Journal of Communication and Management

ISSN: 2583-617X (Online)

Journal home page: https://jcoma.com

Research Article

DOI: 10.58966/JCM2023243

Unmasking the Deepfake Infocalypse: Debunking Manufactured Misinformation with a Prototype Model in the AI Era "Seeing and hearing, no longer believing."

Tendral Rajagopal[1*], Velayutham Chandrashekaran[1], Vignesh Ilango[2]

[1]Anna University, Chennai, India
[2]Microsoft India, Hyderabad, India

ARTICLE INFO

Article history:
Received: 13 October, 2023
Revised: 16 October, 2023
Accepted: 13 November, 2023
Published: 18 December, 2023

Keywords:
Fake news, Deepfakes, Media literacy, AI, Machine learning.

ABSTRACT

Machine learning and artificial intelligence in Journalism are aid and not a replacement or challenge to a journalist's ability. Artificial intelligence-backed fake news characterized by misinformation and disinformation is the new emerging threat in our broken information ecosystem. Deepfakes erode trust in visual evidence, making it increasingly challenging to discern real from fake. Deepfakes are an increasing cause for concern since they can be used to propagate false information, fabricate news, or deceive people. While Artificial intelligence is used to create deepfakes, the same technology is also used to detect them. Digital Media literacy, along with technological deepfake detection tools, is an effective solution to the menace of deepfake. The paper reviews the creation and detection of deepfakes using machine learning and deep learning models. It also discusses the implications of cognitive biases and social identity theories in deepfake creation and strategies for establishing a trustworthy information ecosystem. The researchers have developed a prototype deepfake detection model, which can lay a foundation to expose deepfake videos. The prototype model correctly identified 35 out of 50 deepfake videos, achieving 70% accuracy. The researcher considers 65% and above as "fake" and 65% and below as "real". 15 videos were incorrectly classified as real, potentially due to model limitations and the quality of the deepfakes. These deepfakes were highly convincing and flawless. Deepfakes have a high potential to damage reputations and are often obscene or vulgar. There is no specific law for deepfakes, but general laws require offensive/fake content to be taken down. Deepfakes are often used to spread misinformation or harm someone's reputation. They are designed to harass, intimidate, or spread fear. A significant majority of deepfake videos are pornographic and target female celebrities.

Introduction

By 2023, around 5.3 billion people, two-thirds of the world's population, will be part of the growing information ecosystem. A few years ago, developing a machine that could think like humans was a dream, far-fetched imagination. A video of President Obama went viral with almost 7.5 million views on YouTube, "You won't believe What Obama Says in This Video" (Schick, 2020). A Deepfake video of Ex-US president Obama circulated, which became sensational in no time. Artificial intelligence-backed fake news characterized by misinformation and disinformation is the new emerging threat in our broken information ecosystem. There are many open AI options and transfer learning opportunities from big giants like Google and Tesla. Creating the deep fake video of Obama is child's play now.

When programmable computers were conceived a few decades ago, nobody could have predicted that computers would open a virtual world that is intellectually difficult for humans to solve but an elementary task for computers. The most difficult mental and abstract task for a human is the easiest for a computer to function. Machine learning

***Corresponding Author:** Tendral Rajagopal
Address: Anna University, Chennai, India
Email ✉**:** rajagopaltendral@gmail.com

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

has entered the field of journalism and has already created many remarkable transformations in the process of news gathering. Machine learning has enhanced the experience and accuracy and has deconstructed the challenge of finding the most sensational news. In 2018, Reuters developed NewsTracker and Lynx Insight to build a cybernetic newsroom, and these are essential tools that use machine learning and artificial intelligence to spot breaking news and identify critical factors. NewsTracker is a tool where the journalist can locate the most viral, breaking stories on Twitter and filter out unreliable sources from them. These tools help journalists flag potential newsworthy stories faster than other journalists.

On the other hand, Lynk Insight emerged to augment traditional journalism by identifying key facts, trends, and patterns, suggesting to journalists the news stories they could potentially write about, and helping journalists draft personalized news stories. Due to its realism and scope of impact, DeepFake (such as bogus photos, audio, and videos) has become a serious menace to our civilization. The issue has been exacerbated worse by the plethora of applications for creating phony images, such as FaceApp and ZAO. DeepFake's backend algorithms rely heavily on generative adversarial networks (GANs), employed to synthesize voices and face images. Several of the most cutting-edge DeepFake techniques now available operate at a level that is difficult for humans to see (Yihao Huang, 2020)

Machine learning and artificial intelligence in Journalism are aid and help, not a replacement or challenge to a journalist's ability. Such tools can reduce the newsroom pressure on the journalist to bring in exciting, viral stories. Facial expression and body language are some properties that machine learning currently cannot detect. Passive information like existing photos and videos of a person is the fundamental data to create deep fakes. Identifying fake news is complex since it is not a simple task. If the fake news is not exposed quickly, people will propagate it, and everyone will begin to believe it. Fake news can harm people, communities, groups, or political parties. The influence of fake news on people's opinions and decisions during the 2016 US election is a well-known adverse effect of fabricated information.

Information warfare results from technological advancement with no or minimal media literacy. Media literacy has to especially reach the Tier 3 and Tier 4 cities of every country, which are the most vulnerable people. Along with the increase in accessibility of technology and smartphones, an increase in awareness of proper usage of these inventions should also be initiated. There is no concrete solution to deepfakes; the only way to tackle deepfakes is a combination of all hybrid solutions. To combat deep fakes, Martijn Rasser emphasizes increased digital literacy. We must advance as a society. As so many people are willing to believe deep fakes and spread them, deep fake circulation has become highly effective. A desirability bias causes us to frequently see what we desire to be true (Martijn Rasser, 2019)

What are deepfakes?

Deep fakes are hyper-realistic videos, audio, or images created with the help of artificial intelligence. The development of advanced neural networks has made the creation of deep fakes simpler to the extent that even sophisticated software fails to identify the difference between fiction and reality. Deep fakes are applications that superimpose and swap the face of the source person with a target person, intending to make the target person say or do things that the source does or as intended by the creator. Deepfakes are artificially synthesized content using a computer graphic approach or deep learning models like GAN (Generative adversarial network). GAN has spiralled the existing problem of revenge porn, fake news, financial fraud, warmongering, election manipulation, defaming a person, and causing disruption to government functioning.

Generation of fake videos

Fake videos, known as deepfakes, are generated using Generative Adversarial Networks (GANs). GANs consist of two neural networks: a generator and a discriminator (Yuvraj Choubisa, 2023). The Generator creates synthetic videos, while the discriminator evaluates the authenticity of the generated videos. The process begins with the Generator creating an initial random sample, usually a low-resolution image or video. The discriminator is then fed

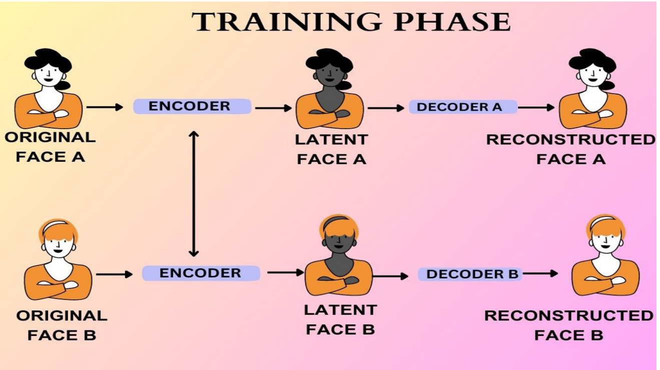


Figure 1: Training phase of deepfakes creation model

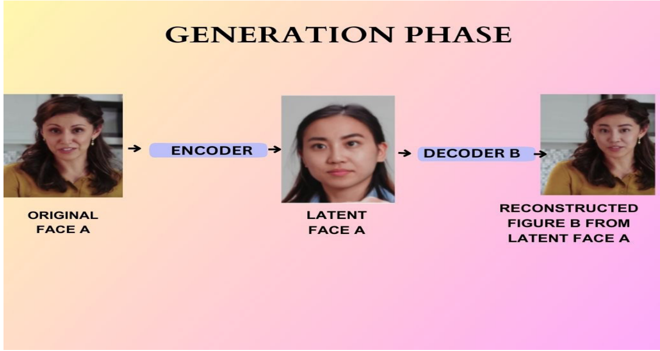


Figure 2: Generation phase of deepfakes

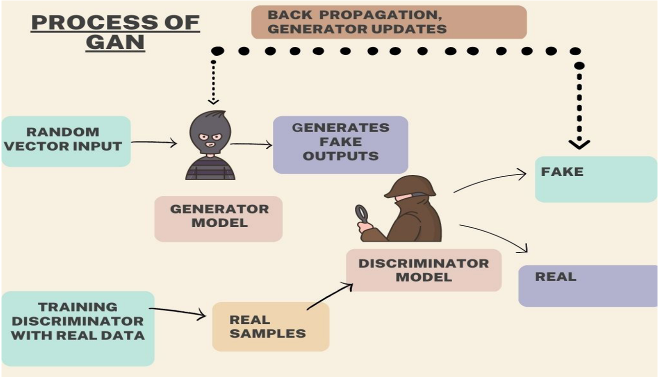


Figure 3: Process of GAN

with real and generated samples and tries to distinguish between them. The Generator takes the feedback from the discriminator and adjusts its output accordingly to create more realistic examples. This process continues until the Generator can create a video indistinguishable from real ones. One of the critical challenges in generating fake videos is the ability to create realistic facial expressions and movements. To overcome this challenge, GANs use facial landmark detection, which identifies and tracks the key facial features such as the eyes, nose, and mouth and then maps them to a digital mesh. This mesh can then be manipulated to create realistic facial movements.

Another technique used in generating fake videos is called style transfer. In this technique, the Generator learns the style of the source video and applies it to a target video. This enables the Generator to create videos with the same visual style as the source video while the content of the video can be manipulated. Once the Generator has made a real video, it can be used to create deepfakes videos by replacing a person's face in a real video with the synthetic look created by the Generator. This can be achieved using techniques such as face swapping or face reenactment. In conclusion, GANs are a powerful tool for generating fake videos, and their sophistication and realism have improved dramatically in recent years. However, using GANs to generate deepfakes raises important ethical and legal concerns, particularly concerning privacy, identity theft, and political propaganda.

Methodology

The researchers have developed a prototype model for detecting deepfake media using a semi-supervised approach. The construction of this prototype model is based on the principles of transfer learning from various existing deep learning frameworks such as DeepFakes, Face2Face, FaceSwap, and NeuralTextures. This involves utilizing pre-trained models and techniques from these frameworks to train a new model specifically designed for detecting fake media. Transfer learning, a powerful technique, plays a crucial role in this process by allowing the adaptation of a pre-trained model to a new dataset. In this particular case, the pre-trained models used are those within the DeepFakes, Face2Face, FaceSwap, and NeuralTextures frameworks, all renowned for generating a wide array of fake media.

Pre-trained models such as DeepFakes, Face2Face, FaceSwap, and NeuralTextures are significant for creating realistic synthetic media by altering facial features and expressions in images and videos. These models employ machine learning and neural network techniques for content manipulation, particularly in facial reenactment or manipulation. Face2Face focuses on real-time facial reenactment by tracking facial expressions from a source video and applying them in real-time to a target face in another video. This technique utilizes convolutional neural networks (CNNs) for convincing facial reenactment, allowing for natural and seamless manipulation of facial expressions. FaceSwap algorithms use deep neural networks to replace one person's face with another in images or videos. They learn the facial features and landmarks of both the source and target faces to create a realistic swap. FaceSwap techniques have both creative and potentially malicious applications. NeuralTextures is a method that focuses on texturing facial surfaces for facial reenactment. It employs neural networks to transfer the facial texture of a source actor onto a target actor's face in videos, enabling more accurate and realistic facial reenactment.

The model is fine-tuned with pre-trained weights to detect fake media in various applications effectively. It undergoes continuous feedback and learning to maintain precision and accuracy in predictions. During the fine-tuning process, the pre-trained model's weights are adjusted to fit the new dataset while retaining the knowledge learned from the original pre-training. This knowledge transfer helps the model learn to detect fake media more effectively. To ensure reliable performance, the model is evaluated using accuracy, precision, recall, and F1 score metrics. The model is deployed post-evaluation in face recognition, emotion detection, and age estimation applications. Social networking platforms can utilize deepfake identification to obstruct the dissemination of manipulated material and safeguard users against misinformation. Journalism and News Platforms can be incorporated into newsrooms to authenticate the veracity of video material prior to its publication, thereby preventing the propagation of fabricated news. Financial and Judicial Institutions can use it to validate the authenticity of audio or video evidence, and one can avert fraud related to deepfakes in financial transactions or legal proceedings. Implementing the deepfake identification model in the aforementioned contexts and continuously monitoring its performance in real-time are essential. Deploying a deepfake identification model necessitates a multidisciplinary approach, which considers technical, ethical, and legal facets to guarantee its effective utilization while mitigating potential risks associated with the proliferation of manipulated content in various societal domains. Ongoing monitoring and refinement of the model may be necessary to ensure it remains effective as new fake media emerge.

Prototype model dataset

Our approach for detecting forgery involves framing it as a binary classification problem where each frame of manipulated videos is classified as genuine or fake. We partitioned the dataset into three fixed sets to conduct these experiments: training, validation, and testing. The present study is focused on collecting videos from diverse sources that depict manipulated footage online. These sets contained 350, 90, and 90 videos, respectively. Our evaluations are solely based on the videos from the test set.

The process involves several steps to enable the system to classify the image accurately. Firstly, the input image is subjected to a robust face-tracking method, effectively identifying and locating the face in the picture. This step is essential to ensure that the classification network receives input only from the region of interest, which in this case is the face.

Once the face has been successfully located and extracted from the image, the region is fed into a learned classification network. This network has been trained on a large dataset of faces, enabling it to accurately predict the identity or characteristics of the person in the image.

The output of the classification network is prediction, which is based on the features extracted from the input image. These features are learned during the training phase of the web and are used to classify new input images. Overall, this process enables accurate and efficient classification of images containing faces and is used in various applications such as face recognition, emotion detection, and age estimation.

Prototype model working

CNN-based classifiers are one of the popular methods used for detecting fake videos. The basic idea behind CNNs is to extract relevant features from the input data using convolutional layers. These features are then passed to fully connected layers for classification.

In detecting fake videos, CNN-based classifiers analyze the video's content frame by frame. Each frame is considered an image and is fed into the CNN for analysis. The CNN then extracts features from the frame, such as texture, edges, and color, that are relevant for distinguishing between real and fake videos. A Convolutional Neural Network (CNN) comprises several layers to extract useful features from input images. The different layers of a CNN are as follows:

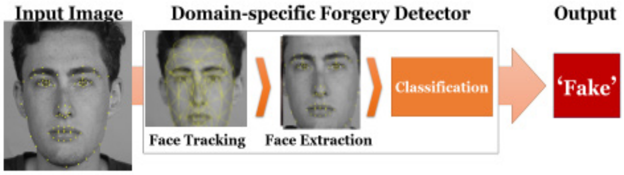


Figure 4: Process of the deepfake detection model

Input Layer: This layer receives the input image and passes it to the next layer for processing.

- Convolutional Layer: The Convolutional Layer performs a mathematical operation known as convolution on the input image using a set of learnable filters. This results in the extraction of features from the input image.

The output feature map Y at location (i, j) is given by the formula:

$Y(i,j) = b + \sum(m=0 \text{ to } M-1) \sum(n=0 \text{ to } N-1) X(i+m, j+n) * W(m,n)$

Where X is the input feature map, W is the set of learnable weights (filters), b is the bias term, and M and N are the dimensions of the filter.

- ReLU Layer: The rectified linear unit (ReLU) Layer applies a non-linear activation function to the output of the convolutional layer, removing negative values and introducing non-linearity into the network.

The output of the ReLU Layer is calculated using the following formula:

$f(x) = \max(0, x)$

- Pooling Layer: The pooling layer reduces the dimensionality of the feature maps obtained from the convolutional layer by downsampling them. This makes the network more computationally efficient and reduces overfitting.

The output of the Pooling Layer is calculated using the following formula:

$Y(i,j) = maxp(m=0 \text{ to } P-1)\ maxq(n=0 \text{ to } Q-1)\ X(istride+m, jstride+n)$

Where X is the input feature map, Y is the output feature map, the stride is the stride length, and P and Q are the dimensions of the pooling filter.

- Dropout Layer: The dropout layer randomly drops out a certain percentage of neurons in the network during training, preventing the network from becoming too reliant on any feature and reducing overfitting.

The output of the Dropout Layer is calculated by randomly dropping out a certain percentage of the neurons in the previous layer during training and scaling the remaining neurons by a factor of 1/(1 - dropout_rate).

- Flatten Layer: A completely linked layer can be utilized as input by flattening the output of the preceding layer into a one-dimensional vector.
- Fully Connected Layer: The fully connected layer performs a matrix multiplication on the input vector and a set of learnable weights, followed by applying a non-linear activation function. This layer enables the network to learn complex relationships between features.

The output of the Fully Connected Layer is calculated using the following formula:

$Y = f(WX + b)$

where X is the input vector, W is the set of learnable weights, b is the bias term, and f is the activation function.

- Output Layer: The output layer produces the network's final output, which can be either a probability distribution over the classes in the case of a classification task or a continuous value in the case of a regression task.

The output layer is calculated by

$Y = 1 / (1 + e^{(-z)})$

where z is the input to the output layer. In a multi-class classification task, the output is calculated using the softmax activation function:

$Y_i = e^{(z_i)} / \sum(j=1 \text{ to } C)\ e^{(z_j)}$

Where Y_i is the output probability for class i, z_i is the input to the output layer for class i, and C is the total number of classes.

These layers work together to extract and process information from the input image, producing an accurate classification or regression output.

Area of improvement

Performance degradation in compressed videos is particularly evident in hand-crafted features and shallow CNN architectures. However, neural networks have shown to be more adept at handling such scenarios, with XceptionNet demonstrating impressive outcomes in the face of weak compression while still maintaining a satisfactory level of performance even with low-quality images. This is attributed to its pre-training on ImageNet and its larger network capacity. Furthermore, there is a need to enhance the model's predictive capabilities when identifying GAN-based fake videos that have not been previously encountered. The research model is not open-sourced and is not publicly available. It runs on a private domain of the researchers, and the prototype model can only detect deepfakes within 30 seconds.

Summary

We have built the CNN network but are making use of transfer learning. The XceptionNet architecture is a standard CNN trained on the ImageNet dataset using separable convolutions with residual connections. We have adopted this pre-trained model for our task by replacing the final fully connected layer with two outputs. The remaining model layers have been initialized with the weights learned from the ImageNet dataset. To set up the newly inserted fully connected layer, we have fixed all the weights to the final layers and pre-trained the network for three epochs. Subsequently, we have trained the network for an additional 15 epochs, and the best-performing model has been chosen based on the validation accuracy. A comprehensive description of our training methodology and hyperparameters can be found in the methodology section of this paper.

Results

Binary classification accuracies are calculated by comparing the predictions made by a binary classification model to the true labels of the corresponding data samples. The accuracy score measures how often the model correctly predicts the class of the data sample.

To calculate binary classification accuracy, the number of correctly classified samples (true positives and true negatives) is divided by the total number of samples in the dataset.

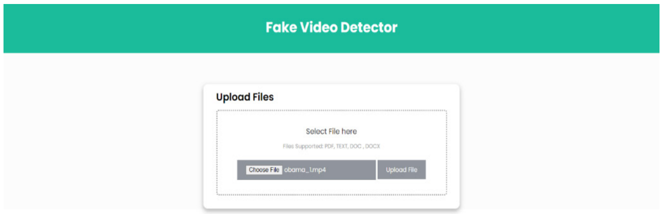
Mathematically, this can be represented as:

Accuracy = (True positives + True negatives)/Total number of samples

Where true positives refer to the number of samples correctly classified as positive, true negatives refer to the number of samples correctly classified as unfavorable, and the total number of samples refers to the sum of true positives, true negatives, false positives, and false negatives.

The XceptionNet model's predictive performance was evaluated across three levels of image quality, namely Raw, HD, and LD. The corresponding accuracies achieved by the model were 92.24, 87.73, and 80.30%, respectively.

Sample prototype deepfake detection model webpage



Fake Video Detector

Upload Files

Select File here

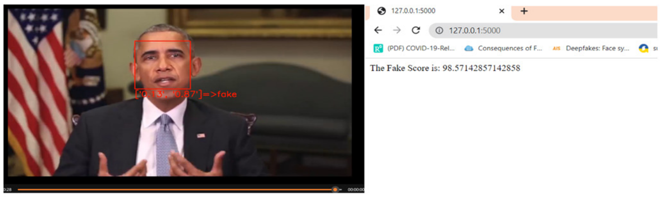
Files Supported: PDF, TEXT, DOC , DOCX

Choose File obama_1.mp4

Figure 5: The prototype model "Fake video detection webpage

Findings

Out of 50 Deepfakes testing videos, the prototype model detected 35 deepfakes videos right out of 50. The accuracy of the prototype model is 70% as of today. The researcher keeps the benchmark of 65% and above as detected as "fake" and 65% and below as detected as "real". 15 videos were detected as real by the prototype model even though they were fake. This could be because of the limitation of the model and the preciousness of those deepfake videos. They are extremely convincing and were a perfect deepfake.



The Fake Score is: 98.57142857142858

Figure 6: The research prototype model detection process and the Fake score

Suggestions for manual detection of deepfakes

- Artificial facial expressions: Most advanced Deepfakes tools attempt to manipulate and morph facial expressions. This is done to make the fake video look most convincing. Artificial facial movements are commonly identified when the source's emotions do not resonate with their speech.
- Pay attention to eye movement: Eyes are crucial in manufacturing deepfakes videos. Any unnatural eye movement. It is not easy to replicate or imitate the natural eye movement of a person while speaking and moving. The eye movement might not match their facial expression, and sometimes, the person in the fake video might not blink their eyes.
- Check the audio quality: A deepfake video focuses on the video rather than the audio. So, the video might be morphed, but the creators often do not focus on the audio quality. The fake videos are packed with bad audio quality, different voice tones, robotic voice, weak or wrong pronunciation, poor lip-sync, background noise, or, in some cases, no audio.
- Unnatural body movement: When deepfake videos move from one frame to another, there will be a glitch or distorted connection of the body movement from the first frame to the next. The continuity of body movements will be disturbed when the source moves their head or turns to the side.
- Abnormal Body posture or physique: Most deepfake instruments concentrate on facial features and movements. Suppose a person's head, hand, or body is positioned strangely and awkwardly, and their body shape appears more unnatural.
- The discrepancy in the lighting: Strange lighting in the video, different lighting from the first clip to the following clip, or oddly placed shadows might indicate that the video is fake.
- Pay attention to the skin and facial hair: If the skin of the person in the video appears to be too smooth, too fair, too dark, too wrinkly, or with different skin textures, then the video might be fake. Comparing the person's skin with the hair, eyes, and body can be an excellent way to analyse the video.

Who is most vulnerable to deepfakes?

"What a perfect tool for somebody seeking to exert power and control over a victim."

-Adam Dodge

Revenge porn targeted towards Women: According to Sensity AI, a research firm that has been monitoring them since December 2018. Most deepfakes target women, between 90% and 95% of deepfake movies on the internet are nonconsensual porn, and 90% of that is nonconsensual female porn. Producing and distributing sexually explicit photos or movies that have been altered using deepfakes technology are known as deepfakes revenge porn. Without the victim's knowledge or consent, the photographs or films are typically made by taking the victim's face and superimposing it onto another person's body in a sexually suggestive context. Deepfake porn reduces women's dignity to a mere sexual object and makes them question their identity. The victim of this kind of crime may experience extreme emotional discomfort due to the grave breach of their privacy and, in some cases, material harm like financial loss or career loss.

Application of Deepfakes

Deepfakes have various drawbacks; Deepfakes can be used to produce false information or propagate fake news. Public opinion, political campaigns, and even national security may all suffer due to this. Without the subject's permission, phony movies or pictures of them can be made using deepfakes. This might be employed maliciously for activities like revenge porn or blackmail, compromising an individual's privacy, dignity, and identity. Fake news can make people lose faith in the media and other channels of Communication, reducing their trust in the fourth pillar of democracy. While deepfakes are known for having adverse effects, they can also be used for good. For instance, deploying deep fakes in the entertainment sector can result in immersive and realistic experiences. For example, a deepfake may reconstruct historical events or stage a virtual performance by a deceased superstar. Deepfakes can be utilized for educational purposes to produce dynamic, engaging instructional content, create art, and magnify awareness.

Conclusion and Discussions

Deepfakes are an increasing cause for concern since they can be used to propagate false information, fabricate news, or deceive people. While Artificial intelligence is used to create deepfakes, the same technology is also used to detect them. However, technological detection is not the permanent solution to the menace of deepfake. Information warfare results from technological advancement with no or minimal media literacy. A desirability bias causes us to frequently see what we desire to be true (Martijn Rasser, 2019).

Cognitive biases theory suggests that when an individual pays attention to news stories that only confirm their ideology, it adds to the problem of deepfakes. In the age of the tabloid news ecosystem, sensational headlines often exaggerate and are the pivotal cause of disinformation. The deepfakes creators take advantage of this formula and create short video deepfakes that are easy to make and forward with maximum impact. With the mistrust, underlying biases and political disagreement can trigger the development of echo chambers and filter bubbles, leading to communal unrest. The social identity theory suggests that people divide society into in-groups and out-groups. Such a perspective leads to political polarisation. The researchers observed that the growing number of deepfakes in social media are mainly targeted toward women and politics. Media literacy is a practical and functional solution to deepfakes, and decisive government intervention will be necessary to counter deepfakes. Education and media literacy continue to be the critical lines of defense against disinformation as there is no simple technology answer in sight, much like disinformation and fake news (Johannes Langguth, 2021**)**. The government, media, and society have to have to work together. We have a higher chance of reducing the deep fake threat if we recognize and correct our biases in ourselves and others (Martijn Rasser, 2019). The spread of a deep fake can be decelerated by teaching individuals to think carefully before mindlessly sharing a disturbing video or questionable media content. To conclude, the researcher highlights the need for psychological change in society, responsible media reporting, timely government intervention, powerful media literacy in all tiers, and a robust technological model, which is suggested as the solution to counterfeit media content.

Further Studies

The researchers believe it would be worth exploring the psychological trauma of the deepfake victims and understanding the underlying issue of deepfake dissemination. The deepfakes detection model is one of the key instruments for detecting deepfakes. However, a long-term solution would be media literacy. A study on the current media literacy and awareness among the public would play a significant role in resolving the foundational issue of deepfakes and giving an effective technological solution. Due to the limitations of resources, the researchers could only propose a working research face detection deepfakes model. A future study could also involve a robust model that will detect even detect body morphing, which is extensively used in revenge pornography.

References

1. Abdullah-All-Tanvir, Mahir, E. M., Akhter, S. and Huq, M. R. (2019). Detecting fake news using machine learning and deep learning algorithms. *7th International Conference on Smart Computing &amp; Communications (ICSCC)*. https://doi.org/10.1109/icscc.2019.8843612
2. Ahmed, Choi, Praveen, Myung, Donepudi, Ayman, A., Aljarbouh. and Ayub, A. A. (2021). Detecting fake news using Machine Learning: A Systematic Literature Review. *Psychology and Education Journal*, *58*(1), 1932–1939. https://doi.org/10.17762/pae.v58i1.1046
3. Ahmed, H., Traore, I. and Saad, S. (2017). Detection of online fake news using N-gram analysis and Machine Learning Techniques. *Lecture Notes in Computer Science*. 127–138. https://doi.org/10.1007/978-3-319-69155-8_9
4. De Lima, O., Franklin, S., Basu, S., Karwoski, B. and George, A. (2020). *Deepfake detection using spatiotemporal convolutional networks*. arXiv.org. https://arxiv.org/abs/2006.14749
5. Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M. and de Alfaro, L. (2018). Automatic online fake news detection combining content and social signals. *22nd Conference of Open Innovations Association (FRUCT)*. https://doi.org/10.23919/fruct.2018.8468301
6. Dewey, C. (2021). *Analysis facebook has repeatedly trended fake news since firing its human editors*. The Washington Post, https://www.washingtonpost.com/news/the-intersect/wp/2016/10/12/facebook-has-repeatedly-trended-fake-news-since-firing-its-human-editors/
7. Durall, R., Keuper, M, Pfreundt, F.-J. and Keuper. (2020). *Unmasking deepfakes with Simple features*. arXiv.org. https://arxiv.org/abs/1911.00686
8. Fagni, T., Falchi, F., Gambini, M., Martella, A. and Tesconi, M. (2021). Tweepfake: About detecting deepfake tweets. *PLOS ONE*, *16*(5). https://doi.org/10.1371/journal.pone.0251415
9. Hao, K. (2021). *Deepfake porn is ruining women's lives. Now the law may finally ban it.* MIT Technology Review. https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/
10. Hao, K. (2022). *A horrifying new AI app swaps women into porn videos with a click*. MIT Technology Review. https://www.technologyreview.com/2021/09/13/1035449/ai-deepfake-app-face-swaps-women-into-porn/
11. Huang, Y., Juefei-Xu, F., Wang, R., Guo, Q., Ma, L., Xie, X., Li, J., Miao, W., Liu. and Pu, G. (2020). Fakepolisher: Making deepfakes more detection-evasive by shallow reconstruction. *Proceedings of the 28th ACM International Conference on Multimedia*. https://doi.org/10.1145/3394171.3413732
12. Kaur, S., Kumar, P. and Kumaraguru, P. (2019). Automating fake news detection system using multi-level voting model. *Soft Computing*, *24*(12), 9049–9069. https://doi.org/10.1007/s00500-019-04436-y
13. Kesarwani, A., Chauhan, S. S. and Nair, A. R. (2020). Fake news detection on social media using k-nearest neighbor classifier. *International Conference on Advances in Computing and Communication Engineering (ICACCE)*. https://doi.org/10.1109/icacce49060.2020.9154997
14. Kurasinski, L. and Mihailescu, R.-C. (2020). Towards machine learning explainability in text classification for fake news detection. *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*. https://doi.org/10.1109/icmla51294.2020.00127
15. Langguth, J., Pogorelov, K., Brenner, S., Filkuková, P. and Schroeder, D. T. (2021). Don't trust your eyes: Image manipulation in the age of deepfakes. *Frontiers in Communication*., *6*. https://doi.org/10.3389/fcomm.2021.632317
16. Maddocks, S. (2020). 'A deepfake porn plot intended to Silence me': Exploring Continuities between pornographic and 'political' deep fakes. *Porn Studies*, *7*(4), 415–423. https://doi.org/10.1080/23268743.2020.1757499
17. Nguyen, T. T., Pham, Q.-V. and Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding. 223*, 103525. https://doi.org/10.1016/j.cviu.2022.103525
18. Ni, B., Guo, Z., Li, J. and Jiang, M. (2020). *Improving generalizability of fake news detection methods using propensity score matching.*, arXiv.org. https://arxiv.org/abs/2002.00838
19. Pratiwi, I. Y., Asmara, R. A. and Rahutomo, F. (2017). Study of hoax news detection using naïve Bayes classifier in Indonesian language. *2017 11th International Conference on Information &amp; Communication Technology and System (ICTS)*. https://doi.org/10.1109/icts.2017.8265649
20. Rastogi, S., Mishra, A. K. and Gaur, L. (2022). Detection of deepfakes using local features and Convolutional Neural Network. *DeepFakes*, 73–89. https://doi.org/10.1201/9781003231493-6
21. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J. and Niessner, M. (*2019).* FaceForensics++: Learning to detect manipulated facial images. *IEEE/CVF International Conference on Computer Vision (ICCV)*. https://doi.org/10.1109/iccv.2019.00009
22. Shin, S. Y. and Lee, J. (2022). The effect of deepfake video on news credibility and corrective influence of cost-based knowledge about deepfakes. *Digital Journalism*., *10*(3), 412–432. https://doi.org/10.1080/21670811.2022.2026797

23. Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H. (2017). *Fake news detection on social media: A Data Mining Perspective*. arXiv.org. https://arxiv.org/abs/1708.01967
24. Siwei Lyu., *Detecting "deepfake" videos in the blink of an Eye*. (2022). The Conversation., https://theconversation.com/detecting-deepfake-videos-in-the-blink-of-an-eye-101072
25. Tan, K. L., Poo Lee, C. and Lim, K. M. (2021). FN-net: A deep convolutional neural network for fake news detection. *2021 9th International Conference on Information and Communication Technology (ICoICT)*. https://doi.org/10.1109/icoict52021.2021.9527500
26. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A. and Ortega-Garcia, J. (2022). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion.*, *64*, 131–148. https://doi.org/10.1016/j.inffus.2020.06.014
27. Wang, W. Y. (2017). "Liar, Liar Pants On fire": A new benchmark dataset for fake news detection. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. https://doi.org/10.18653/v1/p17-2067
28. Whittaker, L., Letheren, K. and Mulcahy, R. (2021). The rise of Deepfakes: A conceptual framework and research agenda for marketing. Australasian Marketing Journal. 29(3), 204–214. https://doi.org/10.1177/1839334921999479
29. Zhou, X., Wang, Y. and Wu, P. (2020). Detecting deepfake videos via frame serialisation learning. IEEE 3rd International Conference of Safe Production and Informatization (IICSPI). https://doi.org/10.1109/iicspi51290.2020.9332419
30. Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M. and Procter, R. (2018). Detection and resolution of rumours in social media. ACM Computing Surveys. 51(2), 1–36. https://doi.org/10.1145/3161603
31. Vatreš, A. (2022). 'Deepfake Phenomenon: An advanced form of fake news and its implications on reliable journalism', *Društvene i humanističke studije (Online),* 6(3(16)), pp. 561–576. https://doi:10.51558/2490-3647.2021.6.3.561.
32. Cover, R., Haw, A. and Thompson, J.D. (2022). 'The visual in an era of hyperreality and disinformation: The deepfake video', *Fake News in Digital Cultures: Technology, Populism and Digital Misinformation.*, pp. 63–76. doi:10.1108/978-1-80117-876-120221005.
33. Kaliyar, R.K., Goswami, A. and Narang, P. (2020). 'Deepfake: Improving fake news detection using tensor decomposition-based deep neural network', *The Journal of Supercomputing,* 77(2), pp. 1015–1037. doi:10.1007/s11227-020-03294-y.
34. Giansiracusa, N. (2021). 'Deepfake Deception', *How Algorithms Create and Prevent Fake News,* pp. 41–66. doi:10.1007/978-1-4842-7155-1_3.
35. Lyu, S. (2021). 'Fighting Ai-synthesised fake media', *Proceedings of the 1st Workshop on Synthetic Multimedia - Audiovisual Deepfake Generation and Detection*., doi:10.1145/3476099.3482881.
36. Hwang, Y., Ryu, J.Y. and Jeong, S.-H. (2021). 'Effects of disinformation using deepfake: The protective effect of Media Literacy Education', *Cyberpsychology, Behavior, and Social Networking*, 24(3), pp. 188–193. doi:10.1089/cyber.2020.0174.
37. Han, B. *et al.* (2021).'Fighting fake news: Two stream network for Deepfake detection via learnable SRM', *IEEE Transactions on Biometrics, Behavior, and Identity Science.*, 3(3), pp. 320–331. doi:10.1109/tbiom.2021.3065735.
38. Jawad, ZA and Obaid, AJ. (2023). 'An overview of rumour and fake news detection approaches', *Advances in Multimedia and Interactive Technologies*, pp. 12–31. doi:10.4018/978-1-6684-6060-3.ch002.
39. Gaur, L., Ratta, M. and Gaur, A. (2022). 'Future of deepfakes and Ectypes', *DeepFakes*, pp. 135–145. doi:10.1201/9781003231493-11.
40. Mishra, S., Shukla, P.K. and Agrawal, R. (2022). Deepfakes, media, and societal impacts', *DeepFakes*, pp. 115–120. doi:10.1201/9781003231493-9.
41. Radoli, L. and Langmia, K. (2023). 'Of deepfakes, misinformation, and disinformation', *Black Communication in the Age of Disinformation*. pp. 1–13. doi:10.1007/978-3-031-27696-5_1.

HOW TO CITE THIS ARTICLE: Rajagopal, T., Chandrashekaran, V., Ilango, V. (2023). Unmasking the Deepfake Infocalypse: Debunking Manufactured Misinformation with a Prototype Model in the AI Era "Seeing and hearing, no longer believing.". *Journal of Communication and Management*, 2(4), 230-237. DOI: 10.58966/JCM2023243

Query Report

Q1 Kindly provide the figure intext citation (An in-text citation is an acknowledgment you include in your text whenever you quote or mention the figure/table or An in-text citation is the brief form of the reference that you include in the body of your work).