POLYTECHNIC UNIVERSITY OF THE PHILIPPINES
STA. MESA MANILA

Researchers:	Gallardo, Matthew Guevarra, John Joemer Rapiza, Leinard		
Program:	Bachelor of Science in Computer Science		
Contact Number:	09663389772 09761545003 09053153139	**Email Address:**	gallardomatthew8@gmail.com johnjoemerguevarra@gmail.com elrapiza@gmail.com
Date of Submission:	June 20, 2023		

Proposed Title:	Deepfake Detection System
Improved Title:	LSTM-Based Deepfake Detection in Videos
Research Themes:	Deepfake Themes, LSTM-based Models, Video Analysis
Keywords / Concepts:	Deepfakes, Artificial Intelligence, Video Manipulation, Long Short-Term Memory
Purpose of the Study:	The purpose of this study is to develop a web-based LSTM deepfake detection system specifically designed for analyzing videos. The system aims to enhance the accuracy and reliability of deepfake detection by leveraging the temporal dependencies and patterns present in video sequences. By combining the power of LSTM-based models with effective feature extraction techniques and face recognition algorithms, the proposed system aims to provide an improved solution for identifying manipulated videos and detecting instances of deepfake content. The study aims to contribute to the field of video forensics and contribute towards mitigating the harmful impacts of deepfake technology in various domains such as media, politics, and cybersecurity.

STATEMENT OF THE PROBLEM (Question Form) / STATEMENT OF PURPOSE (Declarative Form)	
PROPOSED	**REVISED**
1. To recognize the growing threat of deepfake videos.	To address the escalating threat posed by deepfake videos and the urgent need for an effective and robust deepfake detection system.
2. To recognize the need for a comprehensive deepfake detection system.	To acknowledge and address the urgent requirement for a comprehensive deepfake detection system that can effectively identify and combat the rising threat of deepfake videos.
3. To develop a comprehensive system to combat the growing threat of deepfake videos.	To develop a comprehensive and advanced system capable of accurately detecting and combating the proliferation of deepfake videos.
4. To assess the performance of the deepfake detection system.	To evaluate the performance and effectiveness of the developed deepfake detection system in countering the increasing threat of deepfake videos.
5. To evaluate the deepfake detection system's performance on how it is helpful to individuals to combat the growing threat of deepfake videos.	To assess the practicality and usefulness of the deepfake detection system for individuals and organizations in their efforts to mitigate the harmful impacts of deepfake videos.

6. To promote awareness and educate individuals on the threat of deepfake videos.		To actively promote awareness and educate individuals about the significant threat posed by deepfake videos in order to mitigate their harmful effects.
Significance of the Study Who and how will they benefit from the study?	**Institutional:**	**Academic Institutions:** The study contributes to the research and development of advanced techniques in the field of deepfake detection, bolstering the knowledge base of academic institutions and fostering innovation in the area of video forensics. **Research Organizations:** The findings of the study can provide valuable insights and methodologies for research organizations working on artificial intelligence, machine learning, and video analysis. It can serve as a foundation for further advancements in deepfake detection systems.
	Local:	**Law Enforcement Agencies:** Local law enforcement agencies can benefit from this study by utilizing the developed LSTM-based deepfake detection system to identify and investigate instances of deepfake videos involved in criminal activities, such as fraud, cyberbullying, and harassment. **Media and Entertainment Industry:** Local media organizations can employ the developed deepfake detection system to authenticate and verify the authenticity of video content, ensuring the dissemination of accurate information and protecting their reputation from potential deepfake manipulation.
	National:	**Government Agencies**: National governments can utilize the research findings to strengthen their cybersecurity frameworks, develop policies and regulations around the usage of deepfake technology, and take necessary actions to combat the potential misuse of deepfakes. **National Security Agencies:** The study's outcomes can aid national security agencies in identifying deepfake videos that may be used for misinformation campaigns, propaganda, or compromising national security interests. **General Public:** The general public can benefit from the study by having access to more reliable tools and technologies for detecting deepfake videos. This can help individuals in identifying and avoiding the consumption of manipulated content, thus reducing the spread of misinformation and potential harm caused by deepfakes.

METHODOLOGY	
Data Science Problem	The data science problem addressed in this study is deepfake detection in videos. The objective is to develop an LSTM-based model that can accurately identify manipulated videos and detect instances of deepfake content.
Process of Collecting Data	To train and evaluate the model, a diverse and representative dataset of videos will be collected from FaceForensics ++, CelebDF and Kaggle's Deepfake Detection Challenge Dataset. The dataset should contain both authentic and manipulated videos, including various types of deepfake techniques and scenarios.
Pre-processing of Data	Load the dataset,Split the video into frames, Crop the face from each frame, Save the face cropped video
***Data Insights**	Exploratory data analysis techniques will be employed to gain insights into the dataset. This may involve visualizing the distribution of authentic and manipulated videos, identifying common patterns or artifacts in deepfake videos, and analyzing any metadata or contextual information associated with the videos.
Chosen Modeling Technique: Algorithms, Problems, Methods	The chosen modeling technique for deepfake detection is an LSTM-based model. Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) that can capture temporal dependencies in sequential data. The LSTM model will be trained using the collected dataset, focusing on learning spatiotemporal features to distinguish between authentic and manipulated videos.

Data Split Percentage	The collected dataset will be divided into training, validation, and testing sets. The data split percentage may depend on the size of the dataset, but commonly used ratios are 70% for training, 15% for validation, and 15% for testing. This split ensures that the model is trained on a sufficiently large portion of the data while having separate sets for evaluating its performance.
Model Evaluation: Methods	Evaluating the LSTM-based deepfake detection model based on its performance in a real-world setting. Deploy the model in a live environment where it can analyze a diverse set of videos, including both authentic and manipulated content. Engage human reviewers or experts to validate the system's detection results and provide feedback on its effectiveness.
Expected Output	The expected output of the study is a well-trained LSTM-based deepfake detection model that achieves high accuracy in identifying manipulated videos and detecting instances of deepfake content. The model should generalize well to unseen data and demonstrate reliable performance in various real-world scenarios.
Possible Model Deployment	The trained LSTM-based deepfake detection model will be deployed as a web-based system.By deploying the LSTM-based deepfake detection model as a web-based system, users can easily access the detection capabilities through a standard web browser, enabling widespread usage and accessibility for individuals, organizations, and platforms seeking to identify and mitigate the risks associated with deepfake videos.
***Model Improvement**	Continuous improvement of the deepfake detection model can be pursued by incorporating additional techniques, such as ensemble methods, transfer learning, or incorporating domain-specific knowledge. Feedback from users and real-world usage can also inform model refinement and updates to adapt to evolving deepfake techniques and challenges.

INITIAL REFERENCES

1. L. Bondi, E. Daniele Cannas, P. Bestagini and S. Tubaro, "Training Strategies and Data Augmentations in CNN-based DeepFake Video Detection," 2020 IEEE International Workshop on Information Forensics and Security (WIFS), New York, NY, USA, 2020, pp. 1-6, doi: 10.1109/WIFS49906.2020.9360901.

   Link: Training Strategies and Data Augmentations in CNN-based DeepFake Video Detection

2. Gerstner, C. R., & Farid, H. (2022). Detecting real-time deep-fake videos using active illumination. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 53-60)

   Link: Detecting real-time deep-fake videos using active illumination

3. D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 2018, pp. 1-6, doi: 10.1109/AVSS.2018.8639163.

   Link: Deepfake Video Detection Using Recurrent Neural Networks

4. D. Liu, Z. Yang, R. Zhang and J. Liu, "A Robust Deepfake Video Detection Method based on Continuous Frame Face-swapping," 2022 International Conference on Artificial Intelligence, Information Processing and Cloud Computing (AIIPCC), Kunming, China, 2022, pp. 188-191, doi: 10.1109/AIIPCC57291.2022.00048.

   Link: A Robust Deepfake Video Detection Method based on Continuous Frame Face-swapping

5. L. S and K. Sooda, "DeepFake Detection Through Key Video Frame Extraction using GAN," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 859-863, doi: 10.1109/ICACRS55517.2022.10029095.

   Link: DeepFake Detection Through Key Video Frame Extraction using GAN

RESEARCH MILESTONES		
DATE	**ACTIVITY**	**PERSON/S INVOLVED**
Week 1 (second week of April)	Introduction to Research Objectives	All of Team Members
Weeks 2-3 (second to third week of April)	Data collection for the deepfake detection dataset	Researcher/Programmer
Weeks 4-5 (third and last week of Apil)	Pre-processing of the collected dataset	Researcher/Programmer 1
Weeks 6-9 (last week of April to the second week of May)	Development and training of the LSTM-based deepfake detection model	Researcher/Programmer 1
Week 9 (third week of May)	Splitting the dataset into training, validation, and testing sets	Researcher/Programmer 2
Weeks 10-11 (third and fourth week of May))	Model evaluation and system-based testing	Researcher/Programmer 2
Week 12 (fist week of June)	Deployment of the web-based deepfake detection system	Researcher/Programmer r 2
Weeks 17-18 (fourth week of July to the first week of August)	User testing and feedback gathering	Researcher/Programmer 3
Week 19 (second week of July)	Model improvement and adaptation to emerging deepfake techniques	Researcher/Programmer 3
Weeks 20-21 (third and fourth week of July)	Finalizing the research findings and conclusions	All of Team Members
Week 22 (last week of July)	Documentation and preparation of research report	All of Team Members

Target Date of Completion:	Third to Last Week of July , 2023

RECOMMENDATIONS:	
FINAL APPROVAL:	
Dr. Juancho D. Espineli	