**DEEPFAKES AS AN EMERGING THREAT IN NIGERIA**

**INTRODUCTION**

1. Technology has brought the world closer than ever, nations across the world are faced with different security challenges with peculiar similarities when it involves cybercrimes. Deepfakes in particular, as emerging threats has the potential to cause major security concerns. More so, as it depicts events of people doing or saying things they never said or being in places they never were.

2. In late 2017, a video showing the face of Gal Gadot superimposed on an existing pornographic video made it appear like the actress was engaged in the act depicted.1 The term “deepfakes” is derived from the fact that the technology involved in creating these specific styles of manipulated content (or “fakes”) involves the use of deep learning techniques. Deep learning represents a subset of machine learning techniques, which are a subset of artificial intelligence. In machine learning, a model uses training data to develop a model for a specific task. The more robust training data, the better the model gets. In deep learning, a model can automatically discover representations of features in the data that permit classification or analysis of the data. They are effectively trained at a “deeper” level. The data which can be examined using deep learning is not restricted to images and videos of people. It can include images and videos of anything, as well as audio and text. 2

3. This technology has been deployed to several ends around the world. Whether it is a video of Barack Obama calling Donald Trump a “complete dipshit”,3 or Mark Zuckerberg bragging about having “total control of billions of people’s stolen data”.4 The novelty of deepfake know-how has been misused to defame an individual’s character, create political havoc, disseminate fake news via social media platforms, and threaten stability and democracy internationally. As recent as Aug 22, the negative use of deepfakes in electoral misconduct presents the foremost concern in Africa. During the final stage of Kenyan presidential campaigns, it witnessed propaganda, manipulated videos and fabricated audios online, signalling a wave of deep fakes. These were meant to misrepresent the candidate’s opinion and fake statements claiming foreign support for a specific candidate, the deep fakes targeted Kenyan voters.5

4. On Sep 21, the Nigerian Minister of Information and Culture, raised the alarm about deepfakes being used to discredit the government, spreading what he termed “deep fake news”. The minister further noted that at the beginning, sources of deep fake news were largely unknown online publications. The minister recalled the aftermath of an interview by Channels Television of Benue Governor, Samuel Ortom and a retired Navy Commodore, Kunle Olawunmi. “After the interview, online and traditional newspapers were inundated with stories that the station had been shut, while some said that the anchor people had been arrested. Some even went to demonstrate the purported arrest of the anchor people, but till today, the station is still open.6 Therefore, Deepfakes may not only pose threats to individuals or groups but also erodes trust in institutions, as well as have national security implications.

5. The purpose of this paper is to highlight the emergence of deepfakes as a tool that could further encourage terrorism, armed banditry, and electoral violence amongst others in Nigeria. The paper will give an overview of deepfakes, and how non-state actors may apply them to further their nefarious activities. Thereafter it would discuss the mode of detection, challenges of combatting deepfakes in Nigeria, and countermeasures. The paper would be limited to the period of 2017 – 2022 as the term deepfake was first used in the year 2017.

**AIM**

6. This paper aims to discuss the challenges of deepfakes as an emerging threat in Nigeria with a view to making recommendations.

**OVERVIEW OF DEEPFAKES AS AN EMERGING THREAT IN NIGERIA**

7. Deepfakes is one of the most rapidly changing technologies available today and they refer to Al-generated synthesised and/or overlaid images and videos utilising existing photos and videos for deception purposes. This is done by matching various biometric data points of a person to the inputs of another person's facial structure and then replacing the target's face with the faces from the training data to create the synthesised/overlaid video. Deepfake technology has seen dramatic growth in popularity as the underlying Al technology has advanced. This can be evident by the number of videos shared on Facebook and other social media platforms which are doctored using this technology. Such videos can then be uploaded onto the platform by the user and subsequently spread rapidly through the network before the content is detected by moderators.7

8. The ability to deceive using these fabricated videos has far-reaching consequences not only for unsuspecting individuals but also for national security since it allows perpetrators to spread misinformation and incite violence on an unprecedented scale. The style is based on generative adversarial networks (GANs) which are a class of machine learning algorithms which consist of two competing neural networks, one which is a generator and the other which is a discriminator.8 The generator attempts to generate an image that the discriminator cannot distinguish from the real image while the discriminator attempts to distinguish between real and fake images. The generator8 attempts to fool the discriminator by mimicking real images more closely. The discriminatorattempts to distinguish real from fake images. The adversarial game between these two networks trains both networks. The generator eventually learns to mimic real images better and the discriminator learns to distinguish real from fake images. This process is repeated until the generator eventually produces images which are almost indistinguishable from real images and the discriminator cannot distinguish between real and fake images. With the proliferation and the availability of sophisticated open-source deep learning tools, individuals can superimpose faces, and facial expressions and synthesise their speech.

9. What once took weeks and sometimes days and significant skills to create a convincing deepfake, now only takes a matter of hours to create with widely accessible online applications and software programs. Although most deepfakes are been identified to have negative impacts, they can also have positive impacts on human development and the global society at large.

10. **Benefits of Deepfakes**. Deepfakes can also have positive impacts on society, even as it has a high tendency to be deployed negatively. It also has legitimate benefits, with some articulated below;

a. **Education**. The use of deepfakes in education as part of classroom teaching is a benefit, especially in underdeveloped countries which do not have adequate resources. Therefore, outdated methods and technologies are used to teach students. However, old lecture recordings of individuals no longer living have been used to emphasize teaching by using the expertise of historical figures in their classroom settings through AI.9 Deepfake technology was used in an exhibition at The Dalí (Salvador Museum) in St. Petersburg, Florida. Using machine learning, the exhibit featured a life-size of The Dalí delivering a variety of quotes he had spoken during interviews throughout his career.10 Deepfakes can also be used to enhance techniques used in modelling and simulation, as well as training and education.11

b. **Entertainment**. With deepfake technology, moviemakers would be empowered to have more freedom of expression to recreate classic scenes or to create new movies using dead actors. Imagine for a moment the late Sherman Kent, who is often described as “the father of intelligence analysis” synthesised into a movie as a super spy. In a 2019 global malaria awareness campaign, the industry used deepfake technology to create a video to attract diverse audiences in different languages. The advertisement featured David Beckham delivering the video in multi-languages created with visual and voice-altering technology.12

c. **Health Care**. The possibilities Deepfake expertise could provide social and healthcare industry could be admirable. With the advances in AI, deepfakes can help individuals diagnosed with Alzheimer’s disease to potentially remember their past with the interaction video contact of them with a younger facial appearance.12 Deepfakes can be used to recreate images through video and audio manipulation of style and speech of recent lost loved ones for individuals to reconnect with them to aid in their grieving process. Scientists are using the benefits of medical diagnosis by exploring generative models to detect medical abnormalities in x-ray images.12

d. **Public Safety & Digital Reconstruction**. Reconstructing crime scene is a forensic science and art, using inductive and deductive reasoning and evidence. AI-Generated synthetic media can help reconstruct the scene with the interrelationship of spatial and temporal pieces. In 2018, a team of civil investigators used cell phone videos, autopsy reports, and surveillance footage to reconstruct a virtual crime scene.130

11. **Social Media and Fakes**. The internet has turned the world into a global village, where a person in Nigeria can converse with another person in Britain instantly through social media. However, social media as a medium has introduced risk to society and has made disinformation, misinformation and fake news a bigger security concern. Some areas of concern such as disinformation, and differentiating between real and fakes as well as misinformation and fake news are highlighted below:

a. **Disinformation**. Disinformation campaigns are pervasive and are often facilitated by social media bots or state-sponsored trolls who create and distribute propaganda and disinformation to influence public opinion. Fake social media accounts with fake profiles, posts and pictures are used as fronts to push particular agendas and promote political interests. In some cases, they are used to infiltrate and spy on groups and organizations in order to expose secrets and spread disinformation among activists. This leaves social media platforms who are not fully vetted and who do not regulate content properly vulnerable to manipulation. For example, Twitter can be used to spread fake news about a particular topic to promote a particular agenda or it can be used to spread rumours to discredit certain groups in society and create chaos. However, prominent personalities such as President Muhammadu Buhari have witnessed censorship on Twitter for publicly addressing internal security matters which appeared to be sabotage within South Eastern Nigeria.

b. **Differentiating Between Real and Fake**. Fakes vs real

c. **Misinformation or Disinformation? and Fake News**. Misinform