Michael Ortega

CS3001, Spring 2023

Term Paper

**Ethics of Deepfake Technology in Entertainment Media**

*Introduction*

It is clear that since the invention of Computer-generated images (CGI) in the 1960s, technology in film and television has been on an upward trend of expanding entertainment beyond what can physically be captured on camera. The rapid evolution of software and hardware has led filmmakers to seamlessly expand their creativity, leaving us as the viewer unsure of how to assess the credibility of what we see on the screen. In the same manner, deepfake technology and its recent yet rapid conception has exacerbated this dilemma; In this paper I will specifically assess the ethics of using deepfake technology in entertainment media, which is still in its early stages, but has heavy and discernable implications in the present to near future.

For the purposes of clarity in this analysis, I will strictly use the term "entertainment media" to refer to widely viewed television programming and movies as opposed to online content such as YouTube videos or other social media-based forms of entertainment. Deepfakes themselves come in five different forms (Masood et al. 2022), but in general refer to audio or video manipulation that "depict people saying and doing things that never actually happened" (Westerlund 2019). This paper still addresses all five forms of deepfakes, including face swaps (forged faces on different bodies), lip-synching (forged lip movement), puppet mastering (forged body motion), face synthesis, and audio-only deepfakes.

Deepfake media has significantly improved since its inception in 1997 with three short videos of former president John F. Kennedy saying statements like "Read my lips" and "I never met Forrest Gump", two things the former president never really said, especially on video (Bregler et al. 1997). Current deepfake technology allows regular people to generate deepfake videos online for as little as four dollars an hour (Deepfakes web n.d.), as well as free on some other sites and apps for shorter clips (Wombo n.d.). Some of these programs boast ethicality by claiming to only produce ethical deepfakes (Synthesia n.d.). Though the free and cheaper deepfake programs are generally worse quality than the more expensive ones, the technology has become widespread, allowing people to use deepfakes for essentially anything they please.

Research has shown that the general public has not yet gotten enough exposure to deepfakes to be able to detect a deepfake from a real video (Korshunov & Marcel 2020). Hence, this paper seeks to address how ethical deepfake use in movies and film is when the majority of viewers won't even know what they are seeing is a deepfake.

*The History of Modern Deepfake Technology*

Though Braegler et al.'s study showed us that it is possible to alter prerecorded videos to lip sync faces to whatever we want them to appear to say, deepfake technology first became especially powerful with the development of Generative Adversarial Networks (GANs) in 2014 (Goodfellow et al. 2014). GANs work by using a 'generative' machine learning model that generates new data in tandem with a discriminative model that checks the accuracy of the generative model by comparing to its training dataset, which allows the models to build upon and improve itself (Goodfellow et al. 2014). This technology paved the path for the modern generation of deepfakes.

Deepfakes first received major media attention when Jordan Peele and Buzzfeed released a PSA video warning people online of the dangers of deepfakes by using a deepfake video. In their PSA, they depicted Obama saying statements like "Killmonger was right" and "Ben Carson is in the sunken place", both statements the former president has never said on camera. Jordan Peele is later revealed in the video to be speaking those statements, showing the viewer that this is, in fact, a deepfake video (BuzzFeed 2018). The video came about around a time when deepfakes started sporadically appearing online, with reddit user "deepfakes" creating and spreading deepfakes on their eponymous subreddit, utilizing the new FakeApp app that allowed users to face swap people onto videos (Cole 2018). Following this, media outlets became afraid of the possibilities of deepfakes infiltrating social media before the 2020 presidential elections, causing people to be confused by fake videos circulating online (Mak & Temple-Raston 2020). Although this paper is about deepfakes in entertainment media, it's important to note that this widespread fear in media signifies the sheer impact deepfakes can have, especially when used for malicious purposes.

In June 2021, deepfakes made their way into mainstream entertainment when the director of *Roadrunner*, a posthumous documentary about Anthony Bourdain's life, used a portion of deepfaked audio of Bourdain's voice (Martin 2021). Though only using about forty-five seconds worth of audio from the deceased Bourdain, Morgan Neville, the director received a myriad of negative reviews about his use of the deepfake (Rosner 2021).

The month after, news spread about YouTuber "Shamook" joining Lucasfilm's visual effects team to work on deepfakes of Luke Skywalker in *The Mandalorian*. Shamook is a YouTuber that, up until that point, had consistently posted YouTube videos of famous actors deepfaked onto scenes they were never in. In particular, their deepfake Luke Skywalker in *The*

*Mandalorian* led Lucasfilm to offer them a position on their VFX team (Sharf 2021). Although this paper is not concerned with online entertainment like YouTube, this example shows that deepfakes are here to stay in the film and television industry.

Later that year, Bruce Willis appeared in a Russian advertisement for Megafon (Reuters 2021). Although Willis declines the claims that he sold the rights to his face to the deepfake company Deepcake to develop this advertisement (Derico & Clayton 2022), the video is clearly deepfaked.

### Arguments Supporting the Use of Deepfakes in Entertainment Media

When an actor dies, suddenly the only new media that can be published using their face, body, and voice would be scenes or dialogue that have already been recorded in their lifetime. With deepfake technology, production crews can feed a deepfake model these prerecorded scenes and audio clips to generate deepfaked video and audio of the deceased actor doing and saying whatever they choose. Such an example of this is Anthony Bourdain's voice being deepfaked in *Roadrunner*.

As aforementioned, critics gave mixed reviews on Bourdain's short deepfake in the film, but director Morgan Neville explained his use of the technology by saying "I wasn't putting words into his mouth. I was just trying to make them come alive" (Martin 2021). Neville noted that the deepfaked lines were those that Bourdain wrote himself, just never spoke or recorded (Rosner 2021). Therefore, this technology opens the door for recreating and bringing to life words or ideas from deceased actors or figures before they could manifest them themselves.

In a sense, this means that dead actors can be brought back to life through deepfakes. Filmmakers have traditionally used CGI to make this happen, such as in *Rogue One: A Star*

*Wars Story* when the VFX team brought a young version of Carrie Fisher back into the film through CGI. However, the VFX team explained that this is particularly difficult and expensive to do with the necessity of having a real actor recreate the scene and using postproduction to replace the actor's face and voice (ABC News 2017). Critics noted that Fisher's appearance was buggy and looked poor overall (Langmann 2020), but YouTuber Shamook used deepfake technology to enhance her appearance, noting that the process only took 24 hours, an $800 PC, and 500 images of Carrie Fisher to produce (Shamook 2020); the transformation was not perfect; notwithstanding, the minimal resources suggest that deepfakes may be cheaper for VFX teams than traditional CGI methods. Bringing dead actors back to life through deepfakes in posthumous entertainment brings younger generations new exposure to actors, which in a way can prolong their legacy and satisfy fans (Dimitropoulos 2020).

In *Welcome to Chechnya*, filmmaker David France sought to document the anti-gay protests in Chechnya through gay and lesbian Chechens. To protect the identity of 23 subjects in the documentary, the filmmaker relied on deepfake technology to swap out these faces (Rothkopf 2020). Face swapping endangered or hunted individuals is similar to using voice modifiers and blurry images to protect the same individuals, but the use of deepfake provides for stronger story telling capabilities. In addition to this, France provided a sample solution to the question of "how can audiences know what they see is real or not?" by using a star towards the subjects' faces to denote that the audience is looking at a synthetic face (Rothkopf 2020).

Extrapolating from the fact that directors can use deepfakes creatively to expand their possibilities and that deepfakes have already been used in pornography (Patrini et al. 2019), it's possible that deepfakes can be used positively in consensual synthetic pornographic situations. For instance, a director may feel that a certain scene is necessary for the development of a story

or is important to showcase for the audience, but the actors involved may not always be comfortable with the idea. Through deepfakes, directors and actors may be able to find this compromise where an actor may not have to commit to a certain scene but rather allow the director to use their face instead, allowing both parties to be satisfied and comfortable in an ideal situation. This, of course, can extend beyond the realm of pornography and may be the case in any other generic scenario that an actor may not be comfortable with.

### *Arguments Against the Use of Deepfakes in Entertainment Media*

Clearly, the ability to use deepfakes puts a lot of power into the hands of directors. Many of the arguments to be made in support of deepfakes in film and television ride along the hope that their directors use it ethically and in good intention. Even working in a team, production crews may not think about some of the negative implications that using deepfakes may have.

For instance, one of the largest arguments by critics of *Roadrunner* was the fact that Anthony Bourdain himself always preached "authenticity", and deepfaking his voice effectively went against his own beliefs (Rosner 2021). This in itself raises a major problem of posthumous production: dead actors cannot provide consent. Traditionally, filmmakers have utilized old footage of actors when they were alive to posthumously release their work, as it works with every other form of entertainment up until this point. The discussion of posthumously released media goes beyond the scope of this paper, but the main idea is that deepfakes raise a further question of ethicality when considering that producers no longer need this pre-recorded footage (with the exception of training date) to bring dead actors back to life. The prolongation of deceased actors is also heavily shaped by how directors use and manipulate their work. Bourdain's legacy and career was shaped by statements against "artifice, phoniness,

inauthenticity, misrepresentation, [and] simulation", so in some ways showcase posthumous harm in his legacy and going against his wishes (Nash 2022). We may ask, "would Anthony Bourdain be okay with this?" but could never get a definite answer.

Additionally, like in *Welcome to Chechnya*, deepfake technology furthers the capabilities of producing synthetic faces and therefore so-called "digital actors". Along with the development of A.I. generated voices, the combination of visual and auditory deepfakes allows for the creation of a digital actor, an actor that filmmakers can use (and abuse) in whatever way they please. Just as Bruce Willis's face was used in a Russian advertisement, actors may have their face or voice licensed out, which may potentially be cheaper for filmmakers than casting the actor themselves. Filmmakers are able to synthesize new actors out a handful of different faces, potentially taking jobs from real actors (Dimitropoulos 2020). Filmmakers may even prefer using digital actors than real ones. As deepfakes may help directors and actors come to a compromise when actors are uncomfortable with performing specific scenes, some directors may find themselves better off working with digital actors who will do and say what the director wants.

As the technology is not widespread enough for us to understand its impact on the entertainment industry yet, we don't know if the creation of jobs resulting from deepfake technology will counterbalance the jobs taken away from actors. Additionally, we don't have any framework or code of conduct regarding deepfakes in entertainment and how licensing of faces and voices would work. A further argument may be made about the shift in expression through film if the cinema workforce shifts due to the prevalence of deepfakes.

*Current Legislation of Deepfake Technology in the U.S.*

As new as the technology is, it's apparent that many countries (including the United States) have not yet adopted a formal policy regarding the usage of deepfakes. Within the U.S., deepfakes may become illegal for the following reasons (Ban 2022):

1. Infringement on intellectual property rights

2. Violation of personal rights

3. Damage to personal reputation (Slander and libel)

4. Compromises data protection and privacy

5. Liability concerns

While no laws have passed in the U.S. regarding the use of deepfakes in entertainment or even solely online, the U.S. government has taken steps to monitor and report the development of deepfakes. For instance, the National Defense Authorization Act is required to submit an annual report on the state of deepfakes in the U.S. (116th Congress 2019). The same congress passed a year later the Identifying Outputs of Generative Adversarial Networks Act, which makes the National Science Foundation responsible for supporting research on the consequences of deepfakes and how to combat them (116th Congress 2020). That same year, Congress passed the Deepfake Report Act of 2019 requiring the Science and Technology Directorate in the Department of Homeland Security to report at specific intervals on the state of digital content forgery technology (116th Congress 2021).

Beyond the U.S., a few countries have taken their stance on deepfakes. In the U.K., the Ministry of Justice has illegalized the sharing of non-consensual deepfake porn (UK Ministry of Justice 2022). In Japan, a news journal reported that two men were arrested for running a porn site with deepfaked porn videos (The Asahi Shimbun 2020).

*The Morality of Deepfaking Actors*

Answering the question of "should we deepfake actors in entertainment media?" is particularly difficult to answer based on where research is and the examples that we have of deepfakes in entertainment. Addressing the advantages, disadvantages, and consequences of using deepfakes leads to a Utilitarianist approach of answering the question. Surely, it could be impossible to quantify the multitude of impacts deepfakes have on all affected parties, beyond just the filmmaker, actor, and audience. Rather, we should ask the question "is this ethically permissible?"

Depicting people doing things they have never done is, in my opinion, a form of deceit. However, in the sphere of entertainment, much work is already fictionalized, and viewers should generally understand that scenes in works of fiction are made up and performed by actors for the purpose of entertainment. By the ethical frameworks of Kantianism and virtue ethics, deepfakes are certainly unethical based on the principle of deceit; however, the context of deepfakes largely influences this morality. For example, a sci-fi movie that depicts deepfaked actors preforming actions real people could never do is clearly using deepfakes to expand creative ideas and manifest them into entertainment. I would argue that the audience likely knows that they're watching a body of fiction, so deepfakes would not be used in a deceitful manner and would be ethical. However, in a body of nonfiction, the lines of deceit become more blurred. As in *Roadrunner* and *Welcome to Chechnya*, directors of documentaries may utilize deepfakes of actors or subjects in a way that makes it more difficult and even impossible for viewers to know if the scene they're watching really took place or if it was synthesized by the producer, especially given that the work is nonfiction. The difference between the ethics of deepfakes in *Roadrunner* and *Welcome to Chechnya* is that France gives a cue to the audience so that they understand the

23 faces they see are purely synthesized, whereas Neville didn't disclose the use of deepfakes until post release. In my opinion, it's more ethical to let the viewer know in some manner that deepfakes are being used in a body of work where everything is assumed to be real.

Besides deceit, how should the families of deceased actors feel when they see their relative depicted saying or doing something they had never really done? Although some actors may have written in their will or expressed to their close ones that they would be okay with posthumous releases of their work, it's very unlikely that any actor has thought yet to write or express that they would be okay with their face and body being used posthumously for deepfakes. Thus, the conversation falls under the idea of posthumous consent and the lack thereof. It would certainly be more ethical knowing that the actor has clearly expressed their approval of the technology, but with the newness of deepfakes, we can only ask if we think they would be okay with it. Surely, the people closest to them would know the answer and should have some say in what their body and voice should and can be used for, but this may not always be the case with actors with little or no family and loved ones left. As mentioned before, the ethics of posthumous production is beyond the scope of this paper, but I believe that the best way to preserve an actor's, or any figure's, legacy is through their published work and what they've done in their lifetime, not what others have produced using their face or body.

### *Closing Remarks*

Deepfake technology has only been around for roughly five years, yet it has improved at such a rapid pace that has led to a myriad of critiques and fear in the U.S., particularly in the entertainment industry. In researching deepfakes, it's important to note that much of the work that has been done to study this topic is largely based on developing technologies to combat and

detect deepfakes. There is a multitude of factors that plays a role in deciding whether it's okay to use deepfakes in entertainment media, and I think as a filmmaker I would want to understand the context in which it would be used before making the decision whether to use it. For instance, I feel that France's use deepfakes in *Welcome to Chechnya* is a prime example of its benefits and how it can be helpful to drive a filmmaker's storytelling abilities whilst avoiding harm. However, in *Roadrunner*, I feel that Neville's use of the technology, albeit brief, goes against Bourdain's ideologies, and the documentary could have progressed without the need for those roughly 50 seconds of dialogue.

As a society, we have much to learn about the power of deepfakes and should look to establish a code or handbook to help guide our filmmakers and actors implement deepfakes into their work. I have thus discussed the ethical controversy of using deepfakes in entertainment media and noted that the future of this technology is unknown but has room to heavily impact this industry.

## References

*3 arrested for defamation over face-swapping porn sites: The Asahi Shimbun: Breaking News, Japan news and analysis*. The Asahi Shimbun. (2020, November 19). Retrieved March 23, 2023, from https://www.asahi.com/ajw/articles/13944568

ABC News. (2017, January 5). How rogue one created princess Leia, Grand Moff tarkin. YouTube. Retrieved April 10, 2023, from https://www.youtube.com/watch?v=xMB2sLwz0Do

Ajder, H., Patrini, G., Cavalli, F., &amp; Cullen, L. (2019, September). The State of Deepfakes. Retrieved April 10, 2023, from https://regmedia.co.uk/2019/10/08/deepfake_report.pdf?module=inline&amp;pgtype=article

Ban, T. (2022, October 26). *Are deepfakes illegal? here are the implications*. Q5iD. Retrieved March 23, 2023, from https://q5id.com/blog/are-deepfakes-illegal-here-are-the-implications#:~:text=Though%20the%20concept%20of%20a,at%20individual%20and%20societal%20scales.&text=Intellectual%20property%20rights%20are%20given,%2C%20films%2C%20and%20computer%20programs.

Bregler, C., Covell, M., & Slaney, M. (1997, August). Video rewrite: Driving visual speech with audio. In Proceedings of the 24th annual conference on Computer graphics and interactive techniques (pp. 353-360).

BuzzFeed. (2018). *You Won't Believe What Obama Says In This Video!* 😉. *YouTube*. Retrieved

    March 23, 2023, from https://youtu.be/cQ54GDm1eL0.

Cole, S. (2018, January 24). We are truly fucked: Everyone is making ai-generated fake porn

    now. VICE. Retrieved April 10, 2023, from https://www.vice.com/en/article/bjye8a/reddit-

    fake-porn-app-daisy-ridley

Derico, B., & Clayton, J. (2022, October 2). *Bruce Willis denies selling rights to his face*. BBC

    News. Retrieved March 23, 2023, from https://www.bbc.com/news/technology-63106024

Dimitropoulos, S. (2022, June 3). *A new film with a dead star? thanks to CGI and deepfake*

    *technology, it's possible*. Experience Magazine. Retrieved March 23, 2023, from

    https://expmag.com/2020/10/a-new-film-with-a-dead-star-thanks-to-cgi-and-deepfake-

    technology-its-possible/

Ethical deepfake maker. Synthesia. (n.d.). Retrieved April 10, 2023, from

    https://www.synthesia.io/tools/deepfake-video-maker

Gamage, D., Ghasiya, P., Bonagiri, V., Whiting, M. E., & Sasahara, K. (2022). Are deepfakes

    concerning? analyzing conversations of deepfakes on reddit and exploring societal

    implications. *CHI Conference on Human Factors in Computing Systems*.

    https://doi.org/10.1145/3491102.3517446

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville,

    A., & Bengio, Y. (2014, June 10). *Generative Adversarial Networks*. arXiv.org. Retrieved

    March 23, 2023, from https://doi.org/10.48550/arXiv.1406.2661.

Jaiman, A. (2020, August 27). *Debating the ethics of deepfakes*. ORF. Retrieved March 23,

    2023, from https://www.orfonline.org/expert-speak/debating-the-ethics-of-deepfakes/.

Korshunov, P., & Marcel, S. (2020, September 7). *Deepfake detection: Humans vs. machines*.

    arXiv.org. Retrieved March 23, 2023, from https://doi.org/10.48550/arXiv.2009.03155

Langmann, B. (2021, November 29). How J.J. Abrams pulled off Carrie Fisher's CGI flashback

    in 'Star wars: The rise of skywalker'. Esquire. Retrieved April 10, 2023, from

    https://www.esquire.com/entertainment/movies/a30429072/was-carrie-fisher-cgi-in-star-

    wars-the-rise-of-skywalker/

Make your own deepfakes [online app]. Deepfakes Web. (n.d.). Retrieved April 10, 2023, from

    https://deepfakesweb.com/

Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2022). Deepfakes

    generation and detection: State-of-the-art, open challenges, countermeasures, and way

    forward. *Applied Intelligence*, *53*(4), 3974–4026. https://doi.org/10.1007/s10489-022-

    03766-z

Martin, B. (2021, July 13). What was Anthony Bourdain searching for? GQ. Retrieved April 10,

    2023, from https://www.gq.com/story/anthony-bourdain-morgan-neville-roadrunner-

    documentary

Nash, K. (2022). Deepfakes, documentary and the dead: "I wasn't putting words into his mouth.

    I was just trying to make them come alive." *Journal of Media Ethics*, *37*(4), 291–292.

    https://doi.org/10.1080/23736992.2022.2118124

*New laws to better protect victims from abuse of intimate images*. GOV.UK. (2022, November 25). Retrieved March 23, 2023, from https://www.gov.uk/government/news/new-laws-to-better-protect-victims-from-abuse-of-intimate-images

Sharf, Z. (2021, July 26). *Lucasfilm hired the YouTuber who used deepfakes to tweak Luke Skywalker 'mandalorian' VFX*. IndieWire. Retrieved March 23, 2023, from https://www.indiewire.com/2021/07/lucasfilm-hires-deepfake-youtuber-mandalorian-skywalker-vfx-1234653720/

S.2065 - 116th Congress (2019-2020): Deepfake Report Act of 2019. (2019, October 29). https://www.congress.gov/bill/116th-congress/senate-bill/2065

H.R.6395 - 116th Congress (2019-2020): William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021. (2021, January 1). https://www.congress.gov/bill/116th-congress/house-bill/6395

S. Rept. 116-289 - IDENTIFYING OUTPUTS OF GENERATIVE ADVERSARIAL NETWORKS ACT. (2023, March 23). https://www.congress.gov/congressional-report/116th-congress/senate-report/289/1

Reuters. (2021, September 25). Deepfake Bruce Willis appears in Russian advert. YouTube. Retrieved April 10, 2023, from https://www.youtube.com/watch?v=Hp4jbs7ivSY

Westerlund, M. 2019. The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review, 9(11): 40-53. http://doi.org/10.22215/timreview/1282