

**Article Citation Format**

Olebara, C.C. (2023): Towards Non-Techie Deepfake Detection: Role of Critical Thinking. Journal of Digital Innovations & Contemporary Research in Science, Engineering & Technology. Vol. 11, No. 2. Pp 1-6  
www.isteam.net/digitaljournal.  
dx.doi.org/10.22624/AIMS/DIGITAL/V11N2P1

**Article Progress Time Stamps**

Article Type: Research Article  
Manuscript Received: 17<sup>th</sup> April, 2023  
Review Type: Blind Peer  
Final Acceptance: 9<sup>th</sup> June, 2023

## **Towards Non-Techie Deepfake Detection: Role of Critical Thinking**

**Olebara, Comfort Chinaza**  
Department of Computer Science  
Faculty of Physical Science  
Imo State University,  
Owerri, Imo State, Nigeria.  
**E-mail:** chiiprime@gmail.com  
**Orcid:**0000-0002-5891-4206

### **ABSTRACT**

This paper is a short report on the practical methods of creating and detecting deepfakes. Most methods captured in research are based on Generative Adversarial networks for the creations, and autoencoder/decoders, which are deep learning-based, or ffmpeg which are code-based running in command prompt. Deepfake technology is the use of deep learning algorithms or media editing tools to create fake content with the aim of disinforming or misinforming the content viewers. The victims of the attack cut across members of society from all career foundations: IT and non-IT. This paper, therefore, aims to give the non-IT members of society, Critical Thinking Approaches through which deepfakes can be detected and also bring to their knowledge everyday tools for their creation and detection. The ethics of creating digital humans is also presented to serve as a guide, so laymen attempting these practical procedures are able to ascertain the acceptance level of digital humans and contexts under which they are accepted such as for learning, entertainment, and other non-disinforming circumstances. In conclusion, the advantages of critical thinking by non-techies as means of deepfake detection is highlighted, and the use of the Toddington CARS checklist of questioning the credibility, accuracy, reasonableness, and supporting information a video, audio, or image media content is recommended.

**Keywords:** Deepfake creation, GAN, Critical Thinking, Deepfake detection, Digital humans.

---

### **I. INTRODUCTION**

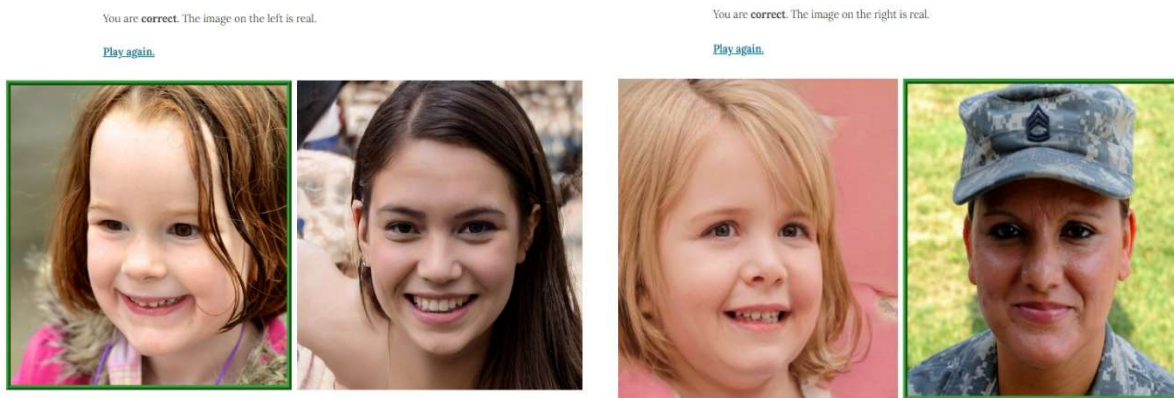
StyleGAN, neural network is used to randomly generate human images. The website, this-person-does-not-exist developed by Nvidia provides glue to identifying deepfake images. Some of these clues include: -analytically scrutinizing symmetrical objects such as eyeglasses and earrings, -scanning the background for out-of-place objects, unrealistic looking backgrounds(bokeh), as well as luminosity spread (vignette). Jevin

West and Carl Bergstrom created a website (which face is real) to further help professionals and non-professionals alike in mastering GAN mistakes that can be used to visually detect deepfake images such as artifact, bokeh, symmetrical objects creation, out-of-place objects, vignettes, etcetera. The learner is presented with a pair of images and the system returns correct/incorrect to his choice of real the image. to Pic 1 and 2 below are some examples of StyleGAN randomly generated images with symmetric object creation(earring) and out-of-place objects used for detection.



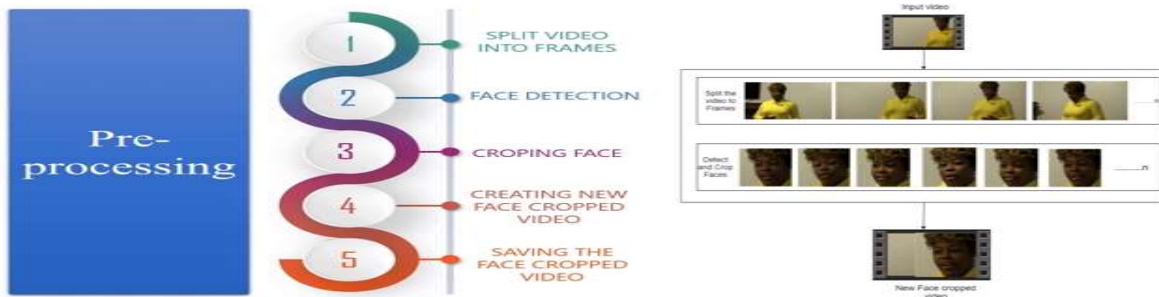
Figure 1: Symmetric object detection(earring)      Figure 2: out-of-place object detection  
**Source:** this-person-does-not-exist.com

The gamified version of this is shown in the Figures below **Source:** whichfaceisreal.com

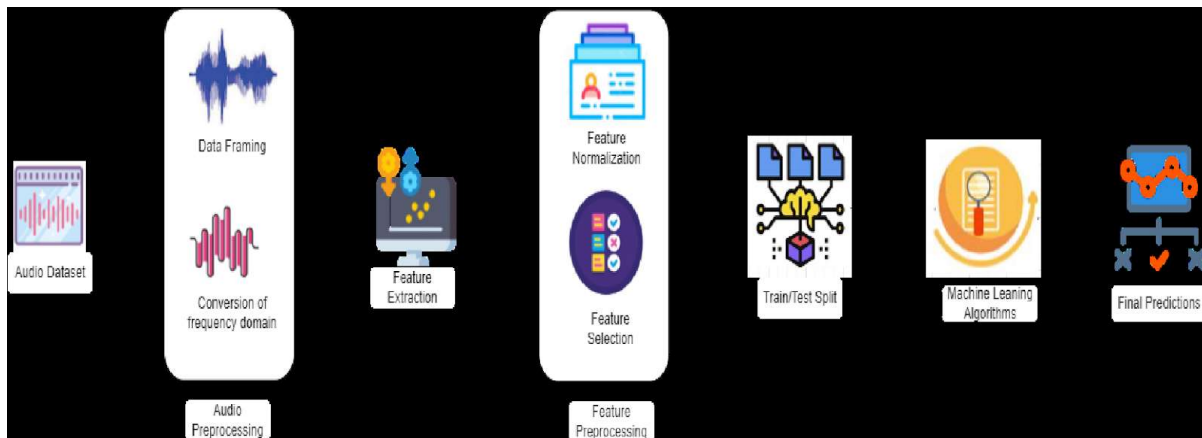


**Figure 3: Earring used for detection**      **Fig. 4 Unnatural background used for detection**  
**Source:** (Bergstrom & West, 2019)

In audio and video detection, the target audio and video is first split into frames using such editing tools as FFMPEG, using code: `ffmpeg -i videoname.mp4 image%2d.png` –for video splitting. Figure 5 and Figure 6 below shows the video and audio frame splits for feature extraction, and subsequent feature selection, as well as the sequence reorganization, all of which are part of the preprocessing step.



**Figure 5: Video Preprocessing**  
**Source: (Murthy, 2021)**



**Figure 6: Audio Preprocessing**  
**Source: (Iqbal et al., 2022)**

In these scenarios, CNN may be used for feature extraction while LSTM is used for sequence processing. Neural networks are the most popular approach for building deepfake.

Deep Neural Network (DNN) is powerful and the best method for finding complex relationships within data. They are based on human learning and can be used for more than classification (data aggregation).

Deep convolutional networks are based on deep neural networks with the convolution of neural representations. They can learn very complex relationships in images and videos. Declarative neural networks are based on network architecture with fully connected components. These are implemented using such trends as:

- ❖ classifies an image/video/audio as genuine or fake using binary classifiers
- ❖ Image Attribution: predicts which of a closed set of GMs generated a fake image
- ❖ Image parsing: infers hyperparameters of GMs used to generate an image for those models not identified during training.
- ❖ Analysis of Photo-Response Non-Uniformity: Analysis of noise patterns of light-sensitive sensors as a result of factory defects
- ❖ Reverse Engineering, etcetera.

However, these methods are outside the scope of this paper and will not be discussed in detail, as the author's intent is to provide a guide for non-IT audience to train their critical thinking ability. David Toddington, an open-source intelligence trainer as well as investigation and intelligence service provider, is the proposer of the Critical thinking concept, and he finds this as the best method for deepfake detection (Toddington International Inc., 2023).

## **2. ETHICS OF DIGITAL HUMANS**

A review of the ethics of this technology is required to ascertain the legality or otherwise of its usage. Digital human is the act of virtualizing people through the use of computer graphics. The introduction of AI and its subsets of machine and deep learning further enhances the look and feel of digitized humans. In (Redman, 2017) an industry veteran expressed concern over the rise and role of virtual people. Some of the roles include videos featuring dead people such as in holograms of Michael Jackson and 2Pac (King & Roos, 2014), (LealsMindless, 2012), is an extreme use of deepfake. In these hologram videos, paintings of dead legends are brought to life using lighting or LED technologies. The inventor of Hologram USA, Alki David in an interview with CNN's Mildones Donia described the hologram process as capturing of the digital head of Michael Jackson and an actor's body in real-time and re-rendering the video to give a synced, smooth effect. He identified the Military, Education, Politics, entertainment, etcetera, as possible application areas of the lighting and LED technology.

Buttressing, the reporter presented the hologram studio as having over 25 cameras and lots of light, (Donia, 2015). Also worrisome is the research by (Thies et al., 2016), which shows how webcams could be used to puppeteer targets from a previously existing video. The webcam uses a live capture as a modifier while the preexisting video of the subject is used. The output video can be re-rendered. This deepfake video can also be presented as a tool to misinform the viewing community. Similar research by (S. Suwajanakorn, S. M. Seitz, 2017) trains weekly speeches of President Obama, then uses neural networks to learn and map audio features into mouth shapes. Applying 3D pose matching to change what he seemed to be saying and syncing this with the audio output results in a near-perfect deepfake video. Of greater concern, therefore, is the ability to make people act and say things they never did or said in real life through AI-based synthesis. According to (Meskys et al., 2020) presently, there is no global regulatory or ethical guidelines for deepfakes creation, and early deployments were for malicious purposes such as in political scenarios.

However, the authors noted that a legal view to deepfake creation would be in the area of privacy invasion where no consensual agreement of the target was obtained, intellectual property disputes such as copyright infringement related to images, videos, speeches, or texts used in the creation of deepfake. They also identified possible exemptions such as obtainable in United States courts, where a liable defendant may invoke fair-use factors if the created deepfake is transformative or intended for morally good purposes. At national, countries like Singapore were found to have passed laws that make online false statements that affect public interest a punishable offense. Most research on ethical guideline for deep fakes, call creators of these contents to follow ethical best practices, and on cloud service providers such as Google, Amazon, and Microsoft, that provide tooling and platforms for scaling and fast creation of deep fakes, to be responsible for monitoring the morality of contents created with their innovations (Meskys et al., 2020), (Lane, 2020) cited in (Jaiman, 2020).

According to (Lane, 2020), today's technologies, though exciting and augmenting human capabilities, tend to infringe on basic human rights and even make light democracy, and without concern for humanity. The author therefore advocates for human-centered design thinking otherwise termed responsible innovation. Social media platforms, media organizations, policymakers, legislators, and journalists were also identified as having ethical responsibility toward the distribution of deep fakes (Jaiman, 2020).

### 3. CONCLUSION AND RECOMMENDATION

Technological innovations in 4IR and 5IR have brought disruption in the normal way of living. The use of digital humans has pros and cons, depending on intent. While both deepfakes and 3D human avatars are termed digital humans, the generation of "fake" representation or an individual's voice, face, video or still image usually end up with the content being put to malicious use. 3D avatar, though digitized, does not attempt to replace or misrepresent the original person nor is it aimed at misinforming or disinforming the viewing audience.

Understanding things to look out for when presented with a suspected face has to do with not taking things at face value in the era of Artificial Intelligence. The adoption of *Critical Thinking* by non-techies, as a means of identifying deepfakes has its high points over AI-based detection methods, when it comes to investigating real media content. Some of these are:

- AI shows great speed at learning from supplied data and making inferences, but cannot get into the field to look for specific pieces of information
- AI does not think critically
- AI cannot change its mind about an ongoing investigation process since it does not have the knowledge and must be fed with human knowledge
- Deep investigations require application of human elements such as instincts, and AI is not capable of applying instincts on its own.

### 4. RECOMMENDATION

Following the CARS list of when faced with deepfake content requires asking the questions:

What?, Where?, & How? Culminating in the Toddington CARS checklist:

The Information received CARS checklist

- ❖ -Credibility: Why should I believe?
- ❖ -Accuracy: How recent?
- ❖ -Reasonableness: Research-based, originator's view, possible hidden intents behind the information
- ❖ -Support: What related information on this target?

## REFERENCES

1. Bergstrom, C., & West, J. (2019). *which face is real*. <https://www.whichfaceisreal.com/>
2. Donia, M. (2015). (49) *How did they create Michael Jackson's hologram~ CNN - YouTube*. YouTube.Com. <https://www.youtube.com/watch?v=fDR3hSeQ8pk>
3. Iqbal, F., Abbasi, A., Javed, A. R., Jalil, Z., & Al-Karaki, J. (2022). Deepfake Audio Detection via Feature Engineering and Machine Learning. *CEUR Workshop Proceedings*, 3318.
4. Jaiman, A. (2020). *Debating the Ethics of Deepfakes*. In a Pandemic World. <https://www.orfonline.org/expert-speak/debating-the-ethics-of-deepfakes/>
5. King, J., & Roos, S. (2014). (41) *Michael Jackson - Slave To The Rhythm - YouTube*. <https://www.youtube.com/watch?v=jDRTghGZ7XU>
6. Lane, M. (2020). *Responsible Innovation: The Next Wave of Design Thinking* | by Mira Lane | Microsoft Design | Medium. Microsoft Design on Medium. <https://medium.com/microsoft-design/responsible-innovation-the-next-wave-of-design-thinking-86bc9e9a8ae8>
7. LealsMindless. (2012). (41) *Tupac Hologram Snoop Dogg and Dr. Dre Perform Coachella Live 2012 - YouTube*. <https://www.youtube.com/watch?v=TGbrFmPBV0Y>
8. Meskys, E., Kalpokiene, J., & Jurcys, P. (2020). *Regulating Deep-Fakes : Legal and Ethical Considerations. January*. <https://doi.org/10.1093/jiplp/jpz167>
9. Murthy, B. (2021). *DEEPAKE DETECTION*. <https://doi.org/10.1016/j.artmed.2021.102017>
10. Redman, R. (2017). *The ethics of digital humans* | Creative Bloq. <https://www.creativebloq.com/news/the-ethics-of-digital-humans>
11. S. Suwajanakorn, S. M. Seitz, I. K.-S. (2017). *Synthesizing Obama: Learning Lip Sync from Audio*. <http://grail.cs.washington.edu/projects/AudioToObama/>
12. Thies, J., Zollh, M., Christian, S., & Matthias, T. (2016). *Face2Face: Real-time Face Capture and Reenactment of RGB Videos*.
13. Toddington International Inc. (2023). *The Global Leader in Open Source Intelligence*. <https://www.toddington.com/>