

Social Influences on Deepfake Advancement and Regulation in the Political Sphere

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Sadhika Dhanasekar

Spring 2024

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Pedro A. P. Francisco, Department of Engineering and Society

I. Introduction

In recent years, advancements in artificial intelligence and deep learning techniques have given rise to a new and concerning phenomenon known as Deepfakes. Deepfakes are synthetic media that have been digitally manipulated to replace one person's likeness convincingly with that of another (UVA, 2022). These videos can be used to enhance entertainment, improve video game graphics, and replicate customer support and caller response applications. However, they can also be used maliciously to spread false information, an increasingly common phenomenon witnessed as the technology becomes more accurate and accessible.

The rapid development and accessibility of Deepfake technology has made it easy to create convincing audiovisual content. Often, this content is indistinguishable from original video evidence even with current detecting technologies. The impact of believing this information could be extremely detrimental, particularly with the new infiltration of Deepfakes in the political sphere. Deepfakes have the potential to become a major problem without proper regulations or without the appropriate technology to detect Deepfakes at hand.

Of alarming concern is the use of Deepfakes in the current United States 2024 political election. The 2024 political election is shaping up to be a pivotal year for US citizens as they become increasingly divided in their political parties. Americans have become increasingly unhappy with the state of the economy and overall direction of the country, with Biden having some of the lowest approval ratings in 50 years (Montanaro, 2024). The growing importance of third party candidates and their outreach through advertisements is also important to be considered as voters attempt to move away from the adopted two-party system.

As visual social media, such as news channels and Facebook, become the foremost source of information for the public, it is crucial that voters are fairly informed about voters to make their decisions. There is large potential for the widespread dissemination of Deepfake content via online candidate platforms to reach vast audiences quickly. From falsifying speeches and debates to fabricating compromising statements from political figures, the potential of Deepfakes are limitless. As such, the threat posed by Deepfakes raises questions about the reliability of information in an increasingly digitized world (Roberts, 2023).

For these concerns, this paper will look at the 2024 United States political election and the current and prospective implications of Deepfake technology. We will explore the ways in which Deepfakes distort truth and accountability in the public. By analyzing real-world examples, potential risks and regulatory responses, we gain a better understanding of Deepfake's impact in the political field, the relationship between technology, politics and society, and propose strategies for mitigating the harmful effects of Deepfakes while preserving the integrity of democratic processes and public trust.

II. Background and Significance

There are 65 elections in 54 countries slated for 2024 (McKenzie, 2023), an alarming number of potential locations where Deepfake videos can mistakenly erode public trust and potentially sway an election with false evidence. These issues by Deepfakes are not new by any means. In Argentina's 2023 election, the winning candidate depicted his rivals as Communists and emperors at the last minute, and won by over 3 million votes (Heath, 2024). In another case,

just hours before Slovakia's parliamentary election in September 2023, an audio recording of the leader plotting to rig the election went viral, causing him a narrow loss (Lindsay, 2024).

Looking internally in America, we can already see critical examples of Deepfakes. The losing candidate for the Chicago Mayor elections this past year narrowly lost due to a fake Twitter post displaying a fake audio recording of him talking about funding the police (Devine, 2024). Even with the platform releasing a post exposing the audio as a lie immediately after, the media had already been propagated through the city. In January 2024, some democrats in New Hampshire received convincing AI-generated robocalls using Biden's voice to urge them not to vote in the state's primary (Lindsay, 2024). The rising prevalence and impact of these videos are only limited by the creativity of malicious actors.

A major part of the concern of Deepfakes arises from the numerous actors and large audience involved in perpetuating or succumbing to Deepfakes. Producers of Deepfake videos want monetary gain for creating them, as well as the ability to develop new technologies or more realistic methods in the entertainment field. Users are typically politicians and/or campaign managers that deploy these technologies. Their goal is to portray themselves positively or portray their competitors negatively (Roberts, 2023). They are tightly coupled with the producers to create the most effective advertisements that supplement politicians who deploy these methods.

These artificial videos differ from normal marketing tactics, such as splicing clips of opposing candidates to promote misinformation, but should be treated and regulated at least as strictly. As these past techniques have been in play for many years, voters know the signs of looking for attacking ads. The regulations for these attack ads have many laws enacted, such as company freedom to censor content or candidates signing off videos to reassure content approval

(Oxenford, 2014). However, Deepfakes are advertised as unedited videos and come in a wide variety of forms, such as ads supporting candidates or implicating videos. In addition, detection technologies for Deepfakes are struggling to keep up with rapidly improving Deepfakes, causing concern about debunking faked advertisements (Barria, 2023).

The lack of regulations for identifying Deepfakes and the potential for them to imitate approvals from political candidates make them elusive and difficult, and can cause harm to the public by companies that exploit this and spread misinformation. There have been many requests for laws or policies enforcement. For example, Google has implemented Deepfake identification technologies to limit the fraudulent behavior of accounts due to these lobbyists (Klepper, 2023). However, many of these regulations are limited to only social media companies, with a lack of attention from the government and other federal advertisements (Cohen, 2023). Current laws that limit impersonation, such as slandering a police officer or practicing law without a license, are not appropriate in regulating Deepfakes. This is partly due to Deepfakes being trained on publicly available videos, of which there are huge and uncredited databases online (Barrier, 2023). Regulations also must be correctly implemented as there are arguments regarding the infringement of free speech in limiting the creation and posting of Deepfake videos (O'Sullivan, 2023).

Lastly, as mentioned before, the public are innocent bystanders who are exposed to these videos. They can play a large role in spreading misinformation and may not exercise the proper validation measures to ensure that their information is reliable (Cohen, 2023). Thus, there is a legitimate concern that these videos can cause the public to lose trust in politicians and the news

system, causing a decrease in voter turnout. This is prevalent among older generations, who make up a larger portion of voters and display a greater lack of internet literacy (Roberts, 2023).

Of even more concern, there is tremendous untapped potential in Deepfake technologies, like videos created from scratch or in live time. As videos inherently feel objectively true and hard to forge, it can be hard to convince the public of what is true or not.

III. Methodology

This paper will conduct a literature review to determine the effect Deepfakes have currently had on previous elections or similar spheres of influence, proposed regulations, and predicted effects on future elections/events. It will also utilize the Social construction of technology (SCOT) theory as a framework, which argues that technology is shaped by human actions and is a response to societal forces. In this framework, Deepfake technologies are created for politicians to spread false narratives about their competitor, creating a new format of attack ads that leave the public unsure of what they can believe. In considering the problem of Deepfake and analyzing its impact, analysis will be done through the key components of SCOT: relevant social groups, interpretative flexibility, problems & conflicts, closure & stabilization, and wider context (Humphreys, 2005).

SCOT states a relevant social group as when “all members of a certain social group share the same set of meanings, attached to a specific artifact” (Pinch & Bijker, 1984). Four groupings occur: producers, users, lobbyists, and bystanders. These were explained in the previous section on the Background and motivations of Deepfake.

Interpretive flexibility is defined by SCOT as the idea “that technology design is an open process that can produce different outcomes depending on the social circumstances of development.” (Klein & Kleinman, 2002). Deepfake has many positive uses in the entertainment sector, such as creating more realistic CGI effects. However, untagged Deepfake videos raise concerns about feeding into public mistrust. In this case, this misinformation and its impact on how people vote or perceive the candidates can cause the technology to develop even further if results favor the candidate who used this Deepfake video.

Closure is when relevant social groups consider their problem with the artifact to be solved. Stabilization is the process by which members of a relevant social group come to communicate definitions and specifications of an artifact similarly over time (Humphreys, 2005). With methods of Deepfake continuously improving, it is hard to see closure or stabilization of this technology happening in the near future.

The last tenant of SCOT addresses wider context. Aspects of wider context were touched upon in the Background section, but involved the future of trust in media and the American government due to increasing corruption in the political scene.

IV. Literature Review

There have been numerous studies to analyze the impacts of Deepfakes on the public’s perception of the political sphere and other overlapping areas. While it is near impossible to replicate the true environment that election season brings to the United States, the rising concern of Deepfakes from the 2020 presidential election as well as in numerous primaries has amplified the research efforts behind it.

One of the first aspects that have been tested is how well the public is able to identify Deepfakes and the impact of misidentification. A 2020 study found that when shown a Deepfake of Obama, approximately 15% of viewers thought it was real (Ray). On the other hand, 50% of viewers were not deceived. The remaining 35% of viewers were not able to come to a conclusion on determining if the video was real or not. Another study by Ahmed (2021) calls attention to the sharing aspect of Deepfakes. Politically-interested or low cognitive users are more likely to inadvertently share Deepfakes, continuing to spread misinformation. The concern here is that the video now appears to be coming from a reliable peer network source, and is more likely to be perceived as true. Furthermore, the more connected a user may be leads to a greater network and sharing size, highlighting the importance of recognizing Deepfakes early.

The accuracy of detecting a political Deepfake also greatly increases with the more political knowledge a person has. Appel and Prietzel (2022) found that the higher a participant scored on analytical thinking and political interest, the better the participant identified Deepfakes, fake news articles or other forms of misinformation used by candidates against their opponents. Another study warned half of its participants about the dangers of Deepfakes before playing either a real or a fake video of the politician making an atypical policy statement for their party (Ternovski, Kalla and Aronow, 2021). The pattern from the previous study stayed true, and warnings about the existence of deepfakes did not enhance the participants' ability to spot manipulated video content. In fact, participants who were unfamiliar with the party's platform were equally likely to think the videos shown were Deepfakes and artificially produced. This study took advantage of many Americans' low levels of policy knowledge, and brought to light increasing concerns about the political literacy of American democracy. Above all else, broad

literacy in politics and digital technology most strongly increases discernment between deepfakes and authentic videos of political elites (Barari et al, 2021).

Other studies call attention to the real issue of Deepfake being the speed at which false narratives and media can be constructed. For example, one study argues that Deepfakes are on the rise and occurring at alarming rates, predicting that more than 50% of future media will consist of Deepfakes in the year 2025 (Barari et al, 2021). This is only after 2 years of Deepfakes being an official form of technology. However, the study goes on to say that this amount of information is no more greater than the misinformation that is conveyed through existing news formats such as textual headlines or audio recordings. Thus, Deepfakes shouldn't be treated as different or more impactful than other forms of media as long as we continue to teach political and technological literacy. Another study states that Deepfakes should be thought of more as an "evolution than a revolution in disinformation techniques" as their threat emerges from the scale and ease to make an organized and dynamic influence (Whyte, 2019). Whyte urges we increase transparency in these forms of media, and work to create regulations early. Lastly, another study highlights the importance of visual and audible media that Deepfakes targets. Vacarri and Chadwick (2020) show that viewers are more likely to accurately recall visual messages, and that these "image bites" are more powerful in shaping voters' opinions and factual recall. Increasingly common Deepfakes can result in more pieces of media the public is exposed to in which they can be confused or incorrectly recall information, even if the same amount of information is available to them in written form.

There are still many advocates for Deepfakes. Some studies argue that, although Deepfakes may play an impact on choosing a candidate at crucial times, the overall political

affiliation and beliefs of most people stays the same. Dobber et. al (2020) constructed a political visual and audible Deepfake and found that, while attitudes to the politician in the video were significantly lower after seeing the Deepfake, the political attitude towards the politician's party remained similar to the control condition. The same study also demonstrated the effectiveness of microtargeting techniques to effectively target and sway critical voter groups. Lastly, Hameleers and Marquart (2023) conducted an experiment where they mislabeled accusations in a video. This exposure of misinformation lowered the perceived credibility of the video, but did not affect the policy preferences related to the content of the video. The effects of the accusations lasted over a period of three days before participants went back to a relatively similar state of policy preferences. Advocates of Deepfake preach the technologies' benefits to communicate candidate information fairly to parties.

Pro-Deepfake enthusiasts believe that Deepfakes can be used to level the playing field and learn to use AI for ad generation and social media content, helping older politicians stay more relevant in younger circles (Ulmer and Tong, 2023). The cost and ease of creating Deepfakes have decreased dramatically in just the past year, making it accessible to any who want to try their hand at politics. For example, cloning a voice used to cost \$10,000 in server and AI-training costs in 2022 and now startups have begun offering it for a few dollars (Kobis et al, 2023). People against stricter regulations for Deepfakes also bring up the numerous positive applications of them. Numinar Analytics, a political data company, has been experimenting with AI content generation to create personalized messaging in a candidate's voice or communicate through language barriers (Ulmer and Tong, 2023). Democratic polling and strategy group,

Honan Strategy Group, is trying to develop an AI survey bot in a female voice to be more inclusive and welcoming.

V. Proposed Regulations

The regulation of Deepfakes is a sensitive topic that suffers from the deployment and identification of the technology being vaguely defined. Even if the Deepfakes themselves are not convincing or malicious, warnings about them can easily be weaponized by politicians and campaigns to dismiss real videos. This can raise questions about the legitimacy of democratic institutions, the power of the citizens, and the power of malicious political actors (Bradshaw and Howard, 2018). Transparency with Deepfakes has led to people feeling uncertain if they are being misled which in turn reduces trust in news on social media. One study states that Deepfakes may contribute toward “generalized indeterminacy and cynicism”, and enlarge challenges in online civic culture (Vaccari and Chadwick, 2020). Another study highlights the biases this reveals in the American democratic system already. Participants viewing a Republican party video were more likely to think it was fake (Schiff et al, 2021). They were also more likely to react with stronger emotions toward Republican party videos. Ultimately, politicians can take advantage of the confusing playing field. Writing about Russian operations, Pomerantsev (2015) notes, “The aim is . . . to trash the information space so the audience gives up looking for any truth amid the chaos.” The plethora of contradictory, nonsense and otherwise disorienting messages that politicians introduce results in a system of uncertainty that we are expecting the public to navigate unknowingly.

This, and the numerous examples of devastating and well-timed uses of fake news, has led to countries developing stricter regulations. The European Union recently developed their Artificial Intelligence Act with safeguards on general purpose AI limiting the use of biometric identification systems, AI scoring and manipulation of user vulnerabilities, and the right of consumers to launch complaints (EU, 2024).

With all of this evidence of incriminating Deepfakes, it's important to observe what the response from the United States has been. Some of the states in America have deployed regulations about the use of artificially generated images during the election season. Minnesota lawmakers enacted a law that bans the publication of "Deepfake media to influence an election" if the media is "made with the intent to injure a candidate or influence the result of an election" or "without the consent of the depicted individual" (Edelmen, 2023). However, the anonymous nature of the internet has made it difficult to find and attest to perpetrators of the crime. Scant evidence of creators of major Deepfakes this election season, as well as a difficulty translating these regulations over state borders, have resulted in scant evidence about the effectiveness of these laws (Swanson, 2024). Arizona's Secretary of State, Adrian Fontes, implemented focus testing on various state stakeholders by creating fake videos of a politician before polling them for preliminary voting results (Sakellariadis, 2023). Fontes found that AI doesn't pose an entirely new threat to the vote as misinformation continuously runs rampant, but strongly encourages the implementation of stricter regulations.

The Federal Election Commission has been slow on their response to regulation of AI-generated Deepfakes in political ads, particularly for the rising concern in the current Primary Election. Arguments over the infringement of the right of speech under the First Amendment

have restricted the potential for regulation. While many of these regulations can be safely used against foreign actors such as other countries or companies, condemning or charging a US citizen may create more cause for concern. There is worry that accusing citizens of releasing a Deepfake could give the impression that the US is trying to influence the election or restrict speech (Devine et al, 2024). Even if the FEC decides to ban AI Deepfakes in campaign ads altogether, they are not able to control outside parties, individual social media users, or Political Action Committees (PACs) from releasing Deepfakes imitating candidates or other concerns (Swenson, 2023). Biden has released statements encouraging watermarks for authenticity and verification on campaign videos (Edelmen, 2023).

Many AI companies have restricted the use of their softwares. OpenAI prohibits its image generator, DALL-E, from creating public figures, with the error message “it may not follow our content policy” (Ulmer and Tong, 2023). OpenAI also restricts any scaled usage of its products, especially for political purposes, disallowing things like sending out mass personalized emails or other aspects. However, there are loopholes and a lack of strong enforcement around these policies such as generating images of lesser known public figures or using third party apps to scale services. META and Google require A leveling policies for their political ads on YouTube, Facebook, Instagram and other popular forms of social media (Klepper, 2023). Any content violating the rule will be forcefully removed.

Other potential AI regulations are actively being worked upon in hopes for future elections beyond the 2024 Presidential election. House Speaker Mike Johnson of Louisiana and House Minority Leader Hakeem Jeffries of New York recently announced a joint task force consisting of a dozen Democrats and a dozen Republicans to work together to regulate the use of

artificial intelligence, especially in politics (Lovato, 2024). Further regulations include drafted bills to regulate use from the developer side, requiring impact assessments on focus groups or a public body, and bills that even forbid the unapproved creation of Deepfakes.

Lack of these regulations could result in large changes to the current democratic process. We may see changes in society encouraging more live events, verification and validation of videos, and marketing strategies as we become more situated with Deepfakes in society. We may also see mass hysteria among the public at being frequently lied to and a rising lack of trust in the government for looking out for citizen well being. In analyzing these issues, there will hopefully be how Deepfake regulations can be properly applied and how Deepfake can be implemented into society fairly.

VI. Conclusion

This paper has demonstrated the rapid proliferation of Deepfake technologies and the significant challenge it poses to the mass public, democratic process and governmental trust. The threat lies in not only the easily generated audiovisual deceptions, but also in the quick spread of false information online and taking advantage of public naivety.

Efforts to combat Deepfakes could lie in initiatives to increase political and media literacy in the public. Another approach is in encouraging transparency and responsibility in political candidates to help combat the spread of misinformation and manipulation of these political figures. Moreover, honesty in media outlets can also help create a more trusted community during political elections.

In light of this, lobbyists and society members must continue to develop regulations to address this threat and safeguard media communication for the general public. These regulations may include restrictions on social media platforms, more robust content moderation, and legal frameworks to hold Deepfake creators accountable. While such measures may be useful, the regulations must maintain a fine line between the protection of free speech and security.

Ultimately, Deepfakes have, and will continue, to play a huge role in political elections from a variety of factors. Thus, it is crucial to develop the appropriate regulations and safeguards now to ensure our democratic system remains just, and to preach political and technological literacy to the public so that there is still trust between our government and the people.

References

Arizona creates own deep-fake election hoaxes to prepare for 2024 - politico. (n.d.-a).

<https://www.politico.com/news/2023/12/22/arizona-ai-election-deep-fakes-00133113>

Artificial Intelligence Act: Meps adopt landmark law: News: European parliament. Artificial

Intelligence Act: MEPs adopt landmark law | News | European Parliament. (n.d.).

<https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>

Cable News Network. (n.d.). *Inside the Pentagon's race against Deepfake Videos.* CNN.

<https://www.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/>

Cohen, B. (2023, October 13). *How the rise of deepfakes will impact the 2024 presidential*

elections. Infosecurity Magazine.

<https://www.infosecurity-magazine.com/opinions/deepfakes-impact-presidential/#:~:text=With%20deepfake%20technology%20becoming%20more,the%20opposing%20candidate%20look%20bad>

Council on Foreign Relations. (n.d.). *Campaign roundup: The Deepfake threat to the 2024*

election. Council on Foreign Relations.

<https://www.cfr.org/blog/campaign-roundup-deepfake-threat-2024-election>

Deepfakes & Elections. Truepic. (2024, April 2). <https://truepic.com/deepfakes-elections-2024/>

Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and Trust in News - Cristian Vaccari, Andrew Chadwick, 2020. (n.d.-b).
<https://journals.sagepub.com/doi/full/10.1177/2056305120903408>

Deepfaking it: America's 2024 election collides with ai boom | reuters. (n.d.-c).

<https://www.reuters.com/world/us/deepfaking-it-americas-2024-election-collides-with-ai-boom-2023-05-30/>

Devine, C., O'Sullivan, D., & Lyngaas, S. (2024, February 1). *A fake recording of a candidate saying he'd rigged the election went viral. experts say it's only the beginning* | CNN politics. CNN.

<https://www.cnn.com/2024/02/01/politics/election-deepfake-threats-invs/index.html>

Do (microtargeted) deepfakes have real effects on political attitudes? - Tom Dobber, Nadia

Metoui, Damian Trilling, Natali Helberger, Claes de Vreese, 2020. (n.d.-d).

<https://journals.sagepub.com/doi/full/10.1177/1940161220944364>

Klepper, D., & Press, T. A. (2023, November 8). *Meta updates AI deepfake policy for political ads, 2 months after google did it*. Fortune.

<https://fortune.com/2023/11/08/meta-label-ai-generated-deepfake-political-ads-2024-election/>

Köbis, N. C., Doležalová, B., & Soraperra, I. (2021, October 29). *Foiled twice: People cannot detect deepfakes but think they can*. iScience.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8602050/>

Lovato, N. (2024a, February 19). *Could ai sway the 2024 election? here's what's being done nationally to stop deepfakes*. USA Today.

<https://www.usatoday.com/story/news/politics/elections/2024/02/19/ai-deepfakes-2024-election/72486000007/>

Montanaro, D. (2024, January 13). *What is the 2024 election all about, anyway?*. NPR.

<https://www.npr.org/2024/01/13/1224412456/2024-election-american-values>

Mshaw. (2023, October 11). *Drawing the line on AI-based Deepfakes proves tricky for Congress*.

Roll Call.

<https://rollcall.com/2023/10/11/drawing-the-line-on-ai-based-deepfakes-proves-tricky-for-congress/>

NBCUniversal News Group. (n.d.). *States are lagging in tackling political deepfakes, leaving potential threats unchecked heading into 2024*. NBCNews.com.

<https://www.nbcnews.com/politics/artificial-intelligence-deepfakes-2024-election-states-r-cna129525>

Oxenford, D. (2014, September 23). *Questions about the truth of political ads, what's a*

broadcaster to do when a candidate complains about an attack ad? – the no censorship rule for candidate ads. Broadcast Law Blog.

<https://www.broadcastlawblog.com/2014/09/articles/questions-about-the-truth-of-political-ads-whats-a-broadcaster-to-do-when-a-candidate-complains-about-an-attack-ad-the-no-censorship-rule-for-candidate-ads/>

O'Brien, M. (2023, October 5). *U.S. lawmakers question Meta and X over AI-generated political deepfakes ahead of 2024 election.* PBS.

<https://www.pbs.org/newshour/politics/u-s-lawmakers-question-meta-and-x-over-ai-generated-political-deepfakes-ahead-of-2024-election>

Swanson, K. (2024, March 8). *Minnesota banned the use of deepfakes in elections. now dozens of other states are, too.* KSTP.com 5 Eyewitness News.

<https://kstp.com/5-investigates/minnesota-banned-the-use-of-deepfakes-in-elections-now-dozens-of-other-states-are-too/>

Swenson, A. (2023a, August 10). *FEC moves toward potentially regulating AI deepfakes in campaign ads.* PBS.

<https://www.pbs.org/newshour/politics/fec-moves-toward-potentially-regulating-ai-deepfakes-in-campaign-ads>

Temir, E. (2020, July 1). *Deepfake: New Era in the age of Disinformation & end of reliable journalism.* Selçuk İletişim. <https://dergipark.org.tr/en/pub/josc/issue/55089/685338>

Washington State's plan to combat election deepfakes and AI ADS - Axios Seattle. (n.d.-e).

<https://www.axios.com/local/seattle/2024/01/24/washington-election-deepfakes-ai-plan>

What the heck is a deepfake?. What the heck is a deepfake? | UVA Information Security. (2022,

November). <https://security.virginia.edu/deepfakes>

Williams, Z. (2023, December 22). *More states to push laws banning AI election deepfakes in*

2024. Bloomberg Law.

<https://news.bloomberglaw.com/artificial-intelligence/more-states-to-push-laws-banning-ai-election-deepfakes-in-2024>

X, S. (2023, August 2). *Study shows speech deepfakes frequently fool people, even after*

training on how to detect them. Phys.org.

<https://phys.org/news/2023-08-speech-deepfakes-frequently-people.html#:~:text=In%20a%20study%20involving%20more,detect%20deepfakes%20had%20minimal%20effects>