A Transformer-Based Framework for Detecting AI-Generated Images and Videos for Digital Media Authentication

[1]Sivalakshmi M, [2]N Saranya
[1]Student, B.Sc Artificial Intelligence and Machine Learning, Final year, Nehru Arts and Science College, Coimbatore, Tamil Nadu, India
[2]Head of the Department, Iot and AIML,, Nehru Arts and Science College, Coimbatore, Tamil Nadu, India

Abstract:
The advancement of artificial Intelligence and deep getting to know technology has considerably modified the way virtual media is created and shared across on-line systems. present day generative models are able to generating noticeably realistic photos, motion pictures, and audio that carefully resemble true content material, making it increasingly hard to distinguish among actual and manipulated media. whilst those technology guide creativity, automation, and virtual innovation, additionally they introduce critical challenges inclusive of incorrect information, identity manipulation, deepfake attacks, and digital fraud. As AI-generated content material will become more common, making sure the authenticity and reliability of digital media has become an crucial difficulty in today's digital environment. the present examine specializes in the development of an AI reality Checker, an clever system designed to come across manipulated or AI-generated media content. The proposed system makes use of synthetic intelligence strategies and picture processing strategies to investigate visual patterns, structural inconsistencies, and hidden artifacts present in digital pictures. by using applying system mastering-based totally detection mechanisms, the gadget ambitions to pick out whether the given media content material is real or artificially generated. The implementation of this device demonstrates that AI-primarily based verification gear can efficiently support the detection of manipulated media and assist lessen the spread of deceptive statistics. The have a look at emphasizes the importance of growing reliable AI-pushed verification frameworks to protect virtual identities, beautify information credibility, and keep agree with in on line media platforms.

Keywords: *Artificial Intelligence, Deepfake Detection, Digital Media Authentication, Image Processing, AI-Generated Content Detection*

I. INTRODUCTION

The rapid growth of digital technology and artificial intelligence has significantly transformed the way information and media are created, shared, and consumed across online platforms. In recent years, advanced AI techniques, particularly deep learning and generative models, have enabled the creation of highly realistic images, videos, and audio content. While these technologies have contributed positively to areas such as entertainment, education, digital design, and media production, they have also introduced serious concerns related to the misuse of AI-generated content. One of the major challenges emerging from this technological advancement is the rise of manipulated media, commonly referred to as deepfakes.

Deepfake technology uses artificial intelligence algorithms to modify or generate digital media that appears authentic but is actually fabricated. Such manipulated content can be used for spreading misinformation, impersonating individuals, creating fake news, and conducting various forms of digital fraud. With the increasing accessibility of AI-based content generation tools, the detection of manipulated media has become increasingly difficult for the general public. As a result, the reliability and authenticity of digital information are being questioned in many sectors including media, politics, cybersecurity, and social communication.

To address these challenges, researchers and technology developers are focusing on designing intelligent systems that can automatically identify AI-generated or manipulated media. Artificial intelligence-based detection systems analyze patterns, inconsistencies, and hidden artifacts present within images and videos to determine their authenticity. These systems make use of machine learning algorithms, image processing techniques, and data analysis methods to identify subtle irregularities that may not be visible to the human eye.

The present study proposes the development of an AI Reality Checker, an intelligent framework designed to detect manipulated or AI-generated media content. The system aims to enhance digital media verification by applying artificial intelligence techniques to analyze and validate the authenticity of images. By identifying signs of digital manipulation, the system helps reduce the spread of misleading content and promotes trustworthy information sharing in digital environments. The implementation of such detection mechanisms is essential for strengthening digital security, protecting individual identity, and maintaining public trust in online media platforms.

II. PROBLEM STATEMENT AND OBJECTIVE

With the rapid advancement of artificial intelligence and digital media technologies, the creation of highly realistic images and

videos has become easier than ever. Modern AI-based generative models are capable of producing synthetic media that closely resembles real content. While these technologies offer many benefits in areas such as entertainment, design, and digital communication, they also create serious challenges related to the misuse of manipulated media. AI-generated images and deepfake content can be used to spread misinformation, impersonate individuals, create misleading content, and manipulate public opinion. In many cases, it becomes extremely difficult for ordinary users to distinguish between authentic media and artificially generated content, as the differences are often subtle and not easily visible.

The increasing presence of manipulated digital media has raised concerns about information credibility, online trust, and digital security. Traditional methods of detecting manipulated images often rely on manual inspection or basic image analysis techniques, which may not be effective against advanced AI-generated content. Furthermore, the continuous development of new generative models makes it challenging for existing detection methods to keep up with the rapidly evolving nature of artificial intelligence technologies. These challenges highlight the need for intelligent systems that can automatically analyze digital media and identify potential signs of manipulation.

In response to these issues, the present study aims to develop an AI Reality Checker, a system designed to detect manipulated or AI-generated images using artificial intelligence and image processing techniques. The primary objective of the proposed system is to analyze digital images, extract meaningful visual features, and apply machine learning algorithms to determine whether the content is authentic or artificially generated. By automating the detection process, the system aims to provide a reliable and efficient method for verifying the authenticity of digital media.

III. Related Work

With the growing use of digital media and artificial intelligence technologies, the issue of manipulated and AI-generated content has gained significant attention from researchers. The ability of modern AI models to create highly realistic images and videos has raised concerns about misinformation and digital authenticity. As a result, many studies have focused on developing techniques that can automatically detect manipulated media using machine learning, deep learning, and image analysis methods.

One of the early contributions in this field was made by Afchar et al. (2018), who introduced the MesoNet model for detecting facial manipulation in images and videos. The study demonstrated that deep learning models are capable of identifying small visual inconsistencies and hidden artifacts that are usually difficult for humans to notice. Later, Rossler et al. (2019) developed the FaceForensics++ dataset, which provides a large collection of manipulated and real videos. This dataset has been widely used by researchers to train and evaluate deepfake detection systems.

Further studies have explored different deep learning approaches for improving detection performance. Tolosana et al. (2020) reviewed various deepfake detection techniques and highlighted the importance of convolutional neural networks in analyzing facial features and identifying unnatural patterns in digital media. Similarly, Verdoliva (2020) discussed the growing challenges of detecting deepfake content and emphasized the need for stronger digital forensic tools to maintain trust in online information.

More recent research has focused on improving detection accuracy and making systems more reliable. Dang et al. (2021) proposed a deep learning framework that analyzes both spatial and temporal features of videos to detect manipulation more effectively. In another study, Chandrasegaran et al. (2022) introduced explainable artificial intelligence methods that allow detection models to highlight the specific areas of an image where manipulation may have occurred, making the detection process more transparent.

Recent developments between 2023 and 2024 have focused on detecting content generated by advanced models such as Generative Adversarial Networks (GANs) and diffusion-based image generators. Researchers have explored hybrid approaches that combine machine learning, image analysis, and digital forensics to detect inconsistencies in texture, lighting, and image patterns. These techniques have shown promising results in identifying AI-generated media.

Despite the progress made in this field, the detection of manipulated media continues to be a challenging task because generative AI technologies are constantly improving. Therefore, developing intelligent and reliable systems for verifying digital media is becoming increasingly important. The proposed AI Reality Checker aims to contribute to this effort by using artificial intelligence techniques to analyze images and assist in determining whether the content is authentic or artificially generated.

IV. System Design

The system design of the proposed AI Reality Checker focuses on building a structured and efficient framework that can analyze digital images and determine whether the content is authentic or artificially generated. With the rapid advancement of artificial intelligence technologies, manipulated media such as deepfakes and AI-generated images have become increasingly common. As a result, it is essential to design systems that can automatically verify the authenticity of digital content. The proposed system is designed to process uploaded images, perform necessary preprocessing operations, analyze important visual features, and provide a prediction regarding the authenticity of the media. The system is organized into several modules, each performing a specific function in the detection process.

4.1 Input Module

The input module serves as the entry point of the system where users provide the image that needs to be verified. A simple and interactive interface is designed to allow users to upload digital images easily. The interface ensures that the process remains accessible even for users who may not have technical knowledge about artificial intelligence or digital forensics. Once the image is uploaded, the system temporarily stores the file and prepares it for further processing. This module plays an important role in establishing communication between

 ISSN: 3107-6513

the user and the detection system while ensuring that the input data is properly received.

4.2 Preprocessing Module

After the image is uploaded, the system performs preprocessing operations to prepare the image for analysis. Images obtained from different sources may vary in size, resolution, and format, which can affect the performance of the detection model. Therefore, preprocessing is required to standardize the input data. During this stage, the image may be resized to a fixed dimension, normalized to maintain consistent pixel values, and cleaned to remove unwanted noise. These operations ensure that the image data is consistent and suitable for further processing. Proper preprocessing improves the accuracy of the detection model and helps the system analyze the image more effectively.

4.3 Feature Analysis Module

The feature analysis module focuses on examining important visual characteristics present in the image. Digital images contain several elements such as textures, edges, color distributions, and structural patterns. AI-generated or manipulated images often contain subtle inconsistencies in these features, which may not be easily visible to the human eye. In this stage, image processing techniques are used to extract meaningful features that can help identify such irregularities. The system carefully analyzes these patterns and prepares them as input for the machine learning model. This module plays a crucial role in identifying the hidden characteristics that differentiate real images from artificially generated ones.

4.4 Prediction Module

The prediction module is responsible for determining whether the uploaded image is authentic or AI-generated. In this stage, the extracted features are processed by a trained machine learning or deep learning model. The model analyzes the patterns within the image and compares them with patterns learned during its training phase. Based on this comparison, the system calculates the probability that the image may have been manipulated or generated using artificial intelligence techniques. The prediction module is considered the core component of the system because it performs the main analysis required for media verification.

4.5 Result Display Module

The final stage of the system design is the result display module, which presents the analysis outcome to the user. After the prediction process is completed, the system generates a result indicating whether the uploaded image is likely to be authentic or AI-generated. The result is displayed in a clear and simple format so that users can easily understand the authenticity status of the media. Providing an understandable output is important because it allows users to quickly interpret the results and make informed decisions about the credibility of the digital content.

Overall, the proposed system design follows a systematic approach for detecting manipulated digital images. By combining user interaction, image preprocessing, feature extraction, machine learning-based prediction, and result visualization, the AI Reality Checker provides an effective solution for verifying the authenticity of digital media. This structured design also allows the system to be further improved and expanded with more advanced detection techniques in the future.

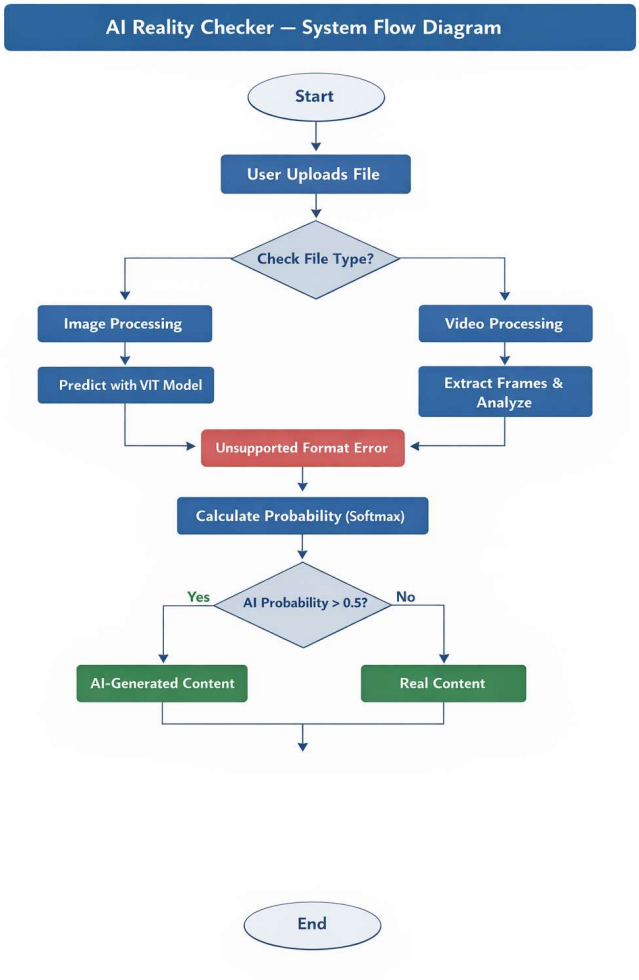

Fig. !. System Design of the Proposed Model

V. ADVANTAGES OF PROPOSED SYSTEM

The proposed AI Reality Checker system offers several advantages in detecting manipulated or AI-generated digital images. One of the major benefits of the system is its ability to automatically analyze images and determine their authenticity using artificial intelligence techniques. This reduces the need for manual inspection and helps users quickly verify whether the digital content is real or artificially generated.

Another advantage of the system is its user-friendly interface, which allows users to upload images and receive results without requiring technical knowledge. The simple interaction process makes the system accessible to a wide range of users, including researchers, students, and general internet users who want to verify the authenticity of digital media.

The system also improves detection accuracy by analyzing hidden visual patterns and inconsistencies that may not be easily visible to the human eye. By using machine learning models and image processing techniques, the system can identify subtle artifacts that often appear in AI-generated images.

In addition, the proposed system follows a modular design, where different components such as preprocessing, feature extraction, prediction, and result generation work together. This modular structure makes the system easier to maintain and update. New detection models or improved algorithms can be integrated into the system in the future without redesigning the entire framework.

Another important advantage is that the system supports efficient and fast analysis of digital images. The automated process allows images to be analyzed quickly, making the system useful for verifying media content in real-time environments.

VI. CONCLUSION

The rapid advancement of artificial intelligence and digital media technologies has created both opportunities and challenges in the modern digital environment. While AI-based tools enable the creation of highly realistic images and videos, they also increase the risk of manipulated content, misinformation, and misuse of digital identities. As a result, the need for reliable systems that can verify the authenticity of digital media has become increasingly important.

The proposed AI Reality Checker system provides an intelligent approach for detecting manipulated or AI-generated images using artificial intelligence and image analysis techniques. The system follows a structured process that includes image input, preprocessing, feature extraction, and machine learning-based prediction to determine the authenticity of digital media. By analyzing hidden patterns and inconsistencies within images, the system helps identify content that may have been artificially generated or altered.

The implementation of this system demonstrates that artificial intelligence can play an important role in improving digital media verification and reducing the spread of misleading information. The user-friendly interface and modular design of the system make it accessible and adaptable for future improvements. Overall, the AI Reality Checker contributes to enhancing trust in digital information and supports the development of more secure and reliable digital communication platforms.

VII. REFERENCES

1. Afchar, D., Nozick, V., Yamagishi, J., and Echizen, I., "MesoNet: A Compact Facial Video Forgery Detection Network," *Proceedings of the IEEE International Workshop on Information Forensics and Security*, pp. 1–7, 2018.
2. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., and Nießner, M., "FaceForensics++: Learning to Detect Manipulated Facial Images," *IEEE International Conference on Computer Vision (ICCV)*, pp. 1–11, 2019.
3. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., and Ortega-Garcia, J., "Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, vol. 64, pp. 131–148, 2020.
4. Verdoliva, L., "Media Forensics and DeepFakes: An Overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910–932, 2020.
5. Dang, H., Liu, F., Stehouwer, J., Liu, X., and Jain, A., "On the Detection of Digital Face Manipulation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5781–5790, 2021.
6. Chandrasegaran, K., Tran, N., and Mahmud, M., "Explainable Deepfake Detection Using Visual Attention Mechanisms," *IEEE Access*, vol. 10, pp. 39520–39532, 2022.
7. Mirsky, Y., and Lee, W., "The Creation and Detection of Deepfakes: A Survey," *ACM Computing Surveys*, vol. 54, no. 1, pp. 1–41, 2022.
8. Wang, S., Wang, O., Zhang, R., Owens, A., and Efros, A., "CNN-Generated Images Are Surprisingly Easy to Spot… for Now," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8695–8704, 2020.
9. Yu, N., Davis, L., and Fritz, M., "Attributing Fake Images to Generative Models," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 7556–7566, 2019.
10. Kietzmann, J., Lee, L., McCarthy, I., and Kietzmann, T., "Deepfakes: Trick or Treat?" *Business Horizons*, vol. 63, no. 2, pp. 135–146, 2020.