Rising Threat - Deepfakes and National Security in the Age of Digital Deception

Dougo Kone-Sow
Old Dominion University

Rising Threat: Deepfakes and National Security in the Age of Digital Deception

Dougo Kone-Sow, Old Dominion University

I. Introduction

As technology advances at an unprecedented pace, what was once confined to the realm of science fiction has now materialized into a disconcerting reality, exemplified by the emergence of deepfakes. Coined by a Reddit user bearing the same moniker, the term "deepfake" blends the concepts of "deep" to signify AI deep learning neural network technology and "fake" to denote fabrication (Somers, MIT). In essence, deepfakes represent digitally altered or entirely synthesized media, encompassing images, videos, and audio recordings, all crafted with the aid of machine learning-based techniques (Galloway, Popular Science).

Though deepfakes have garnered significant attention in recent years, the roots of this technology trace back to the early developments of disinformation campaigns, before the eras of electronic computers. Deepfakes are disinformation involving the spreading of false information with intent. In the late 1800s, limitations in camera technology led to a mixture of authentic and fabricated footage from the Spanish-American war, intended to stoke patriotism among American viewers (Galloway, Popular Science). In 2021, the U.S. Capitol faced an invasion by a mob of Trump supporters who, influenced by former President Trump, falsely believed that the election results were fraudulent despite evidence to the contrary (BBC News). The trajectory of disinformation campaign continued well into the 1990s, as other forms of disinformation techniques developed within academic research circles and, subsequently, among enthusiasts in online communities (qtd. in Porter, The Verge).

Deepfakes rely on the integration of the target subject and the source material in a media form, which the creator intends to have digitally manipulated (Galloway, Popular Science). The implications of deepfake technology resonate far beyond individual cases, extending to potential ramifications for American national and economic security. In light of the susceptibility of Americans to foreign information warfare influence through deepfakes, and the looming concerns for the American national security and economy, strategies for countering this emerging threat is essential. With the new Presidential Executive Order 14110 focusing partially on deepfake threats, the U.S. government indicates its interest to mitigate threats from this novel use of technology.

At a minimum, understanding deepfakes' capabilities is paramount to ensuring the safety and security of all individuals in this rapidly evolving digital landscape.

II. Capabilities of Deepfakes

Deepfakes use artificial deep neural network technologies to develop false but realistic looking text, images, videos, and audio media. Operating on a machine learning-based system, deepfakes require an array of media encompassing videos, images, and/or audios from both the target subject (i.e., individual whose identity and likeness will be used for the synthetic media) and source material to craft convincingly realistic but fake media (Galloway, Popular Science).

For instance, a Twitch streamer—a video platform and community for gamers—who goes by the username QT Cinderella, discovered that someone appropriated her likeness to create a deepfake video that was explicit content (Farokhmanesh, Wired). Such creation of deepfake video is possible when the deepfake creator possesses an ample collection media depicting the target individual and the backdrop source the creator uses to "insert" the target individual into

(Galloway, Popular Science). Notably, anyone can become a target, irrespective of their online presence, because deepfake creators can use images and videos depicting the individual from other websites to build a dataset for creating deepfakes. Also, numerous deepfakes *mobile* applications are entering the market, making it easy for anyone to create deepfake media to commit cybercrimes, such as phishing attacks.

Traditionally, phishing attacks involve usually impersonating authoritative figures via phishing emails for inducing a target individual to commit actions that later will be detrimental to the individual's organization. But audio deepfakes allow threat attackers to impersonate someone over phone, allowing the phishing attack to be more effective.

Notable applications capable of creating deepfake media include LOVO.ai—an AI-based voice generator—DeepFaceLab—an open-source system for manipulating video imagery— and Faceapp—a revolutionary facial and body modifying Russia-based application.

With the proliferation of easily accessible deepfake applications, the risks of misinformation and manipulation grows, causing great concern due to their profound effects particularly within the realm of American national security.

III. Threats to National Security and the Potential Consequences of Disinformation

The fallout from a deepfake attack can have devastating effects on an individual's mental well-being and public perception of them. Should a scammer succeed in using AI for deceit and extortion, the consequences could lead to severe financial repercussions, potentially causing bankruptcy for the victim, impacting both their livelihood and financial stability. These crimes

are not limited to individuals but can also target companies and the national treasury, posing a threat to overall security.

During the 2020 elections, deepfake videos were utilized to defame President Joe Biden, portraying him as afflicted by dementia during a speech. The video circulated widely online, leading millions to believe its authenticity (O'Sullivan, CNN). If the debunking of this video hadn't occurred, it could have significantly disrupted Biden's successful election as President of the United States. Such dissemination of misinformation by a prominent political figure resulted in civil and public unrest in America.

Civil unrest within a country can have adverse implications for national security, potentially encouraging foreign adversaries to exploit the turmoil. Illustrating this point, Langa notes that:

> …bad actors could use deepfakes to impersonate military or intelligence officers "ordering the sharing of sensitive information or taking some action that would expose forces to vulnerability."[51] Combined with a cyberattack, such as a hack of a news organization's website or of a trove of government documents, a deepfake could be widely disseminated and threaten an entire governmental regime or national economy (Langa).[52]

In summary, the profound impacts of deepfakes and the nefarious use of artificial intelligence technologies on American national security are likely extensive and must be addressed effectively. Recognizing this urgency, President Joe Biden issued Executive Order 14110, among many things, to manage the risks that artificial intelligence poses to national and economic security.

IV. The Implication of Executive Order 14110 with Respect to Deepfakes

President Biden's Executive Order 14110 is poised to create significant ripples across all national industries. It focuses on managing the substantial risks associated with AI. Building upon previous federal government's efforts on AI, Executive Order 14110 "…establishes new standards for AI safety and security, protects Americans' privacy, advances equity and civil rights, stands up for consumers and workers, promotes innovation and competition, advances American leadership around the world, and more." (The White House) Concerning AI safety and security, several steps were listed by the Executive Order to mitigate potential risks. This includes the requirement for companies developing foundational models to share their safety results with the American government, aimed at averting severe risks to overall national security. Moreover, the Executive Order proposes steps to further develop safety standards, tools, and tests to ensure the trustworthiness of AI systems. Additionally, there's a focus on establishing standards and best practices to combat deepfake occurrences, preventing disinformation and safeguarding the privacy of Americans. (White House)

Efforts are underway to ensure that all AI systems function as intended and are thoroughly vetted to prevent malicious use. Plans involve the development of effective methods and mechanisms for watermarking and labeling AI-generated content that are mainly originated by federal government agencies. These precautions aim to counteract disinformation spread by adversaries both within and outside the nation's borders. While primarily aimed at protecting American national security, the Biden Administration's actions set examples for companies and businesses to follow in the evolving AI landscape. President Biden emphasized in section 2, part e that "[t]he Federal Government will enforce existing consumer protection laws and principles

and enact appropriate safeguards against fraud, unintended bias, discrimination, infringements on privacy, and other harms from AI." (The White House) These safety measures will fortify the protection of Americans and consumers across various industries, including healthcare, financial services, and law.

In the realm of AI advancements, President Biden highlighted in Executive Order 14110 how AI has facilitated deceptive, manipulative, and unauthorized access to civilians' private information. To counter the risks of exploitation and unauthorized exposure, section 2, part f mandates federal agencies to utilize ". . . available policy and technical tools, including privacy-enhancing technologies (PETs) where appropriate, to protect privacy and to combat the broader legal and societal risks — including the chilling of First Amendment rights — that result from the improper collection and use of people's data." (The White House) Executive Order 14110 requires that within 240 days, the Secretary of Commerce report to the Director of Office of Management and Budget (OMB) and the Assistant to the President for National Security Affairs on standards, methods, practices, and science-approved techniques, encompassing content verification, tracking, and preventing deepfake explicit material involving children and non-consenting adults. Moreover, within 270 days of the signing of the Executive Order, government agencies in the national security arena must addressing potential misuse of AI systems by adversaries threatening national security. Additionally, the Executive Order mandates that agencies must limit specific AI generative platforms based on risk assessments and, within 150 days, issue best practices for financial institutions to manage AI-specific cybersecurity risks. (White House)

President Biden's Executive Order demonstrates a comprehensive approach to addressing AI and deepfake-related concerns, ensuring effectiveness in national security, financial stability,

and public policy. It reinforces existing laws and policies, doubling down on the prevention of privacy infringement while taking further steps to ensure national security. The government's involvement in countering disinformation is crucial, complementing the need for Americans to educate themselves about deepfakes and understand mitigation strategies against potential attacks.

V. Mitigating Threats and Possible Solutions

While the steps taken by the Biden Administration are significant strides toward improving national security and the safety of all citizens, an overarching solution to address deepfake-related concerns involves proactively predicting and preventing deepfake attacks. This proactive stance aims to curtail disinformation affecting citizens and governmental sectors. One potential approach lies in developing specialized sensors or an intricate process explicitly designed to detect deepfake content online (Langa).

These sensors would discern between malicious and benign use of such content, enabling immediate removal from the internet and aiding in tracing perpetrators, irrespective of whether they are American or foreign assailants. This preemptive measure could substantially reduce the spread of harmful disinformation, safeguarding individuals and government entities from the adverse effects of deepfake attacks. Considering the current absence of such a solution, promoting media literacy and critical thinking becomes even more urgent (World Economic Forum).

It's crucial to identify signs in multimedia such as unnatural blinking, poor lighting, robotic voices, misaligned facial features, and discrepancies in voice-to-lip synchronization (Rekhi, ETtech). Additionally, it's vital to approach online information with skepticism and seek

corroborating evidence. Often, individuals and media outlets fall into the trap of "click-baiting" by using sensational titles or thumbnails to boost consumption, even if the content is false. By implementing such strategies and equipping the average citizen with the ability to spot deepfake content, the potential consequences for victims of deepfake attacks can be mitigated.

Although individuals or organizations may still fall victim to deepfakes, empowering people with the knowledge to detect such manipulations can minimize the impact on the victim. Combatting disinformation necessitates ongoing research on deepfakes and collaborative efforts involving academic institutions, industries, and government agencies to continually evolve countermeasures against advancing deepfake technology.

VI. Importance of Further Research and Attention

As society and technology continues to advance rapidly, concerns arise about whether the appropriate laws and countermeasures are being effectively analyzed and enacted. While Biden's executive order signifies a step in the right direction, further legislative actions and extensive research are still necessary to keep pace with the rapid advancement of AI. Hence, collaborative efforts are required from academia, industry, and government agencies to outpace the evolving technology of deep fakes.

Professor Edward Delp, a distinguished professor of Electrical and Computer Engineering at Purdue University, leads the Semantics Forensics Program developed by the Defense Advanced Research Projects Agency (DARPA) for the U.S. Department of Defense. His research focuses on developing algorithms employing sophisticated techniques based on artificial intelligence and machine learning to detect deep fakes (Huchel, Purdue University). Although commonly associated with facial and audio impersonation, deep fakes extend beyond

these domains, tampering with evidence such as street scenes and biological imagery. The overarching aim of Delp's research is to detect disinformation regardless of its nature (Huchel, Purdue University).

Renowned tech giant Google has also been actively conducting research on effective methods to detect deep fakes. Cade Metz, a technology reporter for The New York Times, highlighted in his article, "For internet companies like Google, finding the tools to spot deep fakes has gained urgency. If someone wants to spread a fake video far and wide, Google's YouTube or Facebook's social media platforms would be great places to do it" (Metz, The New York Times). Hence, it is crucial for platforms like Google, where deep fakes are commonly shared, to take responsibility and collaborate on effective countermeasures against deep fakes.

VII. Conclusion

As more Americans become increasingly aware of the capabilities of deep fakes and with ongoing initiatives from the American government, various industries, universities, companies, and others, the likelihood of a concrete and effective solution against the imminent rise of deep fakes becomes more probable and imminent. This critical conversation should no longer be disregarded or postponed for future consideration. It is imperative to intensify research efforts and broaden awareness across all domains so that more citizens comprehend the evolving landscape of our society. Taking President Joe Biden's initiative as an example, it's essential to acknowledge the current scenario in America and realize that this demands a collective effort for the safety and advancement of all Americans and global citizens.

VIII. Works Cited

AI Voice Generator: Realistic Text to Speech & Voice Cloning (lovo.ai)

BBC News. "Capitol riots timeline: What happened on 6 January 2021?" 2 Aug., 2023. Capitol riots timeline: What happened on 6 January 2021? (bbc.com)

Buermann Gretchen, Perucica Natasa. "How can we combat the worrying rise in the use of deepfakes in cybercrime?" 19 May, 2023. How can we combat the worrying rise in deepfake content? | World Economic Forum (weforum.org)

Deepfake - Wikipedia /// (Wikipedia used this source) Another convincing deepfake app goes viral prompting immediate privacy backlash - The Verge

FaceApp: Face Editor

Farokhmanesh Megan. "The Debate on Deepfake Porn Misses the Point." 1 Mar., 2023. The Debate on Deepfake Porn Misses the Point | WIRED

Galloway, Maggie. "Deepfakes may use new technology, but they're based on an old idea." 4 Mar., 2022. A history of deepfakes, misinformation, and video editing | Popular Science (popsci.com)

GitHub - iperov/DeepFaceLab: DeepFaceLab is the leading software for creating deepfakes.

Huchel Brian. "Purdue professor leads international team's research into deepfakes, manipulated media." 16 Mar., 2021. Purdue professor leads international team's research into deepfakes, manipulated media - Purdue University News

Langa, Jack. "Deepfakes, Real Consequences: Crafting Legislation to Combat Threats Posed by Deepfakes." Boston University Law Review, vol. 101, no. 2, Mar. 2021, pp. 761–801. EBSCOhost, search.ebscohost.com/login.aspx?direct=true&db=lgh&AN=150097538&scope=site.

Metz Cade. “Internet Companies Prepare to Fight the ‘Deepfake’ Future.” 24 Nov., 2019. Internet Companies Prepare to Fight the ‘Deepfake’ Future - The New York Times (nytimes.com)

O’Sullivan Donie. “False video of Joe Biden viewed 1 million times on Twitter.” 2 Nov., 2020. False video of Joe Biden viewed 1 million times on Twitter | CNN Business

Rekhi Dia. “Real or not: How to spot a deepfake.” The Economic Times. 7 Nov., 2023. how to spot a deepfake: ETtech Explainer | Real or not: How to spot a deepfake - The Economic Times (indiatimes.com)

Somers, Meredith. “Deepfakes, explained.” MIT Management Sloan School. 21 July, 2020. Deepfakes, explained | MIT Sloan

The White House. “Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.” 30 Oct., 2023. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence | The White House

The White House. “FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.” 30 Oct., 2023. FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence | The White House