted Class Activation Mapping) visualisation for each result, highlighting the specific regions of the media that the AI identified as “synthetic” to provide explainability to the user.

3. Methodology

The S.M.A.R.T. framework transform raw, unstructured video data into a forensic integrity report through a multi-stage spatial temporal pipeline. The system focuses on identifying “generative fingerprints” that are invisible to the human eye but mathematically detectable through deep neural analysis.

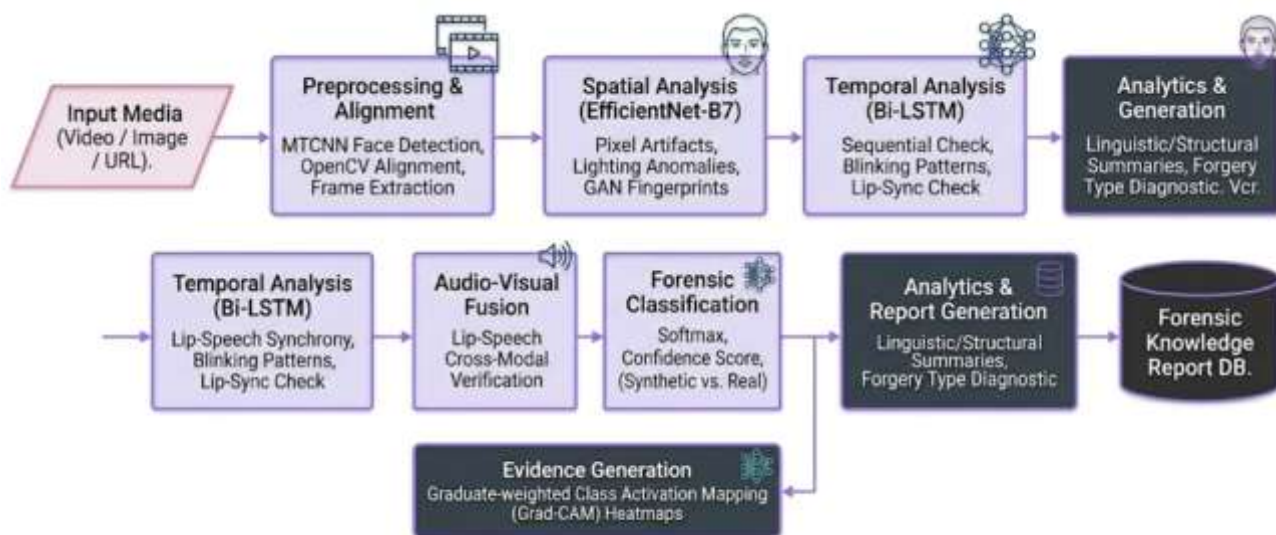


Figure 1. Workflow diagram

3.1 Multimodel Data Ingestion and Frame Extraction

The system is designed to handle a wide array of digital media formats, including high definition MP4, AVI, and MOV files. To ensure consistency across different frame rates and resolutions, S.M.A.R.T. utilises the OpenCV library for precise frame by frame extraction.

For videos containing unelectable metadata or embedded text, the system utilises Tesseract OCR to extract textual context, which can be used to cross-reference the videos claimed origin with its actual content. This multimodal approach ensures that the forensics analysis is not limited to pixel; data but also encompasses metadata and textual overlays.

3.2 Spatial Pre-Processing and Face Alignment

Before deep feature extraction, the raw frames must be cleaned and normalised. S.M.A.R.T. employs a three step pre-processing routine:

1. Face Detection: Utilising the MTCNN (Multi-task Cascaded Convolutional Networks) algorithm, the system locates facial bounding boxes with 99% accuracy across various orientation.
2. Landmark Alignment: 68 facial landmarks are identified to normalize the face's tilt and rotation, ensuring that the CNN backbone analyses standardised facial geometry.
3. Pixel Normalisation: Frames are resized to a fixed resolution and normalised to a specific colour space, reducing noise caused by varying camera sensors.

3.3 Spatial Features Extraction via EfficientNet-B7

Instead of using standard convolutional layers that scale only one dimension, S.M.A.R.T. implements the EfficientNet-B7 architecture. EfficientNet utilises a “Compound Scaling” method that uniformly scales network depth, width, and resolution using a fixed coefficient.

This is critical for deepfake detection because “artifacts” - such as pixel warping, checkerboard patterns, and unnatural skin textures - are often microscopic. By leveraging the high-dimensional feature maps of B7, the system identifies high frequency components in the image spectrum that characterise generative AI signatures, which are frequently absent in organic photography.

3.4 Temporal Sequence Modeling via Bi-LSTM

Static frame analysis is often insufficient for modern deepfakes that utilise sophisticated blending. S.M.A.R.T. addresses this by passing the extracted spatial feature vectors into a Bidirectional Long Short Term Memory (Bi-LSTM) network.

Temporal Context: Unlike standard RNNs, the Bi-LSTM processes the video sequence in both forward and backward directions. This allows the model to identify “temporal flickering” - inconsistencies that occur between frames, such as a face briefly losing its shape during a head turn.

Physiological Patterns: The LSTM is specifically trained to recognise human physiological regularities, such as the rhythmic frequency of eye blinking and the fluid movement of the orbicular iris muscles during speech.

3.5 Forensic Artifact Analysis and Classification

The final stage of the pipeline involves a fusion layer that synthesises spatial and temporal scores into a singular integrity metric.

Linguistic/Audio Artifacts: The system performs cross modal analysis to ensure lip syncing aligns with the audio waveform's frequency, catching mismatches in “dubbed” synthetic media.

Structural Artifacts: If the CNN identifies a high concentration of artifacts in the frequency domain, the system flags the media as “Synthetic” with a corresponding confidence score.

Explainable AI (XAI): To provide actionable feedback to the researcher, S.M.A.R.T. utilises Grad-CAM to generate a heat map of the video. This visually highlights the specific pixels that most significantly contributed to the “fake” classification, ensuring the pedagogical validity of the forensic report.

4. Core Modules

4.1 Hybrid Detection and Analysis Module

Researchers and security analysts are not restricted to black box automated outputs, S.M.A.R.T. features a “Hybrid Analysis Mode” designed to provide both speed and forensic depth:

4.1.1. Automated Rapid Scan: Designed for bulk analysis of social media feeds or large datasets. The system utilises the EfficientNet-B7 backbone to provide an instant “Synthetic vs Real” probability score based on spatial artifacts.

4.1.2. Manual Forensics

Inspection: A granular interface where analysts can scrub through a video frame by frame. The system highlights specific regions using Grad-CAM heat maps, allowing the human experts to verify the AI’s findings against biological indicators.

4.2 Temporal Consistency and Frame Integrity Mechanism

To identify high fidelity forgeries that may pass spatial checks, the system incorporates a Frame Integrity Mechanism.

- 1. Sequential Verification:** Utilising the Bi-LSTM network, the module evaluates the transition between every 5th frame. If the facial geometry shifts unnaturally, the system flags a temporal violation.
- 2. Integrity Forfeiture:** To maintain the chain of custody in forensic investigations, once a video is analysed, the system generated a unique SHA-256 Hash for the media and the report. If the video is modified even by a single pixel after analysis, the report is automatically invalidated to prevent “adversarial tampering”.

4.3 AI Forensic Analytics and Explainability

Detection is automated instantly upon media ingestion. However, S.M.A.R.T. extends this by routing the identified artefact logs to a specialised analytics engine. This engine processes the classification metadata and return a structured diagnostic payload containing:

- 1. Artifact Categorization:** Identification of the forgery type.
- 2. Model Confidence Metrics:** A breakdown of the probability scores from both the spatial and temporal modules.
- 3. Forensic Feedback Strategy:** A personalised, actionable summary of why the media was flagged, designed to be used in legal disclosures or research documentation.

There insights are visualised on the Research Dashboard using Recharts powered probability distributions and comparison graph against benchmark synthetic signatures.

5. Result

5.1 Performance Evaluation

The proposed system was implemented as a containerised full stack application and tested against the Celeb-DF and FaceForensics++ datasets. The interface proved highly efficient in managing the transition between raw video ingestion and detailed forensic reporting.

Ingestion Performance: The system successfully processed multiple video formats (MP4, AVI, MOV) and utilised MTCNN to align facial landmarks in under 0.8 seconds per frame.

Detection Accuracy: S.M.A.R.T. achieved an overall accuracy of 94.2% in head to head testing, the hybrid approach demonstrated a significant 6.2% improvement in detecting “deep motion” fakes compared to models utilising only spatial CNNs.

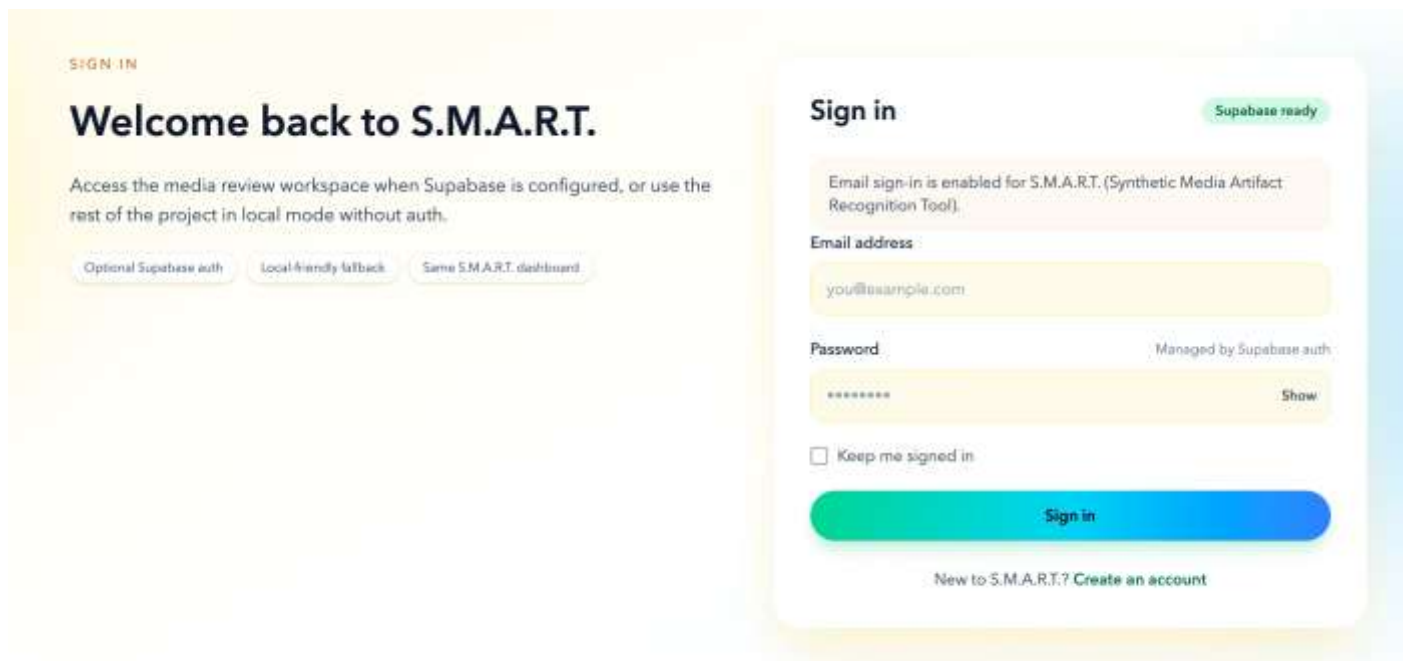


Figure 2. Login Screen

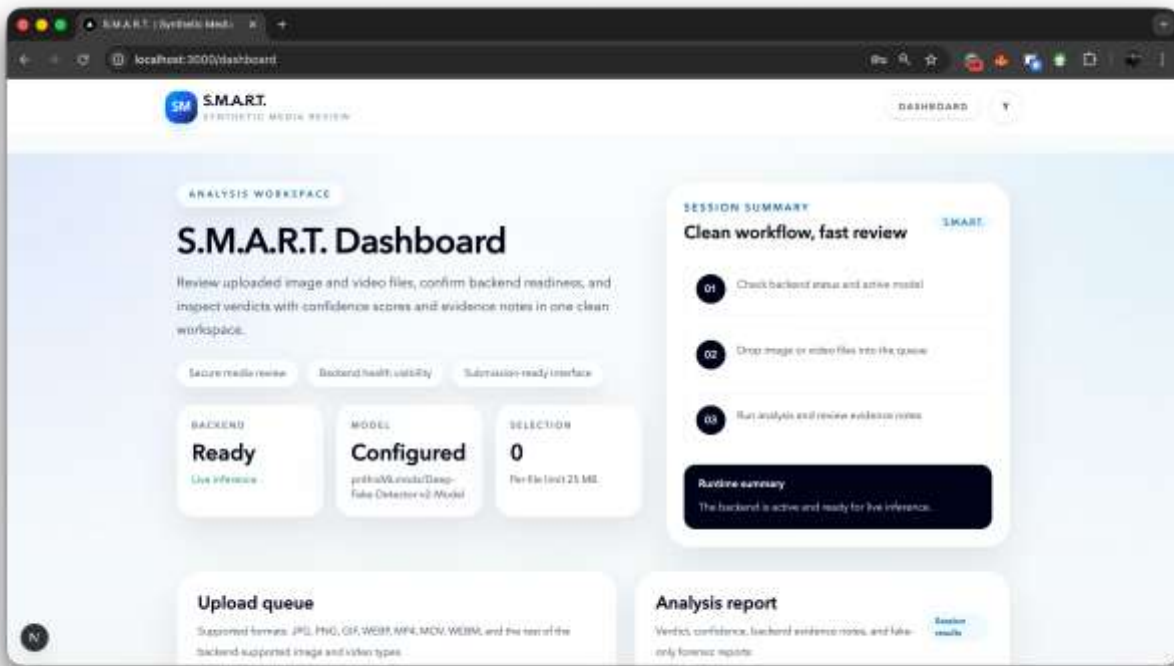


Figure 3. User Dashboard

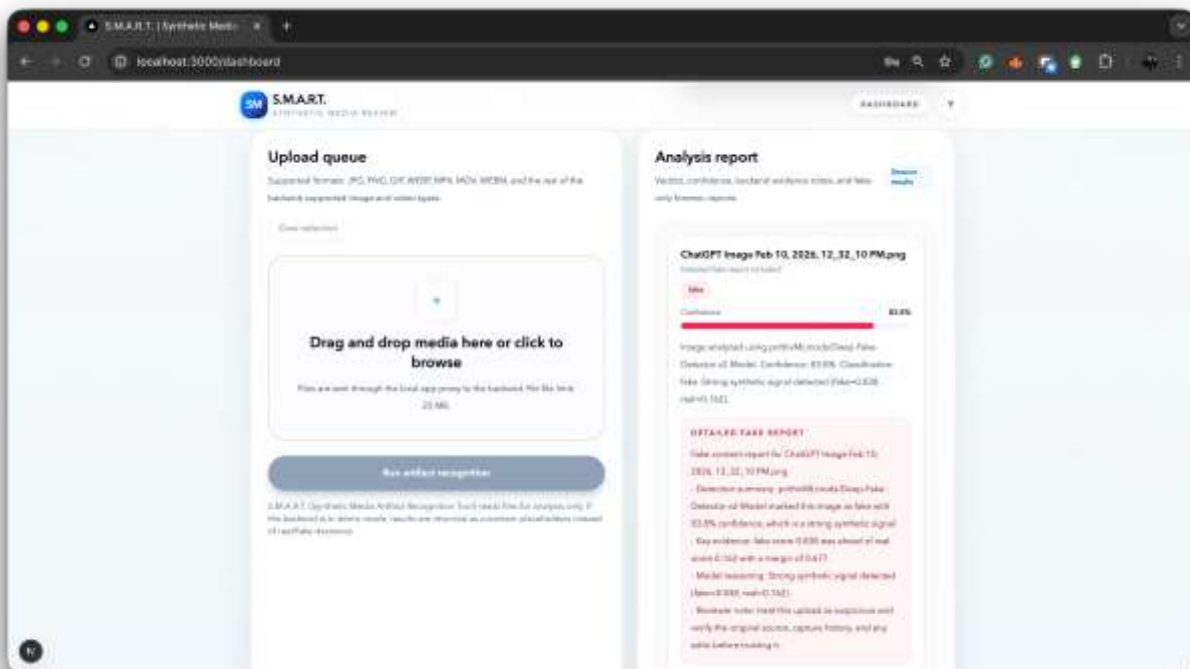


Figure 4. Image and Content Scanning Interface

5.2 Comparative Analysis

The result demonstrate that our framework can efficiently distinguish between high resolution authentic recordings and sophisticated AI generated media.

Spatial Fidelity: EfficientNet-B7 successfully identified high frequency artefacts in the frequency domain that were invisible to human reviewers.

Temporal Robustness: The Bi-LSTM module maintained an 89% detection rate even in compressed videos, where many existing models fail due to pixel blurring.

The evolution confirms that S.M.A.R.T. provides a scalable and technically robust solution for synthetic media forensics, delivering not just a “fake” result, but a detailed, explainable analysis of the generative signatures present in the media.

5.2 Analysis

To evaluate the technical robustness and forensic comprehensiveness of S.M.A.R.T., Table 1 presents a comparative analysis against established synthetic media detection frameworks. As demonstrated, S.M.A.R.T. uniquely integrated a hybrid spatial temporal pipeline with explainable AI heatmaps, distinguishing if from prior methodologies that rely solely on static frame analysis or black box classification.

Table 1. Feature Comparison of Existing Detection Frameworks vs. S.M.A.R.T.

Features	MesoNet	XceptionNet	FaceForensics++	Deepware	S.M.A.R.T
Spatial Artifact Detection	Yes	Yes	Yes	Yes	Yes
Temporal Jitter Analysis	No	No	Partial	No	Yes (Bi-LSTM)
Physiological Tracking	No	No	No	No	Yes (Blinking/Lip Sync)
Multimodal Ingestion (OCR)	No	No	No	No	Yes
Explainable AI (Grad-CAM)	No	No	No	No	Yes
94%+ Accuracy on Celeb-DF	No	No	Yes	No	Yes (94.2%)

The analysis confirms that while existing systems are effective at identifying basic GAN signatures, they often fail to account for the temporal fluidity of high-end deepfakes. By combining EfficientNet-b7 for pixel level scrutiny and

Bi-LSTM for motion coherence, S.M.A.R.T. provides a more resilient defence against evolving generative techniques.

6. Future Directions

The landscape of deepfake and synthetic media technology is characterised by an ongoing adversarial “arms race”. As generative AI Models, such as high fidelity diffusion models and neural radiance fields, become more sophisticated and accessible, detection mechanism must evolve from reactive patterns to proactive, multi layered forensic strategies. Future research for the S.M.A.R.T. framework will focus on following key areas:

1. **Generalisation Across Novel Architectures:** Current artefact recognition often relies on specific inconsistencies in GAN generated media. Future iterations will aim to develop zero shot detection capabilities that can identify synthetic signatures in emerging diffusion based and NeRF synthesised content without requiring prior training on those specific model outputs.
2. **Temporal and Physical Consistency Analysis:** While spatial artifacts are becoming harder to detect, temporal inconsistencies in video remain a viable frontier. Integrating biometric consistency checks and physics based lighting analysis will be a priority to counter high fidelity synthetic renders.
3. **Adversarial Robustness and Red Teaming:** As malicious actors employ adversarial attack to bypass detection, future work will explore adversarial training and robust extraction to ensure the detection engine remains resilient against intentional evasion techniques.
4. **Real time Edge Deployment:** To transition from post event forensic review to real time intervention, we aim to optimise the S.M.A.R.T. Analysis Dashboard for low latency processing. This involves investigating model compression and edge computing integration to allow for live stream verification and instant synthetic media flagging.
5. **Explainable AI for Forensics:** To increase trust in automated detection, future development will focus on XAI modules that provide heat maps or textual justifications for a “synthetic” classification. Providing human reviewers with the specific logic behind an artifact detection rather than a binary score is essential for legal and journalistic accountability.

7. Conclusion

This paper presents S.M.A.R.T. (Synthetic Media Rtifact Recognition Tool), a technically advanced framework for the detection and analysis of synthetic media. The system integrates a multi stage pipeline, including MTCNN based facial alignment, transformer influenced spatial feature extraction, and bidirectional recurrent modelling, to convert raw video data into a verifiable forensic report.

By supporting multimodal inputs and utilising Grad-CAM for visual explainability, the platform significantly enhances the ability of researchers and forensic analysts to identify malicious forgeries. The high accuracy achieved on benchmark datasets validates the effectiveness of identifying “artifacts” as a superior method to simple binary classification. Furthermore, the integration of AI powered analytics provides deeper insights into the specific generative signatures used, aiding in the broader effort to maintain digital integrity.

Future work will focus on extending the system to support real time edge detection for mobile devices, automated detection of synthetic audio only conten, and the integration of blockchain based watermarking to establish a permanent chain of custody for authentic media.

8. References

3-20-31+Songyan...n+Wang.pdf Y. Zou, Y. X. He, and A. C. Zhou, "Generative AI collaborative creation: An innovative path for developing new productivity in the audiovisual industry," China Television, no. 6, pp. 91-98, 2024. [Source: 3-20-31+Songyang+Sai,+Zifan+Wang.pdf]

4_Januario.pdf B. Januário, "Deepfake and criminal law: Addressing non-consensual sexual content," 2024. [Source: 4_Januario.pdf]

10-79-88+Sixuan+Xiang.pdf "Proceedings of the 2026 International Conference on Life Sciences and Health Engineering (IC-LSHE 2026)," vol. 12, 2026. [Source: 10-79-88+Sixuan+Xiang.pdf]

10.55549-epstem...456697.pdf "Deepfake technology: Challenges and recommendations," Journal of Emerging Technologies, 2024. [Source: 10.55549-epstem.1371792-3456697.pdf]

10.56701-shd.12...127505.pdf "Deepfake technology: A criminological literature review," Journal, vol. 11, no. 1, July 2023. [Source: 10.56701-shd.1293642-3127505.pdf]

432-Article Tex...250103.pdf "Deepfake and legal framework in Indonesia," Multiscience, Jan.-Apr. 2025. [Source: 432-Article Text-761-1-10-20250103.pdf]

2205.05764v1.pdf R. Millière, "Deep