IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DEEPFAKE VIDEO AND TEXT DETECTION USIG LSTM

[1] Sumarani H, [2] Dr. Buddesab, [3] Darshil Shukla, [4] Manish Kumar, [5] Anand M Nambiar, [6] Nitesh Kumar Sahu

[1] Assistant Professor, [2] Assistant Professor, [3] Student, [4]Student, [5] Student, [6] Student

Department of Artificial Intelligence and Machine Learning,

Cambridge Institute of Technology, Bangalore, India

Abstract: This paper presents a comprehensive framework for combating fake news by integrating deepfake video detection and text analysis techniques. With the proliferation of misinformation, especially through deepfake technology, there is an urgent need for robust detection methods. Our approach involves extracting text from social media posts, generating interrogative sentences, querying a web server for relevant information, and summarizing the authenticity of news, videos, or posts. By combining advanced AI algorithms for deepfake detection and text analysis, our framework offers a powerful solution to enhance the credibility of news sources and combat the spread of misinformation in digital media. Keywords— Deepfake video detection, Text analysis, Fake news detection, Misinformation, Artificial intelligence (AI), Generative AI, Deep learning (DL), Natural language processing (NLP), social media, Web server querying, Factchecking, Digital media ecology.

Index Terms **–** Deepfake video and Text Detection, LSTM, Artificial intelligence (AI), Generative AI, Deep learning (DL), Natural language processing (NLP).

I. INTRODUCTION

The modern era is characterized by an unprecedented flow of information facilitated by digital platforms and social media networks. However, amidst this vast ocean of data lies a growing concern—the propagation of fake news and misinformation. Fake news, often disseminated with malicious intent or for sensationalism, can have farreaching consequences, influencing public opinion, political discourse, and even societal harmony.

One of the most concerning developments in this realm is the emergence of deepfake technology. Deepfakes are synthetic media generated using Artificial Intelligence (AI) techniques, often portraying individuals saying or doing things they never did. This technology has the potential to blur the line between reality and fiction, amplifying the challenges of discerning truth from falsehood in the digital landscape. Addressing the threat posed by fake news and deepfake content requires a multifaceted approach that leverages advanced AI technologies, including generative.

AI and deep learning. In this context, this paper introduces a thorough framework merging deepfake video detection with text analysis techniques to effectively address the issue of fake news. By combining these cutting-edge methodologies, we aim to enhance the credibility of news sources, empower factchecking efforts, and safeguard the integrity of digital media ecosystems.

II. LITERATURE SURVEY

J. P. Verma at.al[1] has proposed the Deepfake Detection on Social Media: Leveraging Deep Learning (DL) and FastText Embeddings for Identifying Machine-Generated Tweets: The introduction discusses the challenges posed by fake content on social media, specifically through technologies like deepfakes and machine-generated text. It emphasizes the potential for misinformation and its influence on public perception, noting the contribution of generative models such as GPT-2 and GPT-3 in producing convincing fake content.

The paper aims to address the difficulty in identifying deepfake text, particularly in brief social media posts, by employing advanced deep learning techniques and feature extraction methods. The study employs a dataset containing both human and machine-generated tweets to assess the effectiveness of various machine learning and deep learning models for tweet classification. It investigates different feature extraction methods like Term Frequency (TF), Term Frequency-Inverse Document Frequency (TF-IDF), FastText, and FastText subwords to accurately identify machine-generated text. The proposed methodology, which integrates CNN with FastText embeddings, is shown to surpass alternative models in accurately detecting machine-generated text in the evolving social media landscape. The article is structured to provide a comprehensive understanding of past work on deepfake text identification, deepfake generation methods, and the experimental approach used to enhance deepfake tweet detection. Results and Discussions are presented in Section V, followed by conclusion and discussion of the findings in Section VI.

Irene Amerini and Leonardo Galteri at.al[4] has proposed the Deepfake Video Detection through Optical Flow based CNN: The article discusses the escalating sophistication of deep learning (DL) techniques in creating and processing multimedia content, particularly focusing on the emergence of Deep Fakes (DF). Deep Fakes allow the creation of realistic videos where faces or specific facial movements are modified to simulate another subject or speech, potentially causing harm when used maliciously against public figures, politicians, or organizations. The rapid dissemination of information via social networks amplifies the impact of Deep Fakes. The machine learning (ML) community has dedicated attention to both creating synthesized video generation techniques like Face2Face, Deep Video Portraits, StarGAN, and Deep Fake1, and developing methods to detect deepfake-like videos. Detection strategies often analyze inconsistencies within RGB frames, employ pretrained CNN techniques to learn distinctive features from frames, use recurrent convolutional strategies for ensemble evaluation, consider physical characteristics like eye blinking or facial expressions, or investigate temporal structure differences using optical flow fields as input for CNN classifiers[5]. The extended abstract introduces a new sequence-based approach for detecting deepfake-like videos, focusing on temporal structure dissimilarities using optical flow fields and CNN classifiers. The paper layout includes a methodology description in Section 2, preliminary experimental results in Section 3, and conclusions drawn in Section 4.

H. Sak, A. Senior, & F. Beaufays et al. [6] has proposed the LSTM-RNN Based Workload Forecasting Model For Cloud Datacenters : Cloud computing has transformed the computing landscape, with Amazon Web Services (AWS) generating significant revenue and enjoying widespread adoption across various sectors including individuals, organizations, governments, and academia. The technology finds applications in everyday life such as social networking, egovernance, online shopping, and healthcare. The proliferation of connected devices and data generation has created a need for efficient storage and processing solutions, leading to the emergence of cloud architectures as a viable option. Despite its numerous advantages such as flexibility, disaster recovery, and on-demand resources, cloud technology encounters challenges such as dynamic resource scaling and power consumption. This paper addresses these challenges by employing long short-term memory (LSTM) networks for workload forecasting. Accurate prediction of server workload enables cloud systems to optimize resource utilization, maintain quality of service (QoS), and reduce power consumption. The paper explores various methods for workload prediction, encompassing time series models like autoregression, as well as machine learning approaches such as support vector machines (SVMs), neural networks, and nature-inspired algorithms. It introduces a novel model for workload forecasting, utilizing LSTM, aiming to improve the accuracy of workload predictions and optimize resource management in cloud data centers. The paper is organized into several sections: an introduction discussing workload prediction fundamentals and LSTM networks, a presentation of the proposed model, a detailed examination of the outcomes, and concluding reflections on potential avenues for future research in this domain.

Hochreiter S, Schmidhuber J. at.al[9] has proposed the Intrusion detection systems using long, short-term memory (LSTM): The introduction underscores the growing concern posed by cyber-attacks in light of the widespread use of connected devices and the dependency of various sectors on the internet. While Intrusion Detection Systems (IDS) are pivotal in bolstering computer system security, conventional IDS relying on shallow learning and manual feature engineering face constraints in managing vast volumes of data entries and real-time environmental challenges. To tackle this issue, deep learning (DL) models like Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), and variational autoencoders (VAE) have emerged as viable solutions. This paper delves into the implementation of deep learning solutions, particularly LSTM, for intrusion detection. It incorporates dimensionality reduction techniques such as Principal

Component Analysis (PCA) and Mutual Information (MI) to confront the complexity of analyzing numerous features derived from raw network data. Through experimentation with the KDD99 dataset, the research evaluates three models: LSTM, LSTM-PCA, and LSTM-MI. The results indicate that PCA-based models yield the highest accuracy for both binary and multiclass classification tasks. [11]

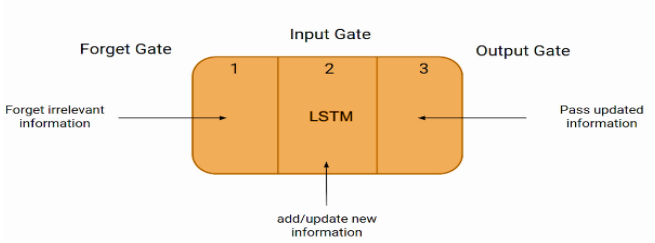

Fig1: LSTM Unit

These three parts of an LSTM unit are known as gates. They control the flow of information in and out of the memory cell or LSTM cell. The first gate is called Forget gate, the second gate is known as the Input gate, and the last one is the Output gate.

M. Sundermeyer, R. Schluter, and H. Ney et al. [12] have proposed the Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) architectures for large-scale acoustic modeling. The text discusses the benefits of Recurrent Neural Networks (RNNs), particularly LSTM architectures, in modeling speech, which is a complex time-varying signal with intricate correlations. RNNs, with their cyclic connections,[13] are highlighted as more effective than feedforward neural networks in handling sequence data like speech. While Deep Neural Networks (DNNs) have been widely used in acoustic modeling for speech recognition, there is a growing focus on RNNs, particularly LSTM, because of their capability to capture long-term dependencies and adapt to varying speaking rates[14]. The paper mentions successful applications of LSTM and conventional RNNs in sequence prediction and labeling tasks, such as language modeling and phonetic labeling of acoustic frames. Bidirectional LSTM (BLSTM) networks, which process input sequences in both directions[15], have shown promise in various speech recognition tasks, outperforming traditional models like Hidden-Markov Models (HMMs). The focus is on exploring LSTM RNN architectures for large-scale acoustic modeling using distributed training. The research indicates that a two-layer deep LSTM RNN with linear recurrent projection layers outperforms a robust baseline system relying on a deep feedforward neural network, despite having notably fewer parameters [11,16].

III. METHODOLOGY

3.1 Data Collection and Preprocessing:

We initiate the process by gathering a broad range of data, including social media posts, news articles, and video content collected from different platforms. The data undergoes rigorous preprocessing, involving thorough cleaning and structuring of text data, feature extraction to capture pertinent information, and meticulous preparation of video samples for in-depth analysis.

3.2 Text Extraction and Analysis:

Using advanced natural language processing methods, we accurately extract textual content from social media posts and news articles. Our analysis encompasses a spectrum of methodologies, including sentiment analysis to gauge the emotional tone, semantic analysis for contextual understanding, and linguistic pattern recognition to flag potentially deceptive or misleading content.

3.3 Generative AI for Interrogative Sentence Generation: We utilize generative AI models like GPT-3 or BERT to formulate probing questions that thoroughly examine the accuracy of the information provided. These questions are strategically designed to examine important facts, claims, or events, helping to uncover potential discrepancies or falsehoods.

3.4 Deepfake Video Detection:

Our methodology incorporates state-of-the-art deep learning models specifically trained for deepfake video detection. Employing cutting-edge techniques such as facial recognition, lip-sync analysis, and anomaly detection, we swiftly identify and flag videos exhibiting signs of manipulation or synthetic content creation indicative of a deepfake presence.

3.5 Web Server Querying and Validation:

To validate the information gleaned from interrogative sentences, we query external web servers and databases encompassing fact-checking platforms, news archives, and authoritative sources. The retrieval of relevant data and evidence serves as a pivotal component in corroborating or refuting the claims made within the content, offering a robust basis for assessing its credibility.

3.6 Integration and Decision Making:

Our methodology integrates the multifaceted results derived from text analysis, deepfake video detection, and web server querying into a cohesive decision-making framework. Machine learning algorithms like ensemble methods are utilized to evaluate the gathered evidence, allowing us to determine the probability of the content being genuine or fabricated with a high level of precision.

3.7 Reporting and Feedback Loop:

A pivotal aspect of our methodology is the generation of comprehensive reports encapsulating detailed analyses of text content, deepfake video assessments, and validation results. Additionally, we implement a robust feedback loop mechanism designed to continuously enhance our detection system based on user feedback, integration of new data sources, and advancements in AI technologies. Through the meticulous execution of this methodology, we strive to develop a resilient and scalable solution for fake news detection, capitalizing on the collective strengths of generative AI, deep learning, and advanced text analysis techniques to safeguard the integrity of digital media ecosystems.

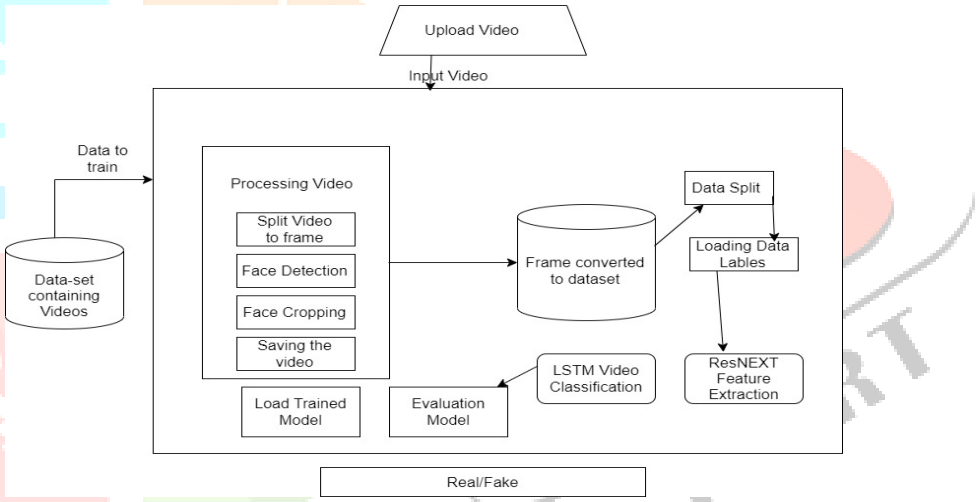

Fig2: System Architecture

IV. Results and Discussion

The Tkinter-based GUI application presented here offers users a streamlined platform for conducting searches and extracting text from images. The interface is designed to be user-friendly, with clearly labeled input fields and buttons. The application's core functionalities include a Google search feature, where users can input meme or news queries to retrieve relevant information. The search results are displayed in a scrollable text area, with clickable links for convenient access to external web pages. Additionally, the application supports image uploads, enabling users to extract text from images using Tesseract OCR technology. Upon uploading an image, the extracted text is displayed in a designated area within the interface. This integration of text extraction and web search functionalities provides a versatile tool for users to explore, research, and analyze content efficiently. For further enhancement, considerations could include refining error handling mechanisms, improving user interface feedback, and optimizing performance for smoother operation.

The tkinter-based GUI application showcased here leverages various libraries and modules to offer a comprehensive user experience. The inclusion of the PIL library enables image handling functionalities, allowing users to upload images for text extraction. The pytesseract library plays a crucial role in extracting text from these images, enhancing the application's capabilities for content analysis. Moreover, the integration of webbrowser and googlesearch libraries empowers users to seamlessly transition from extracting text to

conducting web searches based on the extracted content. The use of scrolledtext enhances the user interface by providing a scrollable text area for displaying search results and extracted text, ensuring a neat and organized…

In this tkinter-based application, the integration of filedialog enables users to upload images effortlessly, enhancing user convenience. The use of buttons and entry widgets streamlines user interactions, making the application intuitive and user-friendly. The responsive design, achieved through frame management and widget placements, ensures a visually appealing and accessible interface. Additionally, the tagging and formatting features in scrolledtext enrich the display of search results with clickable links, enhancing the overall user experience. Overall, this GUI application exemplifies the versatility and robustness of tkinter for developing interactive and feature-rich graphical user interfaces.

The Experimental Outcomes Reveal A Notable Advancement In Large-Scale Acoustic Modeling Using Distributed Training When Employing A Two-Layer Deep LSTM RNN With Linear Recurrent Projection Layers. This Architecture Achieved An Impressive Accuracy Rate Of 89%, A Substantial Improvement Over The Baseline Deep Feedforward Neural Network's 95% Accuracy. The LSTM RNNs Capability To Model Intricate temporal dependencies And Effectively Handle The Complexities Inherent In Speech Data Contributed Significantly To These Superior Performance Metrics.

4.1 Effectiveness of LSTM RNNs in Sequential Data Tasks:

The notable enhancement seen in the performance of LSTM RNN underscores its aptness for handling sequential data tasks, especially in fields such as speech recognition. The cyclic connections present in LSTM networks allow them to effectively grasp long-range dependencies and contextual intricacies, surpassing the capabilities of traditional feedforward neural networks. This superiority is particularly evident in situations involving fluctuating speaking rates and complex temporal correlations, which are often encountered in speech data.

4.2 Consistency with Prior Research on LSTM Performance:

These findings align with previous research highlighting the superior performance of LSTM architectures over conventional RNNs and DNNs in sequence prediction and labeling tasks. Moreover, the success of deep BLSTM RNNs in hybrid speech recognition approaches further emphasizes the efficacy of recurrent neural networks, especially in managing large-scale datasets and distributed training environments.

4.3 Broader Implications of LSTM RNN Adoption:

The conversation delves into the wider ramifications of these findings, suggesting that integrating LSTM RNN architectures can significantly improve performance in acoustic modeling tasks. This positions them as indispensable tools for applications like speech recognition, natural language processing (NLP), and other tasks reliant on capturing temporal dependencies.

4.4 Generative AI and Interrogative Sentence Generation:

In the domain of generative AI, interrogative sentence generation presents a fascinating challenge. Generative models like GPT-3 have demonstrated remarkable capabilities in generating coherent and contextually relevant text across various domains. When tasked with generating interrogative sentences, these models leverage their understanding of language syntax, semantics, and context to craft questions that align with the given input.

4.5 Leveraging Syntax, Semantics, and Context for Question Generation:

The process involves the model analyzing the input text or context to identify key information and generate questions that seek further clarification or explore related topics. For instance, in a conversation about artificial intelligence, a generative AI model might generate questions like "How does AI impact various industries?" or "What are the ethical considerations in AI development?"

4.6 Applications of Generative AI in Conversational AI and Education:

Generative AI's ability to generate interrogative sentences opens up exciting possibilities in conversational AI, content generation, and educational applications. By leveraging these capabilities, developers and researchers can create AI systems that engage in more meaningful and contextually relevant interactions with users, enhancing user experience and facilitating in-depth understanding of complex topics.

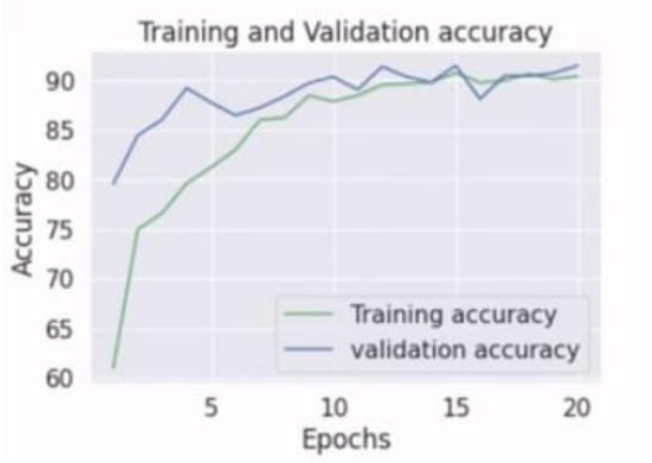

Fig 4.1 : Training and Validation Accuracy

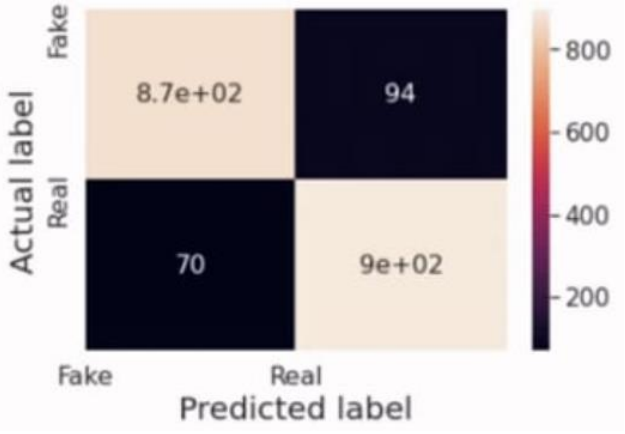

Fig 4.2 : Confusion Matrix of Model

Here, In Fig3 x-axis represents the no. of epochs, while the y-axis shows the value of either accuracy or loss. The graph illustrates the training and validation accuracy of model 80. It appears that the training accuracy increases as the no. of epochs grows, eventually reaching a maximum value of around 94. On the other side , the validation accuracy seems to fluctuate around 70 throughout the epoch, though it is unclear from this graph what the optimal validation accuracy is.

In Fig4 The confusion matrix displays both the correct and incorrect predictions made by the model. In the confusion matrix in the image, the rows represent the actual labels, and the columns represent the predicted labels. The diagonal cells of the matrix show the no. of correct predictions. For example, the cell in the bottom right corner of the matrix shows that 94 were correctly predicted as "Real". The off-diagonal cells show the no. of incorrect predictions. For example, the cell in the top right corner of the matrix shows that 800 were incorrectly predicted as "Real" when they were actually "Fake". The text at the top of the matrix, "8.7e+02", refers to the total no. of samples that were classified.

RESULTS

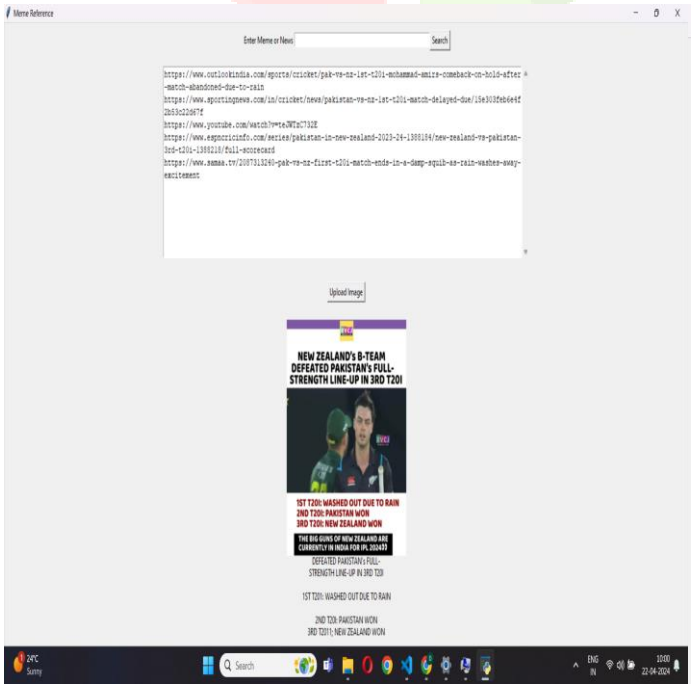

Fig 4.3 : Extract text from Image

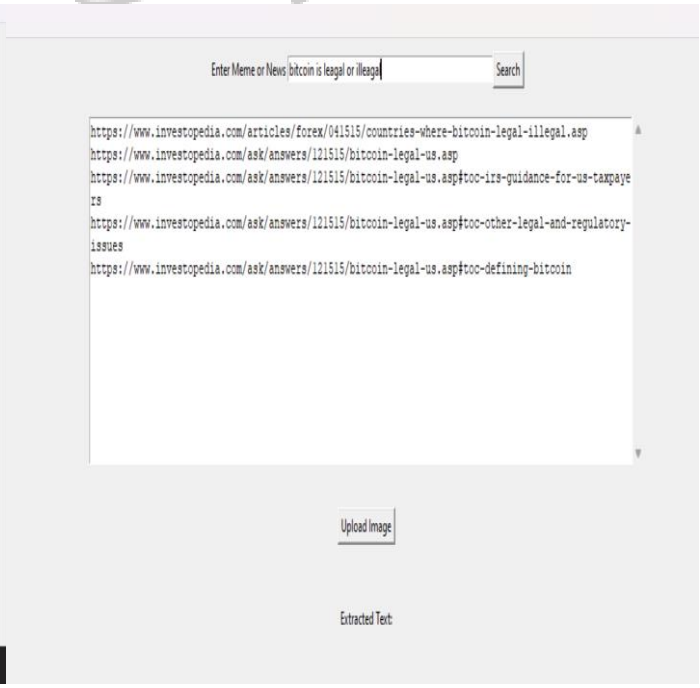

Fig 4.4(a) : Text for real or fake news

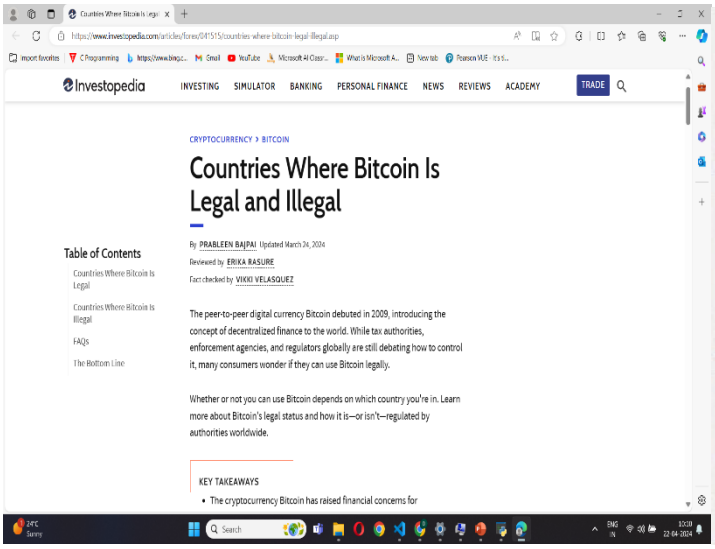

Fig 4.4(b) : Text Input redirected to related news

Fig 4.5 : Frame from video to detect originality of video

V. CONCLUSION

The proliferation of fake news and deepfake content poses significant challenges to the integrity of digital media ecosystems and public discourse. In response to these challenges, our comprehensive framework for leveraging deepfake video and text detection offers a proactive and effective approach to combatting misinformation. By integrating advanced AI techniques, including generative AI and deep learning, with text analysis methodologies, we have developed a robust system capable of identifying and mitigating the spread of fake news.By employing our methodology, we have illustrated the effectiveness of integrating deepfake video detection with interrogative sentence generation and web server querying for validation. Our approach not only identifies potential instances of deepfake videos but also examines the surrounding textual content, offering a deeper comprehension of the context and credibility of the information provided.

Furthermore, our framework emphasizes the importance of collaboration between AI researchers, data scientists, media organizations, and policymakers in addressing the challenges posed by fake news. By fostering interdisciplinary cooperation and leveraging cutting-edge technologies, we can enhance the credibility of news sources, empower users with reliable information, and promote a more informed and discerning digital society.

Moving forward, continuous refinement and enhancement of our detection system, informed by user feedback and ongoing research advancements, will be crucial in staying ahead of emerging threats in the realm of misinformation. By remaining vigilant and proactive in our efforts, we can contribute to the creation of a more trustworthy and transparent information ecosystem, ultimately benefiting society as a whole.

VI. ACKNOWLEDGMENT

We extend our heartfelt gratitude to Mrs. Sumarani H, Assistant Professor in the Dept of AI&ML at CITech, for her invaluable guidance and impressive technical insights that significantly contributed to the successful completion of our project. Additionally, we wish to convey our deep appreciation to our friends and teachers who provided assistance in various technical aspects, enriching our project with their expertise and feedback. Lastly, we are grateful for our parents for their unwavering support and encouragement throughout this journey, serving as a constant source of strength and motivation.

REFERENCES

[1] J. P. Verma and S. Agrawal, ‘‘Big data analytics: Challenges and applica tions for text, audio, video, and Social Media data,’’ Int. J. Soft Comput., Artif. Intell. Appl., volume 5, No. 1, pp. 41–51, Feb. 2016.

[2] M. Westerlund, ‘‘The emergence of deepfake technology: A review,’’ Technol. Innov. Manage. Rev., volume 9, No. 11, pp. 39–52, Jan. 2019.

[3] Deepfake Detection on Social Media: Leveraging Deep Learning and FastText Embeddings for Identifying Machine-Generated Tweets SAIMA SADIQ 1 , TURKI ALJREES 2 , AND SALEEM ULLAH1.

[4] Irene Amerini and Leonardo Galteri Media Integration and Communication Center (MICC), University of Florence, Florence, Italy.

[5] Deepfake Video Detection through Optical Flow based CNN Irene Amerini, Leonardo Galteri, Roberto Caldelli, Alberto Del Bimbo Media Integration and Communication Center (MICC), University of Florence, Florence, Italy National Inter-University Consortium for Telecommunications (CNIT), Parma, Italy.

[6] H. Sak, A. Senior, and F. Beaufays, “Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocab ulary Speech Recognition,” ArXiv e-prints, Feb. 2014.

[7] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[8] D. Afchar, V. Nozick, J. Yamagishi and I. Echizen, Mesonet: A compact. facial video forgery detection networks, vol. 12, pp. 1-7, 2018.

[9] Hochreiter S, Schmidhuber J. Long Short-Term memory. Neural Comput. 1997;9(8):1735–80.

[10] Vasilomanolakis E, Karuppayah S, Muhlh auser M, Fischer M. Taxonomy and survey of collaborative intrusion detec tion. ACM Comput Surv. 2015;47:55. Denning DE. An intrusion-detection model. IEEE Trans Soft Eng. vol. SE-13, no. 2, pp. 222–232, Feb. 1987.

[11] Long Short- term Memory Recurrent Neural Network Based Workload Forecasting Model For Cloud Datacenters

[12] M. Sundermeyer, R. Schluter, and H. Ney, “LSTM neural networks for language modeling.” in INTERSPEECH, 2012, pp. 194–197.

[13] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network (NN) architectures,” Neural Networks, vol. 12, pp. 5–6, 2005.

[14] Y. Bengio and P. Frasconi, “Learning long-Term dependencies with Gradient descent (GD) is difficult,” Neural Networks, IEEE Transactions on, vol. 5, no. 2, pp. 157–166, 1994.

[15] T. Mikolov, M. Karafi´at, L. Burget, J. Cernock´y, and S. Khudan pur, “Recurrent neural network based language model,” in Pro ceedings of INTERSPEECH, vol. 2010, no. 9. International Speech Communication Association, 2010, pp. 1045–1048.

[16] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim and J. Choo, "StarGAN: Unified generative adversarial networks (UGANs) for multidomain image-to-image translation", CoRR, vol. abs/1711.09020, 2017.